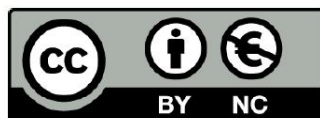


Els sistemes d'informació de qualitat en ciències aquàtiques: la interoperabilitat semàntica

Carmen Reverté Reverté



Aquesta tesi doctoral està subjecta a la llicència [Reconeixement- NoComercial –3.0. Espanya de Creative Commons](#).

Esta tesis doctoral está sujeta a la licencia [Reconocimiento - NoComercial – 3.0. España de Creative Commons](#).

This doctoral thesis is licensed under the [Creative Commons Attribution-NonCommercial- 3.0. Spain License](#).



Universitat de Barcelona. Departament de Biblioteconomia i Documentació.
Programa de Doctorat "Informació i Documentació en la Societat del
Coneixement"

Els sistemes d'informació de qualitat en ciències aquàtiques: la interoperabilitat semàntica

Carmen Reverté Reverté

Tesi doctoral presentada per a optar al títol de doctora en Informació i
Documentació per la Universitat de Barcelona, Abril 2014
Presentada per Carmen Reverté Reverté
Dirigida per Ernest Abadal Falgueras

SUMARI

0. INTRODUCCIÓ	18
0.1. Presentació	18
0.2. Marc conceptual	20
0.4. Mètode d'investigació	29
0.5. Estructura	33
0.6. Mapa conceptual	35
CAPÍTOL 1. EVOLUCIÓ DELS SISTEMES D'INFORMACIÓ DE QUALITAT...	36
1.1. Introducció	36
1.2. Ecologia de la informació.....	37
1.3. Sistemes distribuïts en un marc especialitzat	50
1.3.1. Tipologia de sistemes distribuïts.....	51
1.4. Sistemes d'informació de qualitat	52
1.4.1. Context i evolució	55
1.4.2. Context europeu.....	58
1.4.3. Context australià.....	60
1.4.4. Context internacional.....	62
1.5. Estat actual dels sistemes d'informació: evolució cap a portals semàntics	63
1.5.1. Tendències actuals: sistemes d'informació semàntics	69
1.5.2. Projectes i casos d'ús de biblioteques i el web semàntic	75
1.6. Conclusions	80
CAPÍTOL 2. COMPARATIVA DE MODELS DE SISTEMES D'INFORMACIÓ DISTRIBUÏTS	84
2.1. Introducció	84
2.2. Casos d'estudi	85
2.2.1. Cas d'estudi 1. INTUTE (2006 – juliol 2011).....	85
2.2.1.1. Descripció	85

2.2.1.2. Responsabilitat	85
2.2.1.3. Cobertura	85
2.2.1.4. Arquitectura de la informació	86
2.2.1.5. Desplegament de serveis	90
2.2.2. Cas d'estudi 2 – Vascoda (2003 - gener 2011)	91
2.2.2.1. Descripció	92
2.2.2.2. Responsabilitat	92
2.2.2.3. Cobertura	93
2.2.2.4. Arquitectura de la informació	93
2.2.2.5. Desplegament de serveis	97
2.2.3. Cas d'estudi 3. AustLII (1994? -)	97
2.2.3.1. Descripció	97
2.2.3.2. Responsabilitat	98
2.2.3.3. Cobertura	98
2.2.3.4. Arquitectura de la informació	98
2.2.3.5. Desplegament de serveis	100
2.2.4. Cas d'estudi 4. LIBRIS (2003? -): Linked data interface to the Swedish LIBRIS union catalog.....	101
2.2.4.1. Descripció	101
2.2.4.2. Responsabilitat	102
2.2.4.3. Cobertura	102
2.2.4.4. Arquitectura de la informació.....	102
2.2.4.5 Desplegament de serveis	104
2.2.5. Cas d'estudi 5. Europeana (novembre 2008 -)	106
2.2.5.1. Descripció	106
2.2.5.2. Responsabilitat	107
2.2.5.3. Cobertura	107

2.2.5.4. Arquitectura de la informació.....	108
2.2.5.5. Desplegament de Serveis	110
2.3. Comparativa entre els diferents sistemes d'informació de qualitat	114
2.3.1. Comparativa entre sistemes distribuïts de tipus passarel·la temàtica ..	115
2.3.1.2. Comparativa dels sistemes a través de la RI.....	119
2.3.2. Comparativa entre sistemes semàntics.....	120
2.3.2.1. Comparativa a nivell de serveis	120
2.3.2.2. Comparativa dels sistemes a través de la RI.....	123
2.4. Avaluacions dels sistemes d'informació de qualitat des de la perspectiva d'IQ	127
2.4.1. Metodologia	127
2.4.2. Nivell d'IQ dels casos d'estudi.....	128
2.5. Conclusions	134
CAPÍTOL 3. INDEXACIÓ I NAVEGACIÓ CREUADES	136
3.1. Introducció	136
3.2. Les passarel·les temàtiques	136
3.2.1. Context	137
3.2.2. Característiques de les passarel·les temàtiques.....	138
3.3. Navegació creuada i cerca creuada	139
3.3.1. Navegació creuada o <i>cross-browsing</i>	139
3.3.2. Mapeig de DDC per a la navegació creuada	141
3.3.3. Navegació creuada i cerca creuada	142
3.4. Indexació creuada	143
3.4.1. Indexació creuada i mapeig.....	144
3.5. Conclusions	146
CAPÍTOL 4. ESTÀNDARDS I FORMATS	149
4.1. Introducció	149

4.2. Estàndards i formats.....	158
4.2.1. Estàndards i formats per a la descripció de recursos i/o objectes digitals..	159
4.2.1.1. MARC	159
4.2.1.2. Dublin Core (DC)	161
4.2.1.3. Altres estàndards i formats d'estructures de dades en entorns digitals	169
4.2.1.4. Anàlisis i comparativa entre els formats.....	172
4.2.1.5 Resultats	179
4.2.2. Estàndards i formats per a l'intercanvi d'informació i interoperabilitat entre sistemes	180
4.2.2.1. Semàntica	180
4.2.2.2 Estàndards per a la descripció de sistemes organització del coneixement (KOS)	182
2.2.2.4. SKOS avançat	209
4.3. Integració dels formats semàntics en biblioteques	220
4.4. Discussió	223
4.5. Conclusions	226
CAPÍTOL 5. INTEROPERABILITAT SEMÀNTICA EN SISTEMES D'INFORMACIÓ DE QUALITAT.....	228
5.1. Introducció	228
5.2. Antecedents de la interoperabilitat semàntica: problemàtiques i solucions	230
5.3. Models d'interoperabilitat semàntica.....	232
5.4. Tendències: estandardització i interpretació.....	235
5.5. Estudis de casos: subject gateways (models distribuïts) i <i>linked data</i> (models ontològics)	237
5.5.1. Les subject gateways com a model d'interoperabilitat semàntica en sistemes d'informació distribuïts de qualitat	238
5.5.1.1. Model d'arquitectura d'informació d'una Subject Gateway	239
5.5.1.2. Requisits generals d'interoperabilitat semàntica en una SG	241

5.5.1.3. Arquitectura de la informació de les <i>subject gateways</i>	242
5.5.2. Estudis de casos en ciències aquàtiques i marines.....	248
5.5.2.1. Anàlisi de l'entorn dels sistemes d'informació en ciències aquàtiques i marines	249
5.5.2.2. Tendències: mapeig de vocabularis controlats en ciències aquàtiques i marines	252
5.5.3. <i>Linked data</i> : model d'interoperabilitat semàntica.....	255
5.6. Consideracions per implementar un prototipus de sistema d'informació de qualitat a nivell d'interoperabilitat	259
5.6.1. Model desitjat d'interoperabilitat semàntica.....	261
5.7. Conclusions	262
CAPÍTOL 6. INFORMACIÓ DE QUALITAT: PROPOSTA D'UN MODEL D'IQ PER UN SISTEMA D'INFORMACIÓ DISTRIBUÏT I ESPECIALITZAT	265
6.1. Introducció	265
6.2. Models d'IQ que s'apliquen des de la perspectiva de l'usuari.....	266
6.3. Models d'IQ no basats només amb l'usuari	269
6.4. Model IQ basat en els documents	269
6.5. Model d'IQ basat en el sistema d'informació	270
6.6. Altres tipus d'avaluacions de la IQ.....	275
6.7. Model IQ en un entorn digital (<i>Subject Gateways</i> , Portals, biblioteques digitals)	279
6.8. Anàlisi i comparativa de models d'IQ.....	283
6.9. Model IQ proposat per a un sistema d'informació de qualitat distribuït.....	285
6.10. Conclusions	291
CAPÍTOL 7. SISTEMES DE MAPEIG DE VOCABULARIS CONTROLATS EN CIÈNCIES AQUÀTIQUES	294
7.1. Introducció	294
7.2. Característiques d'alineació de vocabularis: vocabularis controlats <i>versus</i> ontologies	296
7.2.1. Funcionament: procés d'alineament.....	298

7.2.2. Problemàtiques del mapeig d'ontologies	299
7.2.3. Factors determinants per a l'alineació de vocabularis	300
7.2.4. Processos d'avaluació d'alineament.....	304
7.2.5. Avaluació de software de mapeig automàtic de vocabularis	306
7.3. Conclusions	309
CAPÍTOL 8. PROTOTIPUS DE MAPEIG: ESTUDI EXPERIMENTAL D'ALINEACIÓ DE TESAURUS EN CIÈNCIES AQUÀTIQUES I AFINS.....	310
8.1. Introducció	310
8.2. Metodologia	312
8.3. Anàlisi de les característiques i paràmetres dels vocabularis controlats: alineacions manuals	314
8.3.1. Estructura i relacions.....	317
8.3.2. Etiquetes (lèxic).....	318
8.3.3. Semàntica	320
8.3.4. Avaluació dels resultats: alineaments manuals	321
8.4. Execució mapeig automàtic/semi-automàtic de vocabularis	323
8.4.1. Fase de pre-mapeig	323
8.4.3. Fase post-mapeig: anàlisis d'alineaments.....	326
8.4.2.1. Estudi de mapeig de tesaurus a nivell monolingüe	327
8.4.2.2. Estudi de mapeig de tesaurus multilingüe.....	328
8.4.3. Avaluació dels resultats: procés de mapeig semi-automàtic	331
8.4.4. Discussió	337
8.5. Conclusions	339
9. CONCLUSIONS GENERALS	341
10. BIBLIOGRAFIA	352
11. GLOSSARI	384
12. ANNEXES	390
Annexes Capítol 1. Evolució dels sistemes d'informació de qualitat	390

Annex 1.1. Llistat de <i>Subject Gateways</i> consultades	390
Annex 1.2. Taula d'iniciatives i projectes de biblioteques i <i>linked data</i> sobre dades bibliogràfiques	392
Annex 1.3. Taula d'iniciatives i projectes de biblioteques i <i>linked data</i> sobre dades d'autoritats	394
Annex 1.4. Taula d'iniciatives i projectes de biblioteques i <i>linked data</i> sobre alineament de vocabularis.....	395
Annex 1.5. Taula d'iniciatives i projectes de dades heterogenies	400
Annex 1.6 Citacions	403
Annex 1.7. Objectes Digitals	403
Annex 1.8. Projectes d'agrupacions de col·leccions.....	405
Annex 1.9. Usos Socials.....	407
Annexes Capítol 2. Comparativa de Models de Sistemes d'Informació distribuïts de qualitat	412
Annex 2.1. Atributs associats a la informació de qualitat	412
Annex 2.2. Definició d'escalles de valors proposades per a un model IQ.....	413
Annex 2.2.1. Escala d'avaluació sumatòria	413
Annex 2.2.2. Escala d'avaluació per distància (absència/presència d'elements)	414
Annex 2.3. Qüestionari d'avaluació	414
Annex 2.4. Cas INTUTE.....	424
2.4.1. Interfícies	424
2.4.3. Cerques INTUTE	425
2.4.4. Registres de Metadades	427
2.4.5. Extracció resultats: e-mail, html, .txt i javascript	428
Annex 2.5. Cas VASCODA	428
2.5.1. Cerques VASCODA	428
2.5.2. Registres Vascoda	430
Annex 2.6. Cas AustLII.....	431

2.6.1. Interfície General	431
2.6.2. Interfície de Cerca Avançada.....	432
2.6.3. Cerques: resultats	432
2.6.4 Registres AustLII	433
2.6.5. Extracció d'Informació.....	434
Annex 2.7. Cas Libris	434
2.7.1. Interfície d'Inici	434
2.7.2. Cerques	434
2.7.3. Registres	438
2.7.4. Extracció d'Informació.....	439
Annex 2.8. Cas Europeana	441
2.8.1. Interfície	441
2.8.2. Cerques	441
2.8.3. Registres	443
2.8.4. Extracció d'Informació.....	445
Annex 2.9. Qüestionaris d'avaluació de sistemes d'informació de qualitat: anàlisi de dimensions d'IQ.....	446
2.9.1. Qüestionari Intute (2006 – juliol 2011)	446
2.9.2. Qüestionari Vascoda (2003-2011)	456
2.9.3. Qüestionari AustLII (1994 -).....	467
2.9.4. Qüestionari LIBRIS (2003 -).....	479
2.9.5 Qüestionari Europeana (novembre 2008 -).....	490
Annexes Capítol 3. Indexació i navegació creuades	502
Annex 3.1. Navegació creuada	502
Annex 3.2. Exemples de navegació creuades.....	504
Annex 3.3. Procés de mapeig intern de Renardus	508
Annexes Capítol 4. Estàndards i formats	509

Annex 4.1. Vocabularis i espais de noms SKOS.....	509
Annexes Capítol 5. Interoperabilitat semàntica en Sistemes d'Informació de Qualitat	511
Annex 5.1. <i>Subject Gateways</i> Interoperability	511
Annex 5.2. Qüestionari realitzat a biblioteques i centres de documentació del àmbit de les ciències aquàtiques i marines	514
Annexes Capítol 6. Informació de qualitat (IQ): proposta d'un model d'IQ per un sistema d'informació distribuït especialitzat.....	516
Annex 6.1. Models IQ.....	516
Annex 6.2. Taules comparatives d'autors i dimensions d'IQ	518
Annex 6.3. Taula de valors de dimensions d'IQ	525
Annex 6.4. Elements Bàsics Dublin Core	527
Annexes Capítol 7. Sistemes de Mapeig de vocabularis controlats en ciències aquàtiques	531
Annex 7.1. Tests amb NeOnToolkit.....	531
Annexes Capítol 8. Prototipus de mapeig: estudi experimental d'alineació de tesaurus en ciències aquàtiques i afins.....	537
Annex 8.1. ASFA, AGROVOC i GEMET	537
Annex 8.1.1. Biocomplexity Thesaurus NBII Thesaurus.....	563
Annex 8.2. Resultats mapeig semi automàtic: anàlisi qualitativa dels alineaments produïts i realitzades per avaluadors humans.....	587
8.2.1. Resultats de l'anàlisi qualitativa de l'avaluador 1	587
8.2.2. Resultats de l'anàlisi qualitativa de l'avaluador 2.....	587
Annex 8.3. Anàlisi qualitativa de l'experiment d'alineament de tesaurus multilingües (Avaluador 1, 10 estrats)	588
Stratum 1: Mostra aleatòria de 40 alineacions.....	588
Stratum 2: Mostra aleatòria de 20 alineacions.....	610
Stratum 3: Mostra aleatòria de 20 alineacions.....	626
Stratum 4: Mostra aleatòria de 20 alineacions.....	636
Stratum 5: Mostra aleatòria de 20 alineacions.....	646

Stratum 6: Mostra aleatòria de 20 alineacions.....	655
Stratum 7: Mostra aleatòria de 20 alineacions.....	665
Stratum 8: Mostra aleatòria de 20 alineaments	675
Stratum 9: Mostra aleatòria.....	685
Stratum 10: Mostra aleatòria.....	695
Annex 8.4. Anàlisi qualitativa de l'experiment d'alineament de tesaurus multilingües (Avaluador 2, anàlisi de 6 estrats o <i>stratums</i>).....	698

Índex de Taules

Taula 1. Models i propostes d'ecologia de la informació	39
Taula 2. Criteris de selecció i avaluació dels recursos d'Intute	88
Taula 3. Fitxa descriptiva: Intute, Vascosa i AustLII	119
Taula 4. Fitxa descriptiva: Libris i Europea.....	123
Taula 5. Intute: avaluació d'IQ.....	131
Taula 6. Vascode: avaluació d'IQ.....	132
Taula 7. AustLII: avaluació d'IQ.....	132
Taula 8. LIBRIS: Avaluació d'IQ.....	133
Taula 9. Europea: avaluació d'IQ	133
Taula 10. Exemple de camp de dades MARC.....	160
Taula 11. Exemple de camps mínims MARC21	161
Taula 12. Comparativa entre Dublin Core (XML) i Marc21.....	164
Taula 13. Elements Qualified Dublin Core (qDC). (Font: DCMI)	169
Taula 14. Interoperabilitat entre els formats d'estructura de dades DC, MODS i METS (Font: DCMI i <i>Library of Congress</i>).....	176
Taula 15. Exemple de l'element Creador (Font: Elaboració pròpia)	177
Taula 16. Subelements de l'element de descripció del format (Font: MODS; Library of Congress).....	178
Taula 17. Subelements de l'element "RelatedItem"	179

Taula 18. Exemple de tipus de document OWL	196
Taula 19. Projectes Desenvolupats sobre <i>linked data</i> i biblioteques.....	223
Taula 20. Exemple de Crosswalk: mapeig entre diferents formats de metadades (Font: Elaboració pròpia)	234
Taula 21. Models d'estandardització i interpretació (Font: Elaboració pròpia)	236
Taula 22. Relació entre la interoperabilitat semàntica i l'arquitectura de la informació (Font: Sebastià-Salat, M.; Reverté, C., 2011)	240
Taula 23. Exemple de nivells d'equivalència entre termes d'un descriptor d'un Tesauro A a un descriptor del Tesauro B (Font: Elaboració pròpia)	244
Taula 24. Tesauro rellevants en ciències aquàtiques (Font: Elaboració pròpia)	251
Taula 25. Característiques d'interoperabilitat semàntica entre diferents tipus de sistemes (Font: Elaboració pròpia).....	259
Taula 26. Model IQ per Stivilla (2007)	274
Taula 27. Dimensions d'IQ bàsiques per a una SG (Font: Elaboració pròpia)	285
Taula 28. Dimensió usuari.....	286
Taula 29. Taula Dimensió Col·lecció	288
Taula 30. Dimensió Sistema.....	289
Taula 31. Dimensió Serveis.....	291
Taula 32. Exemples d'alineaments de vocabularis en sistemes d'informació	295
Taula 33. Nombre de descriptors per cada concepte objecte d'estudi del mapeig manual.....	315
Taula 34. Alineacions resultants del procés de mapeig manual.....	317
Taula 35. Exemple d'heterogeneïtat entre estructures i relacions entre vocabularis controlats (tesaurus).....	318
Taula 36. Exemple sobre diferències entre etiquetes (ús diferent de singulars i plurals) i nivell de cobertura temàtica entre termes equivalents.	319
Taula 37. Exemple de termes preferits i no preferits en diversos idiomes en un mateix tesauro (AGROVOC) i en diversos tesaurus.....	320
Taula 38. Exemple de termes preferits i no preferits dintre un mateix idioma, i entre diferents vocabularis	320

Taula 39. Exemple d'equivalència simple a múltiple a través del multilingüisme	322
Taula 40. Exemple de diferències d'equivalències entre termes genèrics en tesaurus multilingües.....	322
Taula 41. Anàlisi estadística pre-mapeig.....	323
Taula 42. Anàlisi de Top Concepts.....	323
Taula 43. Diferències estructurals a nivell d'etiquetes alternatives i preferides	324
Taula 44. Diferències estructurals a nivell de relacions associatives	324
Taula 45. Nivell de profunditat jeràrquica dels tesaurus.....	325
Taula 46. Propietats objecte dels tesaurus.....	325
Taula 47. Propietats descriptives dels tesaurus	325
Taula 48. Mapeig de tesaurus monolingüe.....	327
Taula 49. Mapeig de tesaurus multilingüe (10 llengües)	328
Taula 50. Nivell d'equivalència o mapeig entre els tesaurus a partir de dos experiments.....	329
Taula 51. Resultats exemple de mapeig multilingüe <i>versus</i> mapeig monolingüe	330
Taula 52. Mostra de resultats dels alineaments del mapeig multilingüe.....	331
Taula 53. Percentatge total d'acord entre els dos avaluadors.....	332
Taula 54. Percentatge d'acord d'"exact match" entre els dos avaluadors	332

Índex de Figures

Figura 1. Representació dels factors IE en els Sistemes d'Informació de Qualitat (Font: Elaboració pròpia).....	46
Figura 2. Flux d'informació i models de comunicació en els sistemes d'informació (Font: Sebastià-Saltat, M.; Reverté, C., 2011)	47
Figura 3. Model de sistema d'informació centralitzat: Components i flux d'informació (Font: Sebastià-Saltat, M.; Reverté, C., 2011)	48
Figura 4. Model de sistema d'informació distribuït (Font: Sebastià-Salat, M.; Reverté, C., 2011).....	48

Figura 5. Representació percentual de SG's analitzades.....	64
Figura 6. Evolució de les SG's des de 1993 fins el 2010	67
Figura 7. Evolució de la literatura científica sobre biblioteques i web semàntica	73
Figura 8. Evolució de projectes relacionats amb el web semàntic i les biblioteques	75
Figura 9. Arquitectura <i>Cloud</i> (Font: Europeana Thought Lab).....	110
Figura 10. Exemple de cerca simple (Font: Europeana)	124
Figura 11. Exemple de cerca semàntica (Font: Europeana)	125
Figura 12. Exemple de relacions semàntiques d'un registre d'Europeana (Font: Elaboració pròpia)	126
Figura 13. Nivell d'IQ dels sistemes d'informació avaluats.....	128
Figura 14. Nivell d'IQ individual per dimensions.....	129
Figura 15. Formats tradicionals i nous formats: compatibilitat en l'àmbit d'aplicabilitat (Font: Sebastià-Salat; Reverté, C., 2011)	158
Figura 16. Exemple de Dublin Core (Font: RECERCAT)	165
Figura 17. Exemple MARC 21 amb sortida de dades Millenium (Font: CCUC)	165
Figura 18. Exemple de interrogació del sistema (Font: kukuxumus.com)	181
Figura 19. Exemple de descripció XML amb DC.....	184
Figura 20. Exemple 1 descripció de contingut amb XML.....	184
Figura 21. Exemple 2 descripció de contingut amb XML.....	184
Figura 22. Exemple 3 descripció de contingut amb XML.....	184
Figura 23. Exemple de graf simple.....	185
Figura 24. Exemple de descripció Pàgina web de la EEA.....	185
Figura 25. Grafo simple de sentència sobre sentència (reidificació)	187
Figura 26. Representació de graf RDFS	189
Figura 27. Exemple de propietat d'unió de classes (owl:unionOf).....	196

Figura 28. Exemple de Propietat d'intersecció de classes (owl:intersectionOf)	196
Figura 29. Exemple de propietat de complement (owl:complementOf)	196
Figura 30. Exemple d'equivalència de propietats (owl:equivalentProperty)....	196
Figura 31. Exemple d'equivalència entre classes (owl:equivalentClass)	197
Figura 32. Exemple d'equivalència entre individus (owl:sameAs)	197
Figura 33. Exemples de representació de conceptes (SKOS)	198
Figura 34. Exemple d'etiqueta "prefLabel" (SKOS)	199
Figura 35. Exemple d'etiquetat RDF literal d'etiqueta multilingüe (SKOS)	199
Figura 36. Exemple t'etiqueta't SKOS inconsistent	200
Figura 37. Exemple d'etiquetat SKOS consistent.....	200
Figura 38. Exemple t'etiqueta't consistent sobre el domini de les propietats SKOS	200
Figura 39. Exemple d'etiquetes alternatives o "prefLabel" (SKOS)	201
Figura 40. Exemple d'abreviatures representades amb "prefLabel" (SKOS) .	201
Figura 41. Exemple d'etiquetes alternatives o "altLabel" (SKOS).....	201
Figura 42. Exemple d'etiquetes lèxiques amb "prefLabel" (SKOS)	202
Figura 43. Exemple d'etiquetes "narrower" i "broader" (SKOS).....	203
Figura 44. Exemple de jerarquies transitives (SKOS)	203
Figura 45. Exemple de jerarquia transitiva amb etiqueta skos:broader.....	204
Figura 46. Exemple de relació associativa o "skos:related".....	204
Figura 47. Exemple de relació transitiva "skos:related"	205
Figura 48. Exemple de "Scope note" provinent de la Llista d'encapçalaments de matèries en català (LEMAC)	206
Figura 49. Exemple d'etiqueta "skos:definition" (Definició extreta de GEMET)	206
Figura 50. Exemple d'etiqueta "skos:example".....	206
Figura 51. Exemple d'etiqueta nota "skos:historyNote"	206
Figura 52. Exemple d'etiqueta nota "skos:editorialNote"	207

Figura 53. Exemple d'etiqueta nota "skos:changeNote"	207
Figura 54. Exemple d'ús t'etiqueta't d'idiomes en etiquetes de notes (SKOS)207	
Figura 55. Exemple d'esquema de conceptes (SKOS)	208
Figura 56. Exemple de propietat "skos:inScheme"d'esquema de concepte... 208	
Figura 57. Exemple de propietat "skos:hasTopConcept" d'esquema de concepte.....	208
Figura 58. Exemple de col·leccions de conceptes (SKOS)	210
Figura 59. Exemple de classe "skos:collection"	210
Figura 60. Exemple d'ordenació de col·leccions (skos:OrderedCollection)....	211
Figura 61. Exemple de vinculació de "skos:collection" amb la propietat "skos:broader"	211
Figura 62. Exemple de vinculació de col·leccions jeràrquiques a través "skos:Concept"	212
Figura 63. Exemple de propietat "skosxl:literalForm"	212
Figura 64. Exemple 1 de propietat "skosxl:labelRelation"	213
Figura 65. Exemple 2 de propietat "skosxl:labelRelation"	213
Figura 66. Exemple de post-coordinació amb consulta SPARQL	214
Figura 67. Exemple 1 de coordinació entre termes amb propietat "ex:coordinationOf" (SKOS)	214
Figura 68. Exemple 2 de coordinació entre termes amb propietat "ex:coordinationOf" (SKOS)	215
Figura 69. Exemple 3 de coordinació entre termes amb propietat "owl:intersectionOf" (OWL).....	215
Figura 70. Exemple de jerarquia transitiva (SKOS).....	215
Figura 71. Exemple 1 de jerarquia transitiva amb super-propietat "skos:broader"	216
Figura 72. Exemple 2 de jerarquia transitiva amb super-propietat "skos:broader"	216
Figura 73. Exemple de notacions CDU	216
Figura 74. Exemple de notació CDU amb SKOS	217

Figura 75. Exemple de notació amb "skos:prefLabel"	217
Figura 76. Exemple de propietat "rdfs:subPropertyOf" amb "skos broader" ...	218
Figura 77. Exemple 1 de combinació de formats semàntics (SKOS i RDF)...	218
Figura 78. Exemple 2 de combinació de formats semàntics (SKOS i OWL) ..	218
Figura 79. Exemple 3 de combinació de formats semàntics (SKOS i OWL) ..	219
Figura 80. Exemple 4 de combinació de formats semàntics (SKOS i OWL) ..	219
Figura 81. Exemple 4 de combinació de formats semàntics (SKOS i OWL) ..	219
Figura 82. Exemple 3 de combinació de formats semàntics (SKOS i OWL) ..	219
Figura 83. Exemple 1: Navegació creuada (Font: Renardus).....	242
Figura 84. Interoperabilitat entre vocabularis controlats (procés de mapeig, Doerr, 2000)	243
Figura 85. Tipologia de vocabularis controlats usats en biblioteques en ciències aquàtiques (Font: Elaboració pròpia)	250
Figura 86. Resultats ús de formats i estàndards en institucions de ciències aquàtiques i marines (Font: Elaboració pròpia)	252
Figura 87. Exemple de procés de mapeig a través de metadades (Font: Elaboració pròpia)	253
Figura 88. Model d'Arquitectura de la informació AGROVOC (Font: AGROVOC Architecture)	254
Figura 89. Model d'Arquitectura de la informació del projecte NeOn (Font: AGROVOC architecture)	255
Figura 90. Exemple de representació de les dades amb estructura de tripleta (Font: Elaboració pròpia)	257
Figura 91. Model d'interoperabilitat semàntica en sistemes d'informació de qualitat (Font: Elaboració pròpia)	262
Figura 92. Atributs d'IQ rellevants (Font: Elaboració pròpia).....	284
Figura 93. Model conceptual d'IQ proposat per a una SG (Font: Elaboració pròpia)	285
Figura 94. Exemple d'estructura de tesaurus (Font: Elaboració pròpia).....	296
Figura 95. Exemple gràfic d'ontologia	298
Figura 96. Exemple d'Skosificació (Font: Elaboració pròpia)	298

Figura 97. Procés de mapeig d'ontologies (Font: Elaboració pròpia)	299
Figura 98. Exemples de diferents tipus d'etiquetes (Font: Elaboració pròpia a partir del thesaurus AGROVOC)	301
Figura 99. Exemple de dificultats estructurals entre diferents vocabularis controlats (Font: thesaurus AGROVOC i GEMET).....	302
Figura 100. Exemple de relació Broader Term transitiva en una ontologia (Font: Elaboració pròpia amb suport guia SKOS-CoreGuide)	302
Figura 101. Tipus de relacions transitives expressades de forma jeràrquica (Font: Elaboració pròpia).....	303
Figura 102. Procés de mapeig amb avaluació dels resultats amb tests o alineaments de referència (R) (Font: Elaboració pròpia).....	305
Figura 103. Procés de mapeig amb un usuari com avaluador (Font: Elaboració pròpia)	306
Figura 104. Resultats de l'evolució de la qualitat del mapeig durant 2006-2008 (OAEI, 2008)	306
Figura 105. Interfície d'avaluació del sistema de mapeig Amalgame (Font: Elaboració pròpia)	307
Figura 106. Resultats de l'avaluació manual per Avaluador 1 (10 estrats).....	333
Figura 107. Resultats de l'avaluació manual per Avaluador 2 (6 estrats).....	333
Figura 108. Exemple de diferència estructural entre thesaurus	334
Figura 109. Exemple 1 de diferència de cobertura entre thesaurus a nivell cultural.....	334
Figura 110. Exemple 2 de diferència de cobertura entre thesaurus a nivell cultural.....	335
Figura 111. Exemple de diferència de cobertura a nivell jeràrquic entre thesaurus	335
Figura 112. Exemple de termes no equivalents	336

0. INTRODUCCIÓ

0.1. Presentació

L'objecte d'estudi de la nostra recerca és el que anomenem sistemes d'informació de qualitat (SIQ) en un entorn científic i distribuït.

La característica principal d'aquests tipus de sistemes és la possibilitat de vincular els recursos d'informació científics dipositats en diferents tipus de sistemes d'informació (biblioteques digitals, bases de dades, etc.) amb altres recursos d'informació relacionats i disponibles a Internet. L'objectiu final és l'obtenció d'una col·lecció d'informació científica d'alta qualitat.

El referent principal d'aquests tipus de sistemes ha estat la passarel·la temàtica INTUTE (ja no operativa), que juntament amb l'evolució del web semàntic, constitueixen els pilars fonamentals per al desenvolupament d'un futur sistema d'informació de qualitat especialitzat en ciències aquàtiques.

La missió principal d'aquesta tesi és fonamentar les bases per al desenvolupament d'un model de sistema d'informació de qualitat per a les ciències aquàtiques per mitjà de l'estudi dels sistemes d'informació de qualitat i dels sistemes d'informació de tipus distribuït.

El context temàtic des d'on s'enfoca la recerca, les ciències aquàtiques, és escollit per ser l'àrea professional de l'autora de la tesi. Aquest àmbit presenta unes necessitats d'informació difícilment comparables amb altres àrees del coneixement i que actualment no estan resoltes ni a nivell nacional ni internacional (Reverté, C. 2008, Reverté i Sebastià-Salat, 2009a). La problemàtica d'aquesta àrea del coneixement, que impulsa l'interès d'estudiar la viabilitat d'un sistema d'informació distribuït i de qualitat és la interdisciplinarietat de la pròpia àrea científica. És a dir, les ciències aquàtiques inclouen disciplines molt diverses i heterogènies i, per tant, un usuari (investigador del camp de les ciències aquàtiques) ha de cercar els continguts científics en múltiples i diversos recursos i sistemes d'informació per satisfer les seves necessitats d'informació. Conseqüentment, hi ha la necessitat de confeccionar eines més avançades de gestió i recuperació de la informació que permetin la col·laboració entre diferents institucions, per poder crear un sol punt d'accés a la informació científica i de qualitat, a través del qual un usuari podria accedir a aquesta informació heterogènia, multidisciplinària i multilingüe.

Pel que fa a l'abast geogràfic, aquesta recerca se centra a Europa, ja que es considera que hi ha un potencial interès en el desenvolupament d'un tipus de sistema d'informació col·laboratiu en l'àrea d'estudi. De fet, ja existeix una xarxa europea de biblioteques i centres de documentació especialitzades en ciències aquàtiques (EURASLIC), de en la qual, l'autora d'aquest treball és membre des de l'any 2006 i va ésser membre de l'executiva durant el període

2011-2013. Els objectius de la xarxa europea EURASLIC encaixen amb l'objectiu d'aquesta tesi, "foment de la cooperació a Europa i establiment de vincles amb altres biblioteques nacionals, regionals i internacionals en ciències aquàtiques i/o xarxes d'informació", i "dur a terme projectes conjunts, per millorar el flux, intercanvi i difusió d'informació en l'àrea de les ciències aquàtiques".

Els principals models de sistemes d'informació científica distribuïts i de qualitat que han existit són les passarel·les temàtiques, que van ser concebudes com a sistemes avançats de recuperació de la informació de qualitat. Apareixen juntament amb el desenvolupament d'Internet a mitjans de la dècada dels noranta i, en el seu recorregut històric, han estat enteses com a infraestructures d'informació acadèmica o de recerca, i posteriorment com a espais de col·laboració entre diferents comunitats científiques. Aquesta caracterització de passarel·les com a models de sistemes d'informació de qualitat, d'una banda, ve donada per la seva arquitectura de la informació basada en la cerca, indexació i navegació creuades, i d'altra banda, per les seves polítiques de cooperació i col·laboració basades en protocols d'intercanvi d'informació de qualitat (ús d'estàndards, criteris comuns d'avaluació de continguts, serveis col·laboratius, etc.). De fet, han estat considerades també com a webs per accedir a col·leccions d'alta qualitat per donar suport a la recerca en una determinada matèria o disciplina (Fischer i Neuroth, 2000).

Els sistemes d'informació distribuïts i de qualitat pertanyen a un sector que ha viscut un enorme creixement en els darrers anys i una important transformació en un entorn marcat pel desenvolupament de les tecnologies de la informació, tal i com es veurà en el primer capítol d'aquesta tesi. Aquests canvis s'han produït també per l'evolució de la pròpia web, marcada per un creixement desbordant (infoxicació, canvi de model dels usuaris: lectors-creadors de informació i l'*Open Access*), i la necessitat d'anar cap a una web semàntica, que pugui aportar informació de valor, rigor i qualitat. Així doncs, les passarel·les temàtiques també varen experimentar canvis en l'aplicació de la seva pròpia tecnologia, tal i com s'observa en la darrera fase de projectes de passarel·les destacats com és el cas d'INTUTE o VASCODA, on començaven a implementar tecnologia semàntica per tal de millorar els mecanismes d'intercanvi i la recuperació de la informació entre passarel·les. No obstant això, aquests projectes van quedar paralitzats al 2011 a causa de l'evolució de la tecnologia cap al *linked data* i també per la falta de finançament, ja que els recursos econòmics públics van prioritzar els repositoris institucionals¹ primer i després es va anar cap a l'evolució de les passarel·les i d'aquestes als sistemes d'informació semàntics de caràcter nacional o europeu (vegeu vegeu

¹Programa Europeu CIP-ICT promou del 2009 al 2013: "Open Access to scientific information, European Digital Library services, Digitising content for Europeana and Use of cultural heritage for education". *CIP-ICT Policy Support Programme*. EC. Information Society and Media, 2009. Url: http://ec.europa.eu/information_society/activities/ict_psp/documents/leaflet_call_3_cip_ict_psp_v.0309.pdf

Programa de finançament europeu CIP-ICT, 2012, on es parla d'*Open Data*, i es cita d'exemple el cas d'Europeana)².

Així doncs, en aquesta recerca es posa l'accent en la capacitat dels sistemes d'organització del coneixement (*Knowledge Organization Systems* o KOS) per vincular els recursos de biblioteques digitals existents amb altres recursos d'informació relacionats i disponibles a l'exterior (Internet), i d'aquesta manera configurar un sistema d'informació distribuït i de qualitat. És a dir, els vocabularis controlats, que donen suport al descobriment sistemàtic de recursos o enllaços disponibles a la xarxa (documents, objectes, llocs o serveis), i els vincula amb els recursos existents en el sistema d'informació, examinant l'estructura temàtica dels sistemes d'organització del coneixement emprats en la indexació i recuperació de la informació. Per tant, els KOS són considerats l'eina clau per a la interoperabilitat semàntica entre diferents tipus de sistemes d'informació.

0.2. Marc conceptual

0.2.1. Els sistemes d'informació científica de qualitat: model tipus passarel·la temàtica

Les característiques de les passarel·les temàtiques aporten grans avantatges a un sistema d'informació digital i distribuït d'un àmbit científic heterogeni i interdisciplinari com són les ciències aquàtiques. És a dir, són models de sistemes d'informació distribuïts i col·laboratius (models d'informació ecològics), que permeten que diferents institucions comparteixin recursos i serveis de informació per mitjà d'una sola interfície, sense haver de construir un nou sistema d'informació, ja que la tecnologia de la passarel·la temàtica permet la interoperabilitat amb els sistemes d'informació d'origen (institucions que integren la passarel·la), tal i com s'exposa en el capítol 1 (Evolució dels sistemes) i al capítol 3 (Indexació i navegació creuades), on es realitzen les revisions bibliogràfiques sobre aquests tipus de sistemes. Al mateix temps, les necessitats de informació d'un investigador, requereixen informació d'alta qualitat (contrastada, de tipus científic, etc.), i per tant, el model de sistema distribuït tipus passarel·la temàtica, que inclouen protocols d'informació de qualitat inherents, és el tipus de sistema d'informació més adequat per desenvolupar un sistema d'informació distribuït de tipus col·laboratiu i cooperatiu en ciències aquàtiques. Els motius són els següents (Reverté; Sebastià-Salat, 2009c):

- Capacitat d'interoperabilitat: aporta la capacitat de treball en xarxa (integració de diferents institucions i/o sistemes d'informació), a través

² CIP-ICT PSP work programme 2012. *CIP-ICT Policy Support Programme*. EC. Information Society and Media, 2012. Url: http://ec.europa.eu/information_society/activities/ict_psp/documents/cip_ict_psp_wp2012_adopied_01022_012.pdf

de l'ús d'estàndards d'indexació i classificació que s'utilitzen en gestió i recuperació de la informació (indexació, cerca i navegació creuada). Per exemple, ús de metadades estàndard (Dublin Core), suport de protocols per a la cerca d'informació (Z39.50, etc.), entre d'altres estàndards internacionals (llenguatges tipus: html, xml, rdf, etc.).

- Accés multilingüe a una col·lecció: a través dels llenguatges documentals (vocabularis controlats) i metadades, és possible accedir a recursos de informació sense la necessitat de traduir el contingut d'un document.
- Accés integrat als recursos impresos i electrònics: la infraestructura de les passarel·les, gràcies a les tecnologies d'aplicacions de protocols d'intercanvi d'informació com el Z39.50, fa possible la integració de la informació, d'una forma més àgil. Per exemple, facilita l'enllaç entre dades provinents d'un OPAC i dades provinents de la xarxa (indexació creuada)
- Recuperació de la informació i selecció de recursos de forma sostenible: aquests tipus de sistemes donen accés lliure a la informació de la xarxa a través d'un sol punt d'accés i fan que aquesta estigui recollida en un mateix sistema amb criteris de selecció comuns, i per tant, es recuperi informació de qualitat, i al mateix temps d'una forma sostenible (treball col·laboratiu).
- Eines de cerca creuada: amb aquesta tecnologia el sistema permet a l'usuari cercar informació provinent de múltiples sistemes d'informació (de procedència diversa: temàtica, geogràfica, etc.), a través d'una sola interfície.
- Navegació creuada: generalment permet a l'usuari cercar informació per matèries des d'una mateixa interfície però a diversos sistemes d'informació i/o passarel·les temàtiques. És a dir, hi ha interoperabilitat semàntica d'un sistema d'organització del coneixement a diversos tipus de KOS que alberguen els sistemes d'informació d'origen.
- Evitar la duplictat d'esforços: s'estableixen mecanismes de cooperació i coordinació per tal d'establir criteris comuns que evitin la duplictat d'esforços entre les diferents institucions que integrin la passarel·la, com per exemple, la catalogació compartida o la identificació i selecció de informació de forma distribuïda.
- Cooperació: les passarel·les temàtiques estableixen sistemes de cooperació i coordinació fins i tot entre diferents països, perquè les institucions integrants comparteixen un mateix objectiu (donar una cobertura temàtica i, en la majoria dels casos, també donar una major cobertura geogràfica a una comunitat d'usuaris), facilitant així la millora de serveis d'informació oferts de manera individual.

Referent a aquest entorn conceptual, en el capítol 2 (Comparativa de models), es caracteritzen els sistemes d'informació de qualitat i, a més, es realitza una

anàlisi comparativa (recerca qualitativa i quantitativa) entre diferents tipus de sistemes d'informació distribuïts que identifiquen a les passarel·les com a sistemes d'informació de qualitat. A més, al capítol 6 (Informació de qualitat), es du a terme un estudi bibliogràfic sobre diferents estudis realitzats en l'àmbit de la informació de qualitat, on també es caracteritzen aquests sistemes d'informació.

D'altra banda, en les revisions bibliogràfiques es van identificar les problemàtiques associades a la interoperabilitat entre els diferents sistemes d'informació que poden integrar una passarel·la, que principalment són problemàtiques de tipus semàntic i terminològic, provinents de l'ús de diferents vocabularis controlats. La utilització de diferents llenguatges documentals (a causa de la heterogeneïtat de la informació), provoca dificultats en el procés de indexació i recuperació de la informació, tal i com s'exposa en el capítol 5. En aquest sentit, es va procedir a analitzar noves tecnologies que suposen una millora a nivell semàntic (sistemes, llenguatges o formats i estàndards semàntics).

0.2.2. La tecnologia semàntica en sistemes d'informació

Els sistemes d'informació semàntics provenen de l'evolució del web semàntic (*Semantic web*, en anglès), un conjunt d'activitats desenvolupades pel *World Wide Web Consortium (W3C)*, que es basen en la creació de tecnologies per publicar dades llegibles per aplicacions informàtiques. Aquesta idea és promoguda per Tim Berners-Lee³ i es basa en la incorporació de metadades semàntiques i ontològiques (vocabularis controlats) que descriuen el contingut, el significat i la relació entre les dades publicades al *World Wide Web* i, evidentment, en la seva extensió es pot aplicar a qualsevol sistema d'informació que sigui publicat al web i que vulgui incorporar aquesta semàntica.

Linked data, un dels projectes més destacats del web semàntic, té per objectiu utilitzar la xarxa per connectar dades que no s'han relacionat anteriorment, o utilitzar la web per reduir les barreres per a la interconnexió de les dades que són vinculades amb altres mètodes. Dins el *Linked Data Group (W3C)*, el *Library Linked Data Incubator Group* té la missió d'augmentar la interoperabilitat global de dades provinents de biblioteques per posar-les a disposició al web. Per fer possible aquesta interconnexió entre dades disponibles al web és necessari utilitzar estàndards aprovats pel W3C com l'ús de URIs (*Uniform Resource Identifier*) per a identificar recursos de forma uniforme, i el format RDF, per a la descripció de recursos a la xarxa, entre d'altres formats semàntics (OWL, SKOS, SPARQL).

³ Berners-Lee, Tim. 1998. Semantic web Road map. Url: <http://www.w3.org/DesignIssues/Semantic.html>

Però el web semàntic no tracta només de fer disponibles les dades al web (accessibilitat i interoperabilitat), sinó que l'objectiu principal és l'establiment de connexions entre les dades, de manera que una persona o una màquina poden explorar les dades existents al web i recuperar la informació d'una forma més precisa.

Així doncs, en aquest estudi s'inclou una presentació general del nou paradigma web, el web semàntic, i de com aquest afecta els sistemes d'informació distribuïts (concretament, els que inclouen dades de biblioteques i centres de documentació). En el capítol 1 (Evolució dels sistemes d'informació) es fa una revisió bibliogràfica sobre el web semàntic i la seva aplicació en sistemes d'informació existents en biblioteques i centres de documentació, així com una revisió bibliogràfica sobre els projectes de sistemes d'informació que integren aquesta tecnologia. En aquest sentit cal destacar que els estudis sobre sistemes d'informació de tipus documental o bibliogràfic són menys abundats que els existents sobre altres tipus de dades com per exemple d'àmbit mèdic, sobre biodiversitat o dades geoespaciales. La revisió bibliogràfica centrada en els aspectes més tecnològics dels sistemes semàntics, s'estudia posteriorment en el capítol 4 (Formats i estàndards) i en el capítol 5 (Interoperabilitat semàntica), on s'explora la potencialitat de la tecnologia a nivell de sistemes d'informació de tipus distribuït.

En el capítol 2 (Comparativa de models de sistemes d'informació) es realitzen diversos estudis de cas on s'analitzen i es comparen diferents tipus de sistemes (semàntics i no semàntics) amb implicació dels professionals de l'àmbit de la biblioteconomia i la documentació, on es defineixen les característiques i el valor afegit dels sistemes d'informació semàntics com són la indexació, navegació i cerca creuades, produïdes pel mapeig de vocabularis controlats, característiques principals que assegurin la recuperació de la informació de qualitat, des d'una perspectiva documental. En aquest sentit, aquest treball, en la línia de les aportacions realitzades pel *Library Linked Data Incubator Group*, du a terme d'una banda, una revisió bibliogràfica sobre la tecnologia de mapeig de vocabularis (capítol 1, part 3), i de l'altra banda elabora un prototip de mapeig de vocabularis controlats en l'àmbit de les ciències aquàtiques i marines capítol 2 part 3). Aquests darrers estudis pretenen ser la base per corroborar que és possible la interoperabilitat semàntica entre sistemes d'informació distribuïts (problemàtiques no resoltes en les passarel·les temàtiques), i també estudiar la potencialitat o usos futurs dels sistemes semàntics, com poder resoldre la necessitat de crear o adaptar llenguatges documentals en el camp de les ciències aquàtiques i marines per tal de millorar o assegurar la indexació i recuperació de la informació de qualitat en sistemes d'informació distribuïts.

Nosaltres considerem que aquestes dues aproximacions de sistemes d'informació distribuïts semàntics i no semàntics de tipus passarel·la temàtica

no són incompatibles. Ambdues línies de recerca poden confluir per al disseny i construcció d'un sistema d'informació distribuït en ciències aquàtiques, ja que les passarel·les temàtiques són models fonamentals com a ús de protocols d'informació de qualitat i de sistemes col·laboratius, i els sistemes semàntics estan desenvolupant una tecnologia basada en el *linked data* que pot ser adaptada en un tipus de sistema de informació de tipus passarel·la temàtica, perquè comparteixen la mateixa metodologia quant a la indexació, cerca i navegació creuades, el mapeig de vocabularis controlats, la diferència, els formats i estàndards emprats, ja que en la web semàntica garanteixen una major interoperabilitat semàntica.

Per últim, cal afegir que el context del web semàntic, així com la bibliografia sobre aquesta línia de recerca, creix constantment i ràpida, de forma que l'evolució de la tecnologia és contínua. No obstant això, dintre del marc de la biblioteconomia i la documentació hi ha escassa literatura al respecte i no s'ha trobat cap estudi similar dintre de l'àmbit internacional i disciplinari estudiat. En el camp de les tecnologies de la informació sí es troben estudis específics sobre la tecnologia semàntica, cal destacar la tesis de *van Hage* (2008) sobre sistemes d'alineament d'ontologies (inclou experiments sobre vocabularis controlats en l'àmbit agroalimentari) i els esforços duts a terme per la FAO (AGRIS i vocabularis controlats) quant al desenvolupament del *linked data* en relació a sistemes d'informació relacionats amb l'àrea d'estudi, quant a sistemes distribuïts del camp de la biblioteconomia i la documentació com és el cas d'Europeana (*Antoine, et, al.* 2007, 2009, 2013, etc.).

0.2.3. Les ciències aquàtiques

El context temàtic on s'emmarca aquest estudi són les ciències aquàtiques (que inclouen també les ciències marines)⁴ que es caracteritzen per un alt grau d'interdisciplinarietat. En estudis previs realitzats (Reverté, C., 2006), ja s'havien identificat com a mínim quatre grans disciplines que formen part de les ciències aquàtiques: biologia, ecologia, enginyeria i medi ambient. A part de les diferències quant a l'abast i definició del propi camp "ciències aquàtiques" exposades també en estudis previs (DMA⁵ i RAMSAR⁶, etc.), que cal tenir en compte també en la indexació i recuperació de la informació.

Aquesta interdisciplinarietat fa que ens trobem al davant d'una àrea científica heterogènia que implica que l'entorn informacional estigui disgregat, tal i com

⁴ La nomenclatura segueix les directrius de la xarxa EURASLIC, que inclou ciències marines en les ciències aquàtiques.

⁵ Directiva Marco del Agua (2000/60/CE) (DOCE, 2000).URL:
<http://mediambient.gencat.net/aca/es//planificacio/directiva/inici.jsp>

⁶ Sistema de Clasificación de Tipos de Humedales Ramsar. (2006). Dins: "Marco estratégico y lineamientos para el desarrollo futuro de la Lista de Humedales de Importancia Internacional de la Convención sobre los Humedales (Ramsar, Irán, 1971)". Apéndice B. URL:
http://www.ramsar.org/key_guide_list2006_s.htm

s'havia demostrat en els estudis previs (Reverté, C., 2006; Reverté, C., 2008; Reverté i Sebastià-Salat, 2009b), on no només s'ha de fer front a la interdisciplinarietat, sinó que també cal fer front a la diversitat de fonts d'informació existents (articles científics, publicacions de congressos, literatura grisa, monografies, etc.), i a la procedència d'aquesta informació (diferents tipus d'organitzacions gestores d'informació: universitats, centres de recerca, organismes governamentals, biblioteques i centres de documentació, xarxes científiques, etc.), assumint l'existència i ús de diferents formats i estàndards com per exemple per a la descripció de dades i/o protocols d'intercanvi d'informació.

A més, aquestes anàlisis permeten observar la manca de sistemes de gestió d'informació i documentació digitals a nivell col·laboratiu, tal i com s'exposa en el capítol 1 (Evolució sistemes), on es realitza una recerca i anàlisi bibliogràfica sobre els tipus de sistemes d'informació distribuïts i de qualitat, concretament dintre l'àmbit científic en general i en l'àmbit de les ciències aquàtiques. Per tant, es manifesta la necessitat de crear i dissenyar sistemes d'informació i documentació de qualitat tipus passarel·la temàtica que inclouen tecnologia semàntica.

Finalment, quant als sistemes d'informació existents en l'àmbit de les ciències aquàtiques, també es mostra la necessitat de crear o adaptar llenguatges documentals d'aquest camp, que ajudin a la indexació i recuperació de la informació en sistemes distribuïts i/o de tipus col·laboratiu, tal i com es presenta en els capítols 5 i 8. En el capítol 5 (Interoperabilitat semàntica, estudi de cas ciències aquàtiques i marines) es realitza un estudi bibliogràfic i una recerca qualitativa i quantitativa a través de qüestionaris realitzats a professionals de la informació especialitzats en ciències aquàtiques i marines. En el capítol 8 (Estudi experimental) s'estudien les diferències i els nivells de interoperabilitat existents entre els diferents vocabularis controlats utilitzats en els sistemes d'informació relacionats amb les ciències aquàtiques.

0.3. Objectius

Un cop presentat l'objecte d'estudi, els antecedents i el marc conceptual, es procedeix a definir els principals objectius d'aquesta recerca.

L'objectiu principal és estudiar i avaluar els sistemes d'informació distribuïts de qualitat i els seus estàndards per tal d'aplicar-lo a l'àmbit de les ciències aquàtiques. Concretament, l'estudi es centra en els factors claus dels sistemes distribuïts, intercanvi de la informació i la recuperació de la informació a través dels sistemes d'organització del coneixement, com a mecanismes per assegurar la interoperabilitat semàntica d'un sistema d'informació distribuït i de qualitat.

Es tracta d'un àmbit on es detecta la mancança i la necessitat de confeccionar eines de gestió i recuperació de la informació que permetin la col·laboració entre diferents institucions per tal de fer front a la dispersió i heterogeneïtat del coneixement científic d'aquesta àrea interdisciplinària. Aquesta necessitat ha estat detectada per la doctoranda en el seu exercici com a professional de la informació en un institut de recerca especialitzat en tecnologia agroalimentària, on ha exercit de documentalista científic en l'àmbit de les ciències aquàtiques des de l'any 2005.

Fins el moment existeixen pocs estudis i projectes centrats en sistemes d'informació distribuïts en l'àmbit de les ciències aquàtiques i marines i/o basats en la intercanvi de la informació i la recuperació de la informació d'aquesta disciplina. És per això que al llarg d'aquesta recerca s'han examinat treballs representatius d'altres disciplines afins i/o d'àmbit general en aquells casos on és destacable el desenvolupament tecnològic de sistemes d'informació distribuïts.

Aquest objectiu general s'estructura en tres **objectius específics** que es defineixen a continuació:

1. Analitzar els antecedents, les característiques, evolució i tendències dels SIQ, per tal d'assentar les bases per al desenvolupament d'un model futur de SIQ de tipus distribuït i en un entorn científic.

Es tracta d'establir quines són les característiques fonamentals dels models de sistemes d'informació distribuïts de qualitat, i disposar de les bases d'un disseny futur de sistema d'informació distribuït i de qualitat en l'àmbit de les ciències aquàtiques. D'aquesta forma, a part de situar el context dels sistemes d'informació existents en l'àmbit temàtic d'estudi (debilitats i fortaleces), els sistemes distribuïts i l'estudi de la pròpia ecologia de la informació que envolta aquests sistemes, es disposa d'un model descriptiu del sector dels sistemes d'informació distribuïts, principalment de tipus passarel·la temàtica. I al mateix temps s'estableixen les principals tendències en l'evolució dels sistemes d'informació de qualitat de tipus distribuït.

Al mateix temps, es realitza una anàlisi comparativa de diferents models existents de sistemes d'informació distribuïts i de qualitat (capítol 2), que permeten observar l'evolució dels sistemes d'informació distribuïts i les tendències futures. Aquesta anàlisi es du a terme des de tres perspectives (la recuperació de la informació del sistema, els serveis oferts als usuaris i la col·lecció), per tal de conèixer les característiques, funcionament i altres elements clau d'aquests tipus de models amb la finalitat de definir quin és el tipus de sistema distribuït de qualitat amb més valor afegit. A més a més,

d'establir un protocol per a l'anàlisi i avaluació de sistemes d'informació distribuïts de qualitat. Finalment, es realitza una proposta pròpia, basada en indicadors, extrets i analitzats del capítol d'Informació de qualitat (capítol 6), que serveixen per a l'avaluació de sistemes d'informació distribuïts i de qualitat en general, per tal de poder ser aplicat en qualsevol àmbit temàtic d'estudi.

2. Estudiar i caracteritzar els principals elements que fan possible el desenvolupament dels sistemes d'informació distribuïts i de qualitat en ciències aquàtiques

Aquests elements són bàsicament quatre: una arquitectura de la informació basada en la indexació i navegació creuades (model tipus passarel·les temàtiques), els formats i estàndards de descripció bibliogràfica i objectes digitals aplicats en l'àmbit de les ciències aquàtiques i altres ciències afins, l'estudi de la interoperabilitat semàntica com a element clau per assegurar l'intercanvi d'informació i la compatibilitat entre diferents sistemes d'informació que puguin estar integrats en un sistema d'informació distribuït i de qualitat, i finalment, l'estudi de context de la "informació de qualitat" com a característica principal dels sistemes d'informació de qualitat.

Pel que fa a la indexació i navegació creuades s'analitzen els processos de recuperació de la informació de les passarel·les temàtiques com a model base o definitori de sistemes de gestió i recuperació de la informació de qualitat. Aquest estudi permet també conèixer l'estat de l'art sobre la indexació i navegació creuades i conèixer la recerca actual sobre aquest camp i les possibilitats de desenvolupament futur d'un sistema d'informació distribuït de qualitat en ciències aquàtiques.

Quant a l'estudi dels estàndards i formats, es tracta de situar l'estat actual d'aquests en els sistemes d'informació en l'àmbit especialitzat d'interès i/o relacionats. A més a més, s'estudien com a factor clau per a la interoperabilitat entre diferents sistemes d'informació, és a dir, quina és la viabilitat de creació d'un sistema distribuït de qualitat en l'àrea objecte d'estudi, segons els estàndards i formats utilitzats. Concretament, aquest estudi se centra bàsicament en els formats i estàndards d'estructura de dades (metadades que afecten al nivell sintàctic i semàntic) per a la descripció i recuperació de la informació, que són essencials per a la interoperabilitat entre els sistemes d'informació. Cal ressaltar que l'anàlisi d'altres tipus de dades claus per a la interoperabilitat entre sistemes d'informació, com per exemple, les dades categòriques o factuais, s'estudien en altres capítols (interoperabilitat semàntica).

Referent a l'estudi de la interoperabilitat semàntica, es tracta d'analitzar els diferents tipus de nivells d'interoperabilitat semàntica, sobretot en sistemes

d'informació objecte d'estudi (àmbit de recerca, distribuïts i de qualitat), i les seves característiques com a factors claus per tal d'assegurar la interoperabilitat en sistemes d'informació distribuïts i de qualitat. Cal tenir en compte, que la recerca actual sobre interoperabilitat es divideix en múltiples contextos, i tots ells tenen com a objectiu comú la integració, gestió i accessibilitat de la informació, però aquest estudi es focalitza en la interoperabilitat semàntica perquè és el nivell d'interoperabilitat corresponent a la vessant de la biblioteconomia i documentació.

Finalment, pel que fa a l'anàlisi de context de la "informació de qualitat", es realitza una avaluació de l'estat de l'art sobre sistemes d'informació de qualitat, així com de l'evolució del concepte d'informació de qualitat al llarg de la literatura científica, per tal de definir les característiques i models d'informació de qualitat (IQ) més adequats per als sistemes d'informació distribuïts en un àmbit disciplinar. Conseqüentment, es desenvolupa una metodologia pròpia d'identificació i avaluació de sistemes d'informació de qualitat, és a dir, s'estableix un model d'informació de qualitat basat en dimensions i atributs d'IQ que s'adeqüen al model de sistema d'informació proposat.

3. Desenvolupar un estudi experimental que demostrï l'eficàcia dels mètodes d'interoperabilitat semàntica en sistemes d'informació de qualitat en ciències aquàtiques.

D'una banda, s'estudia la tecnologia de mapeig de vocabularis controlats, presentada com a mecanisme de connexió entre diferents sistemes d'organització del coneixement, per tal d'estudiar i analitzar tant la tecnologia de mapeig de vocabularis controlats com els vocabularis controlats més destacats de l'àrea d'estudi (tesaurus en ciències aquàtiques i/o ciències afins), per tal d'avaluar la viabilitat del desenvolupament d'aquesta tecnologia en un sistema d'informació distribuït.

D'altra banda, es realitza un estudi experimental d'alineació de tesaurus objecte d'estudi, per tal de corroborar la hipòtesis que l'alineament de vocabularis controlats (prèviament convertits en formats ontològics) funciona com un mecanisme de connexió entre diferents models conceptuals d'organització del coneixement. Aquesta recerca es du terme des del punt de vista de la interoperabilitat entre els tesaurus multidisciplinaris i multilingües. Cal afegir que s'escullen els tesaurus com a vocabularis controlats per als experiments d'alineaments, perquè són els principals sistemes d'organització del coneixement emprats per a la indexació i recuperació d'informació en sistemes d'informació especialitzats. Aquest objectiu persegueix també estudiar els sistemes de mapeig i tècniques d'alineaments d'ontologies, com a mètodes i eines que poden ajudar i/o millorar els sistemes d'informació i documentació en general i els

sistemes d'informació de tipus distribuït. Al mateix temps que es proposen recomanacions per tal de desenvolupar o millorar eines futures de mapeig de vocabularis controlats, des de la perspectiva de les ciències de la informació i la documentació.

0.4. Mètode d'investigació

Al llarg d'aquesta tesi s'han utilitzat diferents metodologies de recerca adaptades als objectius presentats.

El mètode seguit és una recerca avaluativa que inclou diferents tècniques:

0.4.1. Revisió bibliogràfica

Les fonts d'informació utilitzades en les revisions bibliogràfiques han estat diverses i s'han centrat en l'àmbit de la biblioteconomia i la documentació però també en l'àmbit de les tecnologies de la informació i la comunicació: Emerald Mangement Xtra, Library, Information Science & Technology Abstracts, LISA: Library and Information Science Abstracts (CSA), Periodicals Index Online (PIO), web of Science (Social Science and Science Citation Index and Proceedings), entre d'altres recursos web que proporcionen informació bibliogràfica com Google Scholar o webs de comunitats científiques específiques (AIMS-FAO, Dublin Core Metadata Initiative Home, ECDL, NKOS, MMI, IAMSLIC i EURASLIC o W3C), entre d'altres recursos citats al llarg del treball.

Aquesta metodologia es utilitza com a tècnica complementària a d'altres tècniques com l'anàlisi de contingut, i s'ha utilitzat per als capítols 1, 3, 5, 6 i 7.

En el primer capítol, d'evolució dels sistemes d'informació de qualitat, es realitzen diferents revisions bibliogràfiques centrades en la definició del context que envolta els sistemes d'informació de qualitat i els sistemes d'informació distribuïts. En aquesta part de l'estudi no es fa restricció de les cerques bibliogràfiques a nivell geogràfic o temàtic. Primerament, es realitza un estudi de l'estat de l'art que permet definir l'evolució i abast geogràfic dels sistemes distribuïts de qualitat, i després, es realitza un estudi que permet determinar l'evolució dels sistemes d'informació de qualitat i de tipus distribuït. D'entrada la literatura científica sobre el tema és molt àmplia, així que s'opta per incloure només aquelles paraules clau que permeten definir què és un sistema d'informació distribuït i de qualitat, i concretament de tipus passarel·la temàtica, és a dir, s'opta per recollir les definicions descrites per Koch (2000) a l'hora de definir les equacions de cerca bibliogràfica pertinents. La segona part de la revisió bibliogràfica, a mode de mètode exploratori, consisteix en identificar si existeixen sistemes d'informació distribuïts i de qualitat en l'àmbit de les ciències aquàtiques, i es corrobora la seva inexistència. Així doncs, s'amplia la

recerca a disciplines afins (agricultura, medi ambient, ecologia, biologia, etc.) sense especificar àmbit geogràfic.

En el tercer capítol (La Indexació i navegació creuades), l'estudi bibliogràfic es centra sobre l'arquitectura de la informació de les passarel·les temàtiques, que és l'element clau que les defineix com a sistemes d'informació de qualitat (la indexació i navegació creuades).

En el cinquè capítol, la revisió bibliogràfica es centra en els sistemes d'interoperabilitat semàntica (antecedents, models i tendències), centrats en sistemes d'informació de qualitat i de tipus distribuït, entre altres projectes representatius a nivell europeus i internacional a nivell d'interoperabilitat semàntica en sistemes d'informació en general.

L'estudi bibliogràfic dut a terme en el sisè capítol, la informació de qualitat, permet identificar els treballs més representatius de l'àrea de la informació de qualitat, com els estudis específics centrats en l'avaluació de la IQ en entorns digitals, essent el coneixement base per al desenvolupament d'un model d'IQ adaptat a l'objecte d'estudi: sistemes d'informació de qualitat en ciències aquàtiques

Finalment al setè capítol (Sistemes de mapeig de vocabularis controlats en ciències aquàtiques) també es realitza una revisió bibliogràfica. Aquest estudi, es centra en la identificació de sistemes d'informació o projectes relacionats amb aplicació de la tecnologia semàntica a través dels vocabularis controlats. La recerca està centrada amb les tècniques d'alineament de vocabularis controlats, que són el mecanisme que permet la connexió entre diferents models conceptuals d'organització del coneixement.

La informació recollida en els diferents estudis bibliogràfics respon a un dels principals objectius perseguits en aquesta tesi: conèixer els antecedents, característiques i tendències dels sistemes d'informació de qualitat distribuïts, els sistemes d'informació en les ciències aquàtiques i la tecnologia aplicada per fer possible la interoperabilitat semàntica en sistemes d'informació de qualitat i distribuïts.

0.4.2. Anàlisi de contingut

Aquesta metodologia s'ha utilitzat en tots els capítols, excepte el darrer capítol sobre recerca experimental. Cada capítol utilitza diversos instruments per realitzar l'anàlisi de contingut, segons la formulació dels seus objectius.

En general la metodologia és emprada per analitzar els models de sistemes d'informació de qualitat, concretament sistemes de tipus passarel·la temàtica i sistemes d'informació semàntics.

En el primer capítol, a través de revisions bibliogràfiques realitzades prèviament sobre els sistemes d'informació de qualitat i distribuïts, es realitza l'anàlisi de contingut que permet descriure cada tipus de sistema segons la seva definició, categorització, abast (temàtic i geogràfic), característiques principals, diferències, tendències, etc. A posteriori, es realitza un anàlisi de contingut sobre la bibliografia recollida sobre l'àrea de les ciències aquàtiques i disciplines afins (agricultura, medi ambient, ecologia, biologia, etc.) sense especificar àmbit geogràfic, degut a que la literatura científica d'aquesta àrea quant a sistemes d'informació és més escassa i es pretén conèixer àmpliament l'estat de l'art sobre els sistemes d'informació distribuïts i/o de qualitat en l'àmbit de les ciències aquàtiques. Finalment amb les dades recollides d'aquest segon estudi exploratori, es recullen les característiques descriptives dels projectes més rellevants identificats, per tal de definir i exemplificar quins són els sistemes d'informació distribuïts models a seguir en l'àmbit de les ciències aquàtiques. En aquesta part de l'estudi, només s'han inclòs casos de projectes o prototipus que tenien una descripció complerta i detallada sobretot dins la descripció pròpia com a sistemes d'informació de qualitat o que contempnen el "control de qualitat" dintre dels seus objectius. Per acabar de concloure el primer capítol en relació al context dels sistemes d'informació de qualitat (SIQ), es van desenvolupar entrevistes via e-mail, a diferents responsables de SIQ de tipus passarel·la, per tal conèixer l'estat actual i futur d'aquests tipus de sistemes. No obstant això, com s'explica en el capítol 1, només es va obtenir resposta per part del govern australià.

Quant al segon capítol, es realitza un estudi descriptiu, on a través d'estudis de casos, s'analitzen els models de sistemes d'informació de qualitat, concretament sistemes de tipus passarel·la temàtica i sistemes d'informació semàntics. A continuació, es realitza una anàlisi comparativa entre els diferents casos estudiats on, per mitjà de mètodes de recerca qualitativa i quantitativa, s'analitzen els sistemes des de tres perspectives: la recuperació de la informació, els serveis oferts als usuaris i la pròpia col·lecció.

Pel que fa a l'avaluació qualitativa, s'utilitzen les tècniques d'observació dintre la dinàmica del comportament del sistema amb l'accés de la informació, és a dir, s'observen bàsicament les dimensions de la recuperació de la informació i de la col·lecció, i s'utilitzen unes fitxes descriptives per recollir les observacions. Després s'utilitza l'avaluació quantitativa per tal d'objectivar i donar una major solidesa als resultats obtinguts en el procés d'observació dels sistemes. En l'anàlisi quantitatiu, s'utilitza una escala d'avaluació on es puntuen diferents indicadors considerats imprescindibles per avaluar la IQ de cada sistema. Les dimensions de la col·lecció o de l'usuari no són avaluades perquè els sistemes analitzats no són comparables, ja que són sistemes d'informació amb especialització temàtica diferent. No obstant això, pel que fa a la col·lecció, si

s'analitza l'adequació de la col·lecció al tipus d'usuari/audiència a la que es dirigeix.

En els següents capítols, es desenvolupen diferents metodologies enfocades a l'anàlisi de contingut per tal d'identificar les característiques principals dels sistemes d'informació de qualitat i de tipus distribuït, i conèixer així les característiques principals d'aquests models, la tecnologia semàntica emprada per assegurar la interoperabilitat d'aquests tipus de sistemes i conèixer les tendències d'aquests models de sistemes d'informació.

En el tercer capítol s'estudia l'arquitectura de la informació dels sistemes d'informació distribuïts de tipus passarel·la temàtica a través d'estudis de casos. Les anàlisis descriptives pretenen analitzar la pròpia tecnologia i veure les possibilitats de desenvolupament futur en un sistema d'informació distribuït i de qualitat de tipus científic. Els resultats obtinguts mostren l'evolució negativa que ha tingut el desenvolupament d'aquesta tecnologia. És a dir, la tecnologia que va obtenir una explosió entre els anys 2000 i 2008, va anar minvant durant els darrers anys 2010-2011, on s'observa que els projectes en curs varen anar tancant majoritàriament per falta de recursos econòmics. Conseqüentment, no s'han pogut analitzar en profunditat casos que mostraven l'evolució d'aquesta tecnologia cap a la web semàntica actual (casos com INTUTE o VASCOSA varen començar a desenvolupar-la).

Pel que fa al quart capítol (Formats i estàndards) es centra bàsicament en una anàlisi descriptiva centrat en formats i estàndards d'estructura de dades (metadades que afecten al nivell sintàctic i semàntic) per a la descripció i recuperació de la informació. A més, es realitza una anàlisi comparativa entre els formats més emprats en l'àmbit especialitzat, concretament en l'àrea de ciències aquàtiques i marines i ciències afins.

Referent al cinquè capítol (Interoperabilitat semàntica en sistemes d'informació de qualitat), es realitza una recerca descriptiva a través de les dades recollides en estudis bibliogràfics realitzats sobre els sistemes d'interoperabilitat semàntica que permeten definir els antecedents, models i tendències actuals sobre aquests tipus de sistemes. Tot seguit, es realitzen estudis de casos sobre projectes europeus i internacionals més representatius, a través dels quals es defineixen les característiques i requisits dels sistemes d'informació amb interoperabilitat semàntica en general. A més a més, es realitza un estudi de casos centrat en l'àmbit de les ciències aquàtiques i marines que permet profunditzar en l'estat actual i tendències futures sobre els sistemes d'informació d'aquest àmbit en relació a la interoperabilitat semàntica.

Un altre instrument d'anàlisi de contingut emprat en el cinquè capítol són les entrevistes. En aquest capítol, a través d'una mostra de biblioteques i centres de documentació en l'àmbit de les ciències aquàtiques, són entrevistats via e-mail, diferents membres de la xarxa europea EURASLIC. I a través de l'anàlisi

quantitatiu i qualitatiu dels resultats de les enquestes, s'avalua l'entorn dels sistemes d'informació de l'àrea d'estudi, centrant-ho en conèixer aspectes claus per a la interoperabilitat semàntica: els llenguatges documentals, l'ús de formats i estàndards, i l'ús de protocols d'intercanvi de dades entre d'altres.

En el sisè capítol, a través dels estudis bibliogràfics previs, es realitzen també anàlisis de contingut (qualitatiu i quantitatiu) sobre treballs relacionats amb l'àrea de la informació de qualitat, els models existents per avaluar la IQ en sistemes d'informació digital, i models d'IQ aplicats en SIQ de tipus passarel·la entre d'altres. Els resultats d'aquestes anàlisis han permès desenvolupar un model propi d'IQ centrat en sistemes d'informació distribuït de tipus passarel·la temàtica.

El capítol setè, realitza l'anàlisi de contingut tant a través de la revisió bibliogràfica com en l'estudi de casos significatius, on s'estudia el funcionament dels vocabularis controlats com a mecanisme de connexió entre diferents models conceptuals d'organització del coneixement, és a dir, estudi de tècniques d'alineament de vocabularis controlats. Aquest capítol permet desenvolupar hipòtesis de treball que es duen a terme en el capítol 8 o estudi experimental.

0.4.3. Recerca experimental

Finalment, el capítol 8, és una recerca experimental que persegueix demostrar el que s'ha exposat al llarg dels capítols: els sistemes d'organització del coneixement o vocabularis controlats són l'eina fonamental per a la indexació de la recuperació de la informació en un sistema d'informació. Al mateix temps, són eines que permeten també l'intercanvi d'informació entre diferents organitzacions i/o sistemes d'informació gràcies a les tecnologies del web semàntic. En el capítol setè es pretenien analitzar diferents casos d'estudi que permetessin corroborar la hipòtesi de que la tecnologia semàntica permet la interoperabilitat entre diferents sistemes gràcies als vocabularis controlats, però el fet de no existir casos d'estudi sobre l'àmbit de les ciències aquàtiques ha fet impossible portar-ho a terme i en el seu lloc hem desenvolupat un estudi experimental a mida en l'àmbit científic objecte d'estudi. En aquest estudi experimental, es realitza el disseny del propi estudi, es du a terme la recollida de dades, i finalment l'execució del mateix. L'estudi experimental pròpiament dit, fa referència a l'alineació semiautomàtica de vocabularis controlats de tipus thesaurus, relacionats en l'àmbit de les ciències aquàtiques i marines, i finalment es realitza l'avaluació dels alineaments per tal d'avaluar al mateix temps la funcionalitat del sistema de mapeig utilitzat per a l'estudi.

0.5. Estructura

L'estructura d'aquesta tesi, s'inicia amb aquest capítol d'introducció on s'exposen els antecedents i el marc conceptual del tema d'estudi, els objectius i el mètode d'investigació seguit.

Tot seguit, la tesi s'estructura en tres parts fonamentals que responen als objectius específics plantejats i inclou vuit capítols:

Una primera part contextual, que engloba els dos primers capítols, on s'exposa l'evolució dels sistemes d'informació distribuïts i de qualitat, i on es du a terme una comparativa entre ambdós models de sistemes d'informació distribuïts i de qualitat: passarel·les temàtiques i sistemes semàntics.

Una segona part, que consta de quatre capítols, on s'aprofundeix sobre diferents aspectes claus que envolten els sistemes d'informació distribuïts i de qualitat tant de tipus passarel·la temàtica com de sistemes d'informació semàntics. El conjunt d'aquests estudis configuren els fonaments per al desenvolupament d'un sistema d'informació distribuït en ciències aquàtiques i marines de qualitat, a través de l'estandardització i la informació qualitat. De fet, és en el capítol sisè on es desenvolupa un model conceptual de sistema d'informació de qualitat per al disseny i avaluació de sistemes d'informació distribuïts especialitzats i de qualitat.

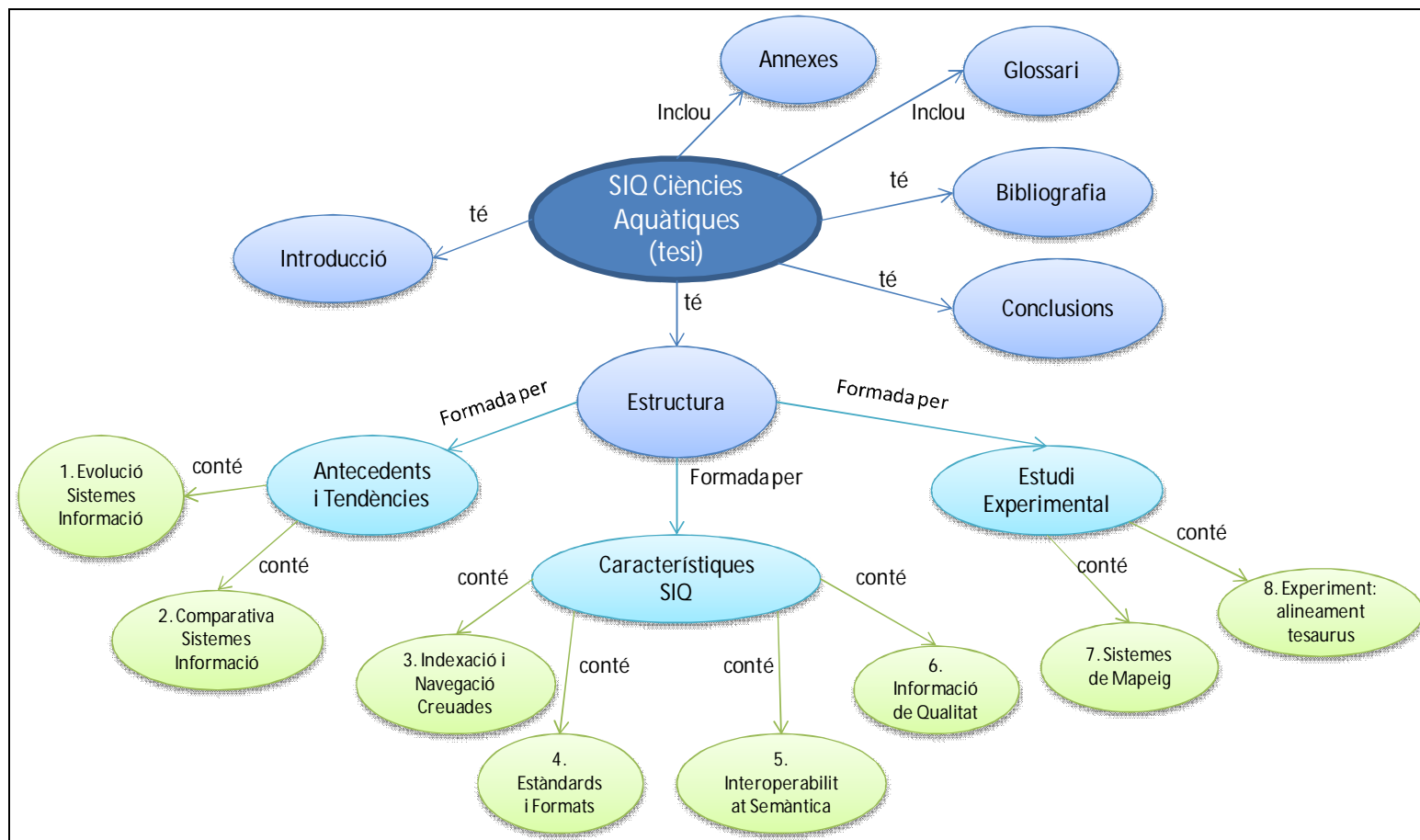
Quant a la tercera part de la tesi, capítol sisè i setè, consta d'un estudi experimental on s'estudia en profunditat les problemàtiques i solucions relacionades amb la interoperabilitat semàntica com a eina fonamental per assegurar l'accés, gestió i recuperació de la informació en un entorn multidisciplinari, multilingüe i heterogeni com són les ciències aquàtiques. Aquest estudi es centra en l'avaluació de la tecnologia d'alineament de vocabularis controlats convertits a ontologies (*linked data*) i inclou una discussió sobre els resultats obtinguts.

El darrer capítol, el formen les conclusions, on es mostren els principals resultats obtinguts al llarg de la tesi.

Finalment, el treball finalitza amb una bibliografia que engloba totes les àrees interdisciplinàries que s'han estudiat al llarg d'aquesta recerca.

Afegir també que s'ha inclòs un capítol de bibliografia general i un capítol d'annexes per cadascun dels capítols amb la seva corresponent numeració.

0.6. Mapa conceptual



CAPÍTOL 1. EVOLUCIÓ DELS SISTEMES D'INFORMACIÓ DE QUALITAT

1.1. Introducció

Aquest capítol mostra l'evolució i l'estat actual dels sistemes d'informació distribuïts i de qualitat de tipus passarel·les temàtiques per mitjà d'una revisió bibliogràfica. S'analitzen les característiques dels prototipus i projectes de sistemes d'informació de qualitat més rellevants així com les seves debilitats i fortaleces per tal de mostrar la seva evolució al llarg del temps i assentar les bases per al desenvolupament d'un model futur de SIQ de tipus distribuït i en un entorn científic. D'una banda, el concepte de sistema d'informació distribuït que es pren de referència a un conjunt de sistemes d'informació independents apareix als seus usuaris com a un únic sistema i servei d'informació coherent i de qualitat. És a dir, a partir d'una única interfície l'usuari navega per col·leccions que estan distribuïdes en diferents sistemes d'informació. D'altra banda, el sistema d'informació de qualitat es determina pel nivell de qualitat que estableix el sistema distribuït. Això esdevé en qüestió als següents aspectes: la selecció i tria de la col·lecció, descripció de recursos amb ús de metadades estàndards, gestió de la col·lecció (actualització, manteniment, etc.), qualitat s'accés a les matèries (indexació i navegació amb vocabularis controlats), ús d'estandards i protocols d'intercanvi d'informació per assegurar la interoperabilitat entre els sistemes i altres serveis de valor afegit (serveis de personalització, d'alertes, etc.). Juntament amb aquestes característiques cal afegir que aquests models de sistemes es regeixen per la cooperació i la col·laboració dels seus participants.

El canvi de paradigma en la gestió de la informació, l'evolució de les xarxes informàtiques i la globalització de la informació han obligat als professionals a adoptar nous mecanismes per gestionar la informació digital i integrar-la en els respectius sistemes d'informació. En aquest sentit, han estat necessari crear polítiques per integrar informació i consolidar aquella disponible amb els sistemes d'informació. Aquestes polítiques fan referència a la integració de la interoperabilitat entre sistemes, per exemple, fer possible tant l'intercanvi d'informació entre dos o més sistemes, i la distribució de recursos garantint l'heterogeneïtat, com la col·laboració entre sistemes, sobretot per definir tasques i processos. Dit d'una altra manera, es refereix a un tipus de sistemes d'informació anomenats sistemes d'informació distribuïts. Un bon exemple de sistema distribuït és la pròpia web, definida com a "xarxa de nodes amb els quals l'usuari pot navegar al seu gust, i on s'accedeix mitjançant una única interfície d'usuari a diversos tipus d'informació: informes, notes, bases de dades, documentació informàtica i ajuda en línia" (W3C, 1990).

Els sistemes distribuïts i les seves eines associades proporcionen solucions a les cinc principals preocupacions de la integració de sistemes d'informació, és a dir: el suport a l'heterogeneïtat, el suport a la interoperabilitat, el suport a aplicacions distribuïdes, el suport als sistemes autònoms, i el suport a l'accés transparent a tota la informació disponible i a la diversitat de recursos existents. Aquestes solucions es produeixen principalment per metadades i sistemes de modelatge que es donen en múltiples bases de dades, oferint el suport requerit

per fer front a la heterogeneïtat de la informació i la interoperabilitat entre sistemes. En particular, el contingut d'una base de dades múltiple (o distribuïda) proporciona mecanismes per descriure l'heterogeneïtat de la informació mentre manté alhora un model d'autonomia de la informació integrada en cadascuna de les bases de dades que el formen.

El propòsit d'aquesta secció és conèixer els diferents sistemes d'informació existents i aplicables en un context especialitzat o científic, i, aprofundir en aquells models que s'identifiquen a internet com a referents de sistemes ecològics (Capurro, 1990; Davenport i Prusac 1997; Nardi i O'Day 1999; Benkler, 2001, 2003 i 2006; Sebastià-Salat, M. 2000 i 2008), i de recuperació d'informació de qualitat. . Les principals característiques d'un sistema d'informació de qualitat i ecològic són: la seva arquitectura de la informació, basada amb els processos de navegació creuada (*cross browsing*), la indexació creuada (*cross indexing*) i la cerca creuada (*cross searching*). Aquests trets són possibles gràcies al treball en xarxa entre diferents sistemes i/o organismes, amb l'objectiu que l'usuari pugui captar informació diversa de qualitat i de varis recursos amb major accés i ús (diferents punts d'accés d'informació dins del mateix sistema).

Abans de passar a l'estudi de l'evolució dels sistemes d'informació de qualitat, es presenten les característiques principals que engloben els sistemes d'informació de qualitat.

1.2. Ecologia de la informació

L'ecologia de la informació (IE) és la base dels sistemes d'informació de qualitat de tipus cooperatiu i col·laboratiu. La IE és l'estudi de sistemes d'informació complexos, dinàmics i l'estudi de la interacció entre els diferents agents que el conformen, així com els espais i processos d'informació que els configuren. L'ecologia de la informació i estudis relacionats amb l'ecosistema de la informació pretenen oferir una millor comprensió de la interacció entre els usuaris, els espais digitals, els serveis d'informació i els entorns en què es produeixen. El concepte d'ecologia de la informació va sorgir de l'àmbit de la gestió de la informació i dels estudis del comportament informacional.

El concepte d'ecologia de la informació que engloba aquest estudi es basa en les idees dels autors de referència, els quals insisteixen en què tots els agents i processos estan involucrats en l'ecosistema que envolta la informació. És a dir, els models d'informació proposats en ecologia de la informació estudien les complexes relacions entre els éssers humans i la tecnologia en termes d'ús de la informació dins les comunitats i organitzacions. Els primers autors són de la dècada de 1990, però en el decurs del temps han anat sorgint noves propostes que també recollim en el quadre següent.

Autors	Models i propostes d'IE
Capurro (1990)	La IE és vista des de dues perspectives principals: les dimensions de la informació (social, lingüística i històrica) i el concepte de contaminació de la informació (s'entén com la bretxa que hi ha entre la informació rica i la informació pobra)
Davenport & Prusak (1997)	Ecologia de la informació des del punt de vista de com fer que la informació sigui significativa. Aquests autors exposen com la tecnologia per si sola no és suficient per tal d'assegurar l'èxit dels productes d'informació, i com les persones o gestors de la informació juguen un paper fonamental.
Nardi & O'Day (1999)	La IE es basa en la relació entre la tecnologia de la informació i la gent, és a dir, la transformació de la informació en coneixement. L'ecologia de la informació és concebuda com el lloc on la gent utilitza les eines i les relacions socials per ajudar en les diferents activitats i processos on es produeix la informació. Amb la IE es cobreixen els procediments, els objectius i els valors de les comunitats amb el suport de la tecnologia.
Sebastià-Salat (2000, 2008)	El concepte d'ecologia s'ha utilitzat també en el camp de la informació per indicar les qüestions relacionades amb la responsabilitat dels especialistes de les ciències de la informació en els processos d'innovació. Sensibilitat que configura una nova ideologia professional i un nou paradigma en el món de la informació digital.
Benkler (2001, 2003, and 2006)	Concepció basada en l'ecologia social comuna i institucional de l'entorn digital (economia de la informació en xarxa). Aquest model proposa la construcció d'infraestructura com un requisit previ per la transició des d'una societat de consumidors d'informació passius cap a uns consumidors d'informació actius. Es parla de desenvolupar una societat on tothom pugui parlar amb tothom i convertir-se en un participant actiu en el discurs polític, social i cultural
Malhotra (2002)	La proposta de " <i>knowledge ecology</i> " tracta la creació de coneixement com un procés dinàmic i evolutiu, en el qual el coneixement es crea i es reutilitza en diferents contextos i en diferents punts en el temps. L'autor proposa el model de " <i>loose-tight systems</i> ", sistemes que fomenten l'aprenentatge i el desaprenentatge simultani per garantir l'optimització de l'eficiència dels processos de recerca. Els estudis d'aquest autor posen l'èmfasi en canvis de l'ecologia de la informació i la necessitat de dissenyar sistemes flexibles.
Williamson (2005)	Model ecològic de cerca i ús de la informació. Es basa en un model on hi ha un actor social involucrat en les necessitats d'informació personals, físiques i en contextos socials i de treball.
Nahl (2007)	Desenvolupament de la base de la psicologia ambiental d'un model constructivista que estudia l'ecologia de la informació des del comportament dels usuaris. És a dir, és un model d'informació ecològic on s'integren les parts cognitives i

	afectives de les activitats d'informació.
Nahl & Bilal (2007)	Aquests autors relacionen també l'ecologia de la informació amb els estudis de les emocions afectives i, com aquestes afecten a la conducta de l'ús de la informació i les relacions amb les tecnologies de la informació. La cerca i ús de la informació estan dirigides per un filtre afectiu que posa l'usuari.
García-Marco (2011)	L'evolució de les biblioteques digitals i dels serveis d'informació utilitzant el concepte d'ecologia de la informació. L'autor identifica les tendències més importants en l'entorn digital i el seu efecte sobre la funció social i econòmica dels sistemes d'informació.

Taula 1. Models i propostes d'ecologia de la informació

El resultat final de tots aquests enfocaments mostra que les diferents estratègies emprades en la societat de la informació per a la transferència de coneixements són molt complexes, amb molts subsistemes i lògiques que competeixen, col·laboren, i molts cops són convergents i/o divergents. Com passa en la web actual, s'està representant un univers de sistemes d'informació amb diferents models que, per tal de poder conviure i ser eficaços, calen esforços per part de tots els agents implicats (professionals, marcs institucionals, investigadors, xarxes socials, etc.).

L'objectiu principal dels models d'ecologia de la informació exposats és la transferència de coneixement a les persones i/o usuaris d'informació. Es fa èmfasi en la importància de reduir les diferències entre les persones i els sistemes d'informació, es podria comparar amb la filosofia de les xarxes socials (compartir informació i fer que aquesta sigui accessible a tothom).

L'estudi del flux de la informació, també anomenat logística de la informació (Grolík, 2007), fa referència als processos que envolten un sistema d'informació, des de la demanda d'informació de l'usuari fins a la relació entre els fluxos d'informació i les fonts d'informació que satisfan aquestes necessitats informatives de l'usuari. La logística de la informació és una àrea que sorgeix dins l'estudi dels fluxos d'informació de les empreses i companyies generalment de telecomunicacions, on el principal problema que hi havia era la "sobrecàrrega d'informació". La logística de la informació identifica quatre paràmetres essencials:

1. **Informació ambigua.** La major part de la informació que té un flux diari és informal. Per tant, la informació és susceptible de diverses interpretacions i es converteix en ambigua. A més, la informació no està organitzada, no hi ha protocols o manuals que la gestionen i molts cops no té significat o els objectius d'aquesta són ambigus, i, el receptor es frustra davant l'accés a aquesta informació. Cal recordar que la informació s'actualitzarà amb el temps a causa dels canvis en les circumstàncies estructurals, tecnològiques i ambientals d'una

organització. Un exemple paradigmàtic d'aquest paràmetre és la comunicació diària amb els sistemes de correu electrònic, que constantment augmenta el nivell d'ambigüitat com a resultat de la informació inesperada o no desitjada, com és el cas del correu brossa.

2. **Informació excessiva.** Les organitzacions estan constantment buscant solucions que permetin millorar la transmissió d'informació. Encara que les organitzacions apliquen les tecnologies de la informació i la comunicació per gestionar la informació i els sistemes de comunicació, aquests tendeixen a ser funcionals, l'usuari encara té accés a molta informació que no pot dirigir fàcilment, i, fins i tot, molts cops es tendeix a descartar o ignorar informació correcta a causa de l'excés d'informació rebuda. Un exemple és la duplicitat d'informació, aquest seria el cas de dos departaments d'una mateixa organització que publiquen o distribueixen informació de diferent contingut però, amb el mateix tema, o el mateix títol i on el receptor o usuari elimina la informació rebuda sense llegir-ne el contingut.
3. **Informació descontextualitzada.** El flux de la informació és un procés complex i sovint no és capaç d'arribar a la persona adequada en el moment oportú, amb la finalitat i el contingut correctes. A més a més, les empreses i les organitzacions tenen objectius i contextos diferents. Una mateixa informació pot ser escrita de moltes formes, un mateix contingut expressat en un context diferent pot donar lloc a informacions incomprensibles, és a dir, esdevé informació fora de context o descontextualitzada. Alguns exemples es troben en els encapçalaments dels *e-mail* o els titulars de notícies, on el contingut no es correspon amb la seva descripció i no compleix amb les expectatives de l'usuari final. Molts cops els títols semblen interessants, però el contingut real és molt diferent a les expectatives dels usuaris. En aquests casos, el títol és el que crida l'atenció de l'usuari, però el contingut no és coherent amb el context.
4. **Informació poc fiable.** L'entorn de la informació, sobretot en entorns digitals, no ha generat una forma correcta de comunicar informació fiable ni per a les organitzacions ni tan sols per a un sol individu. D'una banda, amb l'aparició d'Internet i la seva ràpida evolució, cada cop més apareixen múltiples vies i mitjans de comunicació de la informació que fa encara més difícil trobar informació veraç. D'altra banda, amb la comunicació quotidiana, la informal, tampoc resulta ser un mitjà suficient per proporcionar informació fiable amb el contingut adequat per a la persona adequada. Un exemple paradigmàtic és una organització que promou un nou servei, i no disposa de suficient informació. Per exemple, una empresa envia un correu electrònic a diverses empreses que utilitzen els seus serveis, però no sap si la persona de contacte té dret o

no a rebre aquesta informació. Un altre exemple és quan una empresa vol promoure un servei i envia la informació a una empresa que ja ha adquirit aquest servei i ja no el necessita, i tants d'altres casos tipificats pels estudis. En aquests casos, l'estratègia de l'empresa en termes de flux d'informació, en general, no s'ha basat en un pla de distribució d'informació adequada, i per això ha fracassat pel que fa a l'execució, la planificació i el control. Aquesta forma de distribució de la informació mostra un entorn típic de canvi, on el desbordament de la informació domina el procés de comunicació, principalment a causa del caos creat pels múltiples tipus de tecnologies de la comunicació i la manca de consciència de les necessitats d'informació dels usuaris finals. Aquest cas exposat és un exemple d'un model de comunicació caòtic amb excés d'informació.

El comportament dels usuaris en l'accés a la informació i la interacció amb els sistemes és fonamental per al disseny d'un sistema d'informació de qualitat.

Conseqüentment, el model de flux d'informació d'un sistema ecològic passa per la convergència en la cerca i recuperació de la informació, i per compartir i aprendre per millorar l'accés a la informació i el coneixement. L'ecologia de la informació es basa, doncs, en les relacions entre les tecnologies de la informació i les persones implicades en la transformació de la informació obtinguda en coneixement. És per això que l'ecologia de la informació constitueix un model híbrid (persona-màquina).

L'ecologia de la informació ha centrat els seus objectius en la gestió i resposta a les interaccions d'un ecosistema d'informació com a un problema complex i com a un procés clau en la relació entre tots els diferents factors que intervenen i que es presenten a continuació:

- **Components dels sistemes d'informació ecològics:** són sistemes adaptatius complexos i compostos formats per molts components heterogenis que interactuen entre sí (per exemple: sistemes i ecosistemes diversos). Per tant, la seva gestió requereix d'una capacitat i flexibilitat que respongui a la retroalimentació de la informació digital que canvia constantment. Els sistemes d'informació ecològics són extremadament dinàmics i els experts destaquen la necessitat d'un aprenentatge continu i un treball de col·laboració mitjançant els professionals per tal de comprendre l'entorn informacional corresponent. El desenvolupament del coneixement i la comprensió dels sistemes, per fer front al canvi i la incertesa que hi ha en aquests sistemes (García-Marco, 2011).
- **Sistema dinàmic:** dins d'un model d'ecosistema digital s'ha d'entendre que aquesta és una visió d'un sistema que està en canvi constant dins d'un model d'ecosistema digital. Per tant, un dels punts forts d'utilitzar un

enfocament basat en l'ecologia és que s'intueix que l'ecosistema és naturalment dinàmic. De forma implícita hi ha la idea que els components del sistema canvien i s'adapten al seu entorn: la comunitat de domini, el paradigma del coneixement, les tecnologies, els mediadors, les necessitats d'informació, els usuaris i els processos de recerca d'informació, i entorns d'informació digital. Aquesta adaptació no es pot produir a una velocitat constant, en tot ecosistema de la informació hi ha components que canvien ràpidament, com per exemple, les tecnologies de la informació. Aquesta adaptació es pot considerar no només com una evolució, sinó també com un inesperada «mutació» provocada per la societat de la informació i/o els factors econòmics, com ara, les xarxes socials i els sistemes distribuïts d'informació. Hi ha molts exemples basats en els canvis ràpids que experimenten els ecosistemes d'informació, però un dels canvis més significatius ha estat l'adaptació de les metadades com a eines de suport a la qualitat en la indexació de la informació. Un altre exemple d'una espècie o component relacionat amb l'evolució d'un sistema d'informació, és la transformació de la web cap a el web semàntic i l'evolució dels seus estàndards, incloent, entre d'altres, RDF, RDF-S, OWL, SPARQL, i molts d'altres.

- **Què passa en la dinàmica de l'ecosistema a la informació?** Els estudis assenyalen dos aspectes importants, en primer lloc, els autors de la gestió del coneixement afirmen que com les dades es mouen per la cadena d'informació, el valor de la informació augmenta a mesura que disminueix el seu volum (per sobre de tot, el disseny d'aquests ecosistemes requereix l'adquisició, processament, interpretació / indexació, síntesi i serveis d'informació), aquestes accions requereixen de temps del personal i per tant una despesa considerable (valors d'informació). D'altra banda, la informació costa d'obtenir, i els prenedors de decisions tenen una quantitat limitada de temps, personal i diners. Per tant, és important determinar per endavant quin tipus d'informació és "decisiva", és a dir, quina és necessària, justa i suficient per satisfer les necessitats d'informació. Aquest concepte d'informació decisiva (així com el concepte) i el de "pertinença" redueixen els costos dels ecosistemes de la informació (pertinença i sostenibilitat).
- La **sostenibilitat** és una qüestió clau per als sistemes d'informació d'aquest segle XXI com ho reconeixen cada vegada més quan les polítiques i pràctiques tenen conseqüències socials (info-coneixement). En conseqüència, molts sistemes implementen elements de sostenibilitat en les seves pràctiques de gestió. En sintonia amb aquesta tendència trobem que el creixent reconeixement dels subjectes relacionats amb la sostenibilitat ha de ser inclòs en els projectes digitals d'informació,

sobretot en el disseny de nous models de sistemes d'informació de recuperació.

- **Escalabilitat / Granularitat:** la granularitat es refereix al grau de detall i precisió continguda en un model o sistema d'informació. Quan estem modelant un sistema d'informació des d'una sola perspectiva, la granularitat no és molt important, però, això canvia totalment quan s'està tractant de combinar la informació amb els sistemes distribuïts o sistemes múltiples, perquè el més probable és que la granularitat no coincideixi (tot depèn de la flexibilitat del model de dades que s'hagi dissenyat). En un sistema dinàmic, i, des d'una perspectiva ecològica, les interaccions i els processos ocorren a escales particulars (granularitat). Els processos del sistema d'informació afecten a diferents nivells, i, per tant, es requereix una acurada separació dels diversos processos per comprendre com i per què funcionen o fracassen els sistemes d'informació. En aquest estudi (dins el capítol Interoperabilitat semàntica), s'exposa que un dels mètodes per aconseguir aquesta granularitat entre sistemes distribuïts és unir els processos del "cicle de vida de la informació" amb els diferents nivells d'interoperabilitat semàntica que existeixen en un sistema (Reverté i Sebastià, 2009).
- **Les entitats i els factors externs.** Un sistema, des d'una perspectiva ecològica, està format per entitats que participen en les interaccions, els fluxos d'informació i en els factors ambientals que engloben el sistema. Exemples d'aquestes interaccions són les activitats entre els usuaris, els serveis, el programari, la personalització, les pràctiques de l'organització i el contingut (els registres i les metadades que els descriuen). En aquest sentit, l'objectiu dels ecosistemes de la informació és centrar-se en entitats específiques i les seves interaccions, que són les combinacions d'entitats i espècies (és a dir, l'usuari expert, dins de l'espècie 'usuaris'). També pot ser útil fer referència a una interacció amb usuaris particulars, com ara, un investigador, un professor o un estudiant (és a dir, una entitat específica que alhora és una instància de l'espècie i una destinació de la informació). Les interaccions, poden indicar simplement que s'està produint un flux d'informació i poden incloure, per exemple, el programari, els correus electrònics, *feeds* (comunicació via RSS), converses, i d'altres. O bé, també es poden referir a cada usuari i entendre els recursos específics implicats en les interaccions. Per exemple, hi ha nombroses interaccions entre estudiants de doctorat i el seu tutor: bases de dades, projectes, esborranys, notes, conferències, debats, correus electrònics, prototips, taules, figures, cronologia, bibliografia, recursos web, mètodes, estudis de casos, l'exploració, els grups d'enfocament, estadístiques i moltes més. En general, els efectes d'aquestes interaccions en l'ecosistema no són causades per altres

entitats o espècies, sinó pel context i les forces externes, com ara l'efecte de la societat de la informació i les tecnologies específiques presents en un ecosistema determinat. Per tant, un factor ambiental és un element que afecta a un ecosistema, organització o comunitat, de manera que és més general que la interacció entre les entitats que el constitueixen. Per exemple, el mandat del consell d'un programa de doctorat és probable que afecti el nombre de tesis disponibles en cada tema en una comunitat particular i de tot l'ecosistema que interactua amb la matèria (bases de dades, repositoris digitals, revistes, accés obert).

En considerar els factors ambientals en el disseny i gestió d'un ecosistema d'informació, és important recordar que els factors ambientals poden sorgir dels participants actius en altres entorns (diferents tipus d'interaccions), el context legal, el context lingüístic i l'entorn cultural. Cal assenyalar, doncs, la importància de l'aplicació d'aquests factors externs en el disseny dels ecosistemes. En tot disseny és fonamental conèixer dos aspectes importants: 1) la influència sobre els ecosistemes no ha de ser modelada com una interacció que afecta a totes les entitats, ja que la interacció és sempre un enllaç directe que connecta dues entitats com a mínim, 2) les tendències web i la influència d'escalabilitat que pot afectar tant a les entitats com també a les interaccions o l'aplicació d'estàndards com l'RDF generat per la *Dublin Core Metadata Initiative* (DCMI).

La metàfora de conceptes ecològics és útil per definir el flux d'informació i essencial quan volem articular les interaccions produïdes en els sistemes o serveis d'informació. Quan es tracta de sistemes digitals, la gent, els usuaris, i les organitzacions com entitats en el mateix entorn, de manera selectiva poden articular les interaccions bàsiques, les dependències i les influències que es produeixen. Ja sigui com a usuaris, administradors o desenvolupadors el que ens permet comunicar i gestionar les dependències és clau dins del sistema, i proporciona una forma alternativa d'anàlisi i desenvolupament dels sistemes i serveis de qualitat.

No obstant això, el concepte d'ecologia de la informació exposat en aquest capítol no només inclou el concepte des del punt de vista del comportament d'informació mitjançant diferents actors socials implicats, sinó que també incorpora les eines que poden ajudar a netejar o filtrar l'entorn informacional per tal de convertir la informació en coneixement. És a dir, com l'ecologia de la informació pot contribuir per crear sistemes d'informació de qualitat. En aquest sentit, les diferents característiques o factors que defineixen a aquests tipus de sistemes són:

1. Sistemes complexos basats en xarxes d'informació:
 - Els sistemes complexos utilitzen sistemes i serveis intel·ligents basats en la interoperabilitat semàntica per tal de reduir el "soroll

documental” o la sobrecàrrega d'informació, com és el cas de les ontologies.

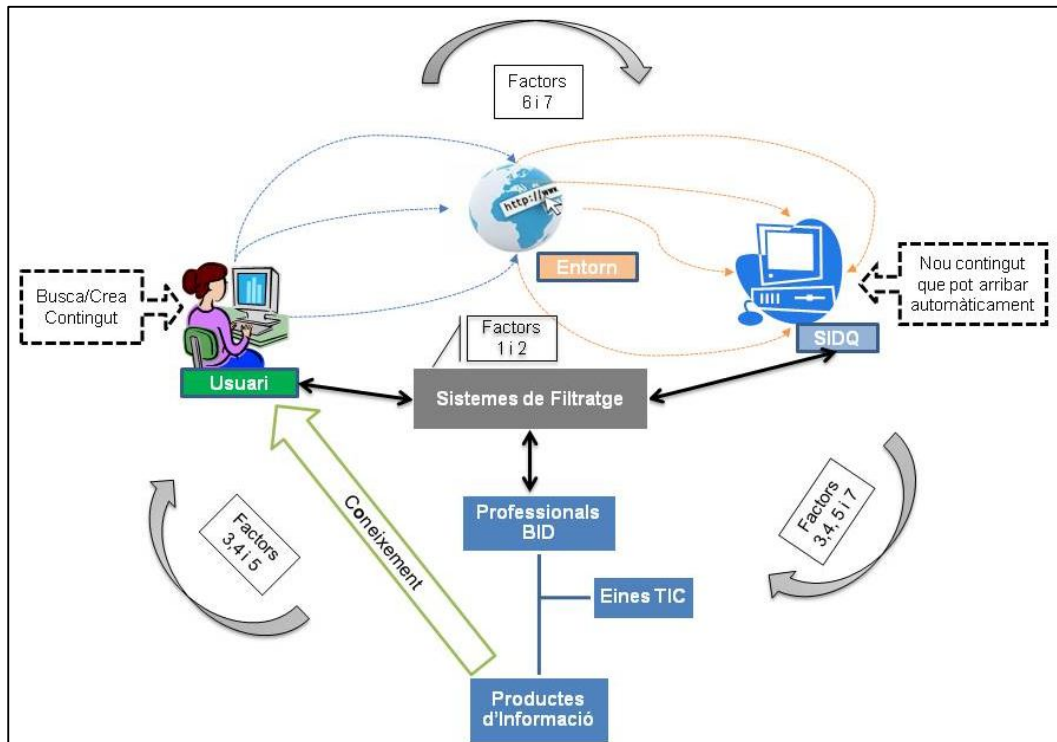
- Descentralització de la informació: fomenten la interacció i intercanvi d'informació

2. Rellevància. És una de les característiques més importants d'aquests sistemes i aporta beneficis, ja que en l'aplicació de l'ecologia de la informació i l'ús d'eines tecnològiques es pot ajudar a la presa de decisions. Per exemple: ús de mapes conceptuals, ontologies, *topic maps*, la construcció d'estils personalitzats de la informació en les interfícies d'usuari, la construcció d'eines per al descobriment, la construcció de l'avaluació del significat per rellevància, integració d'estils d'informació amb les estructures de dades (recursos descobriment, e-memòria) i funcions d'informació digital diverses.
3. Organització del coneixement basada en el comportament de la informació. Els vincles entre el coneixement intern i extern, les emocions i les xarxes socials. Suport al comportament de la rellevància a través de les característiques visuals dels sistemes i noves eines d'organització del coneixement que formen part de la nova filosofia de productes d'informació “ecològics”. Com per exemple, els núvols de *tags* (etiquetes socials) ajuden als usuaris a la presa de decisions alhora de seleccionar recursos per la seva rellevància. O bé, reutilitzacions i *mash-ups* poden ser exemples d'una nova direcció per al desenvolupament de productes d'informació.
4. Nova filosofia de fonts d'informació i serveis d'informació: Integració de noves estructures d'informació en serveis digitals i interfícies. Els serveis d'informació i productes basats en part de l'entorn global de la informació com per exemple les xarxes socials que permeten: navegar, interactuar, enllaçar, relacionar, visualitzar o personalitzar la informació i crear serveis d'alerta entre d'altres.
5. La reutilització d'objectes, serveis, i productes. Ús d'informació ecològica: eines per a la neteja o filtratge d'informació (vocabularis controlats: ontologies, taxonomies, mapes conceptuals, etc.)
6. Integració de les necessitats dels usuaris en funcions (satisfacció dels usuaris i optimització de recursos). És a dir, permetre la integració dels usuaris en els sistemes. Les xarxes i les eines socials són un exemple, perquè permeten llegir, escoltar, crear, escriure o cercar informació. Des d'aquesta visió les necessitats dels usuaris són més fàcils d'atendre amb la seva participació.
7. Continguts generats pels usuaris com poden ser aspectes de creativitat: classificació, etiquetatge, comentaris, anotacions i discussions entre d'altres (exemples d'aplicació: xarxes socials, blogs, wiki's, etc.). Al mateix temps aquestes funcionalitats redueixen esforços. Un cas exemplar és l'etiquetatge social.

Tots aquests factors representats mostren una interacció entre els diferents fluxos d'informació i els agents que intervenen (informació-usuaris-sistemes informació, professionals-informació-sistemes i informació-professionals-tractament informació-sistemes-usuaris)

Figura 1. Representació dels factors IE en els Sistemes d'Informació de Qualitat (Font: Elaboració pròpia)

En conclusió, l'estratègia fonamental d'una organització implica saber com planificar correctament, executar i controlar els seus processos. El flux d'informació és una part d'aquesta estratègia, de la mateixa manera que els



propis consumidors i creadors d'informació. No obstant això, com a resultat dels canvis constants en la tecnologia i la quantitat de tecnologia utilitzada en els entorns digitals, és evident que molts cops no s'aplica una estratègia adequada per a cada flux d'informació corresponent, perquè sovint aquests cops són molts i diversos, també canviants i difícilment controlables.

Des del nostre punt de vista, el concepte de l'ecologia de la informació pot ajudar a respondre als desafiaments tècnics i culturals del comportament en front de la informació i el seu ús. Però, també caldria tenir present que el treball conjunt i la integració del coneixement de tots els agents que envolten un entorn digital és l'eina principal per tal d'aconseguir l'efectivitat de tot projecte que implementa un sistema d'informació de qualitat.

Existeixen diferents sistemes d'informació, cadascun amb diferents fluxos d'informació i models de comunicació: models col·laboratius, centralitzats, distribuïts i piramidals (centrals).

Els models que més s'adapten als sistemes d'informació complexos són els models distribuïts (apartat 3) tot i que, exceptuant el model piramidal, en la resta de models de comunicació existeix o es pot establir algun tipus de comunicació o sistema col·laboratiu.

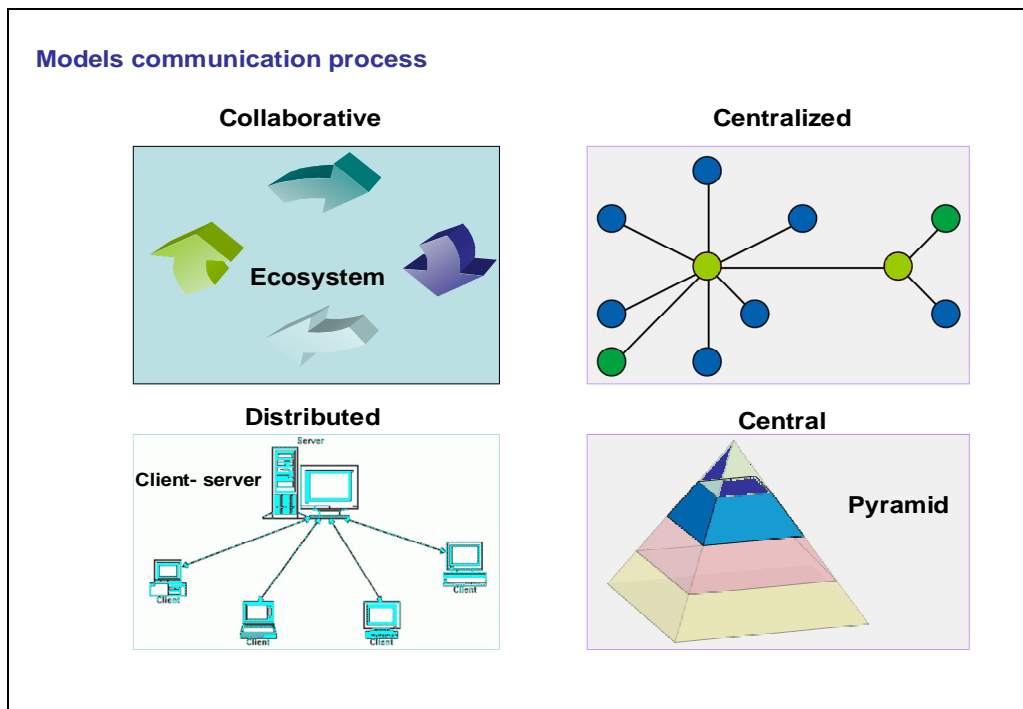


Figura 2. Flux d'informació i models de comunicació en els sistemes d'informació (Font: Sebastià-Saltat, M.; Reverté, C., 2011)

El model piramidal apareix als anys vuitanta (Laudon, KC i Laudon, JP, 1988). A diferència de la resta de models, aquest reflecteix la jerarquia dels diferents sistemes d'informació que coexisteixen en una mateixa organització (sistemes de processament de transaccions a la base de la piràmide, seguits per sistemes de gestió de la informació, seguits per sistemes de suport a la presa de decisions i acabant amb els sistemes d'informació executius en la part superior). Encara que el model de la piràmide segueix vigent i continua sent útil han anat apareixent altres sistemes i models d'informació que porten grans avantatges.

La resta de models exposats (figura 1) són dissenyats i/o s'adapten a altres models de sistemes d'informació on estan implicats diferents organitzacions amb un objectiu comú.

Quant al model de gestió de la informació centralitzada, l'avantatge principal és que maximitzar l'ús eficient dels recursos, la informació i els serveis són compatibles en quant a la gestió i presentació de la informació, els riscos són més baixos, el control de qualitat és eficient, i, tot amb un cost relativament baix. A més, és un model que requereix d'un suport per al seu desenvolupament. Per contra, el desavantatge principal d'aquest model és que tendeix a ser conservador, la qual cosa pot crear un coll d'ampolla en el flux d'informació, a més a més de ser un model variable, canviant i divers, per dependre molts cops de la resta de sistemes associats que l'integren. Les limitacions són percebudes com una amenaça per a la llibertat dels usuaris i, finalment, no sempre apliquen una bona escalabilitat.

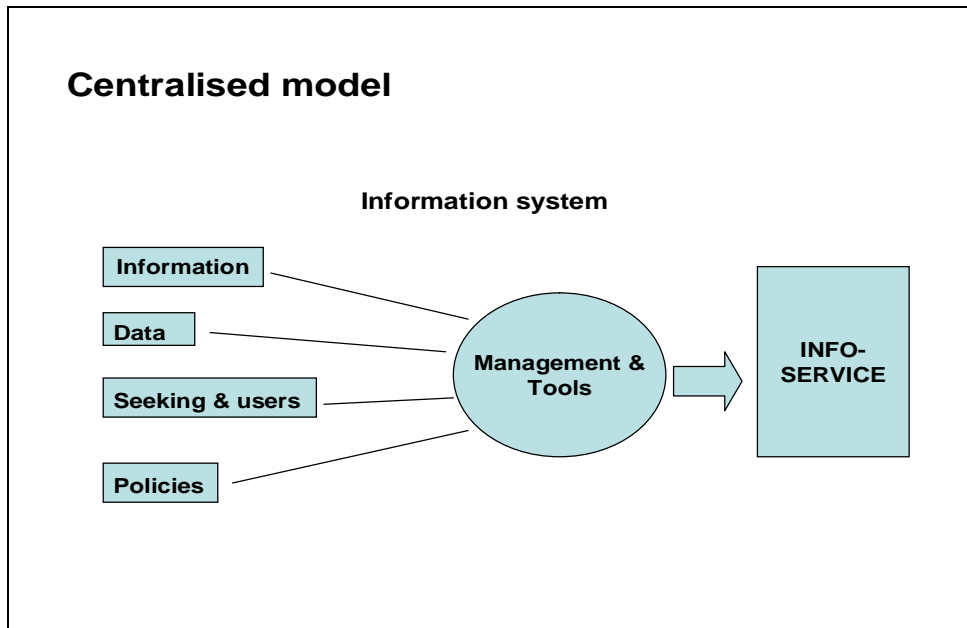


Figura 3. Model de sistema d'informació centralitzat: Components i flux d'informació (Font: Sebastià-Saltat, M.; Reverté, C., 2011)

És un model molt més sensible i flexible que el distribuït i els seus avantatges principals són: la càrrega de treball, assegurar la millora de la “consciència” del treball des de cadascuna de les organitzacions implicades en la gestió de la informació, i gestionar el fluxos d'informació a diferents nivells (departaments o grups “de control”). No obstant això, el model de gestió distribuïda també mostra diverses debilitats. Per exemple, es necessita un grup gestor/responsable per controlar el flux d'informació i una política d'informació adequada per evitar que es converteixi en un model de "no gestió".

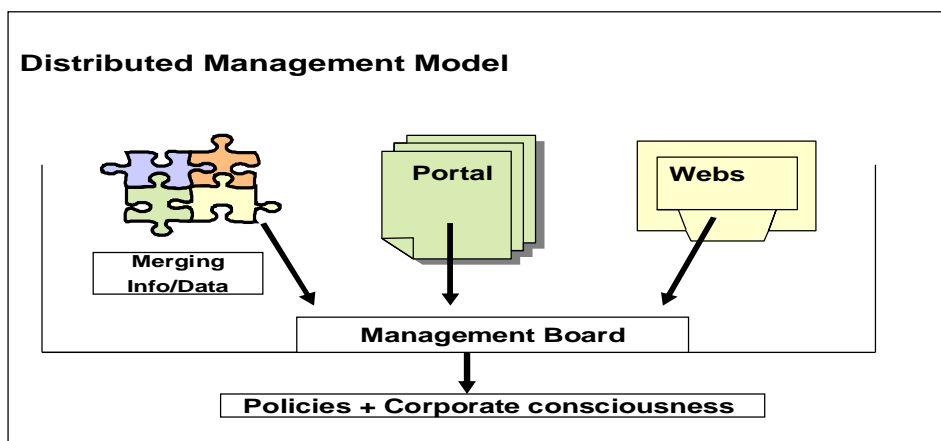


Figura 4. Model de sistema d'informació distribuït (Font: Sebastià-Salat, M.; Reverté, C., 2011)

El procés de construcció d'un model de flux d'informació està subjecte a la influència de factors molt variables i de vegades impredecibles. Els factors més importants són els documents i el seu cicle de vida, el model d'informació descriptiva (el conjunt de propietats que descriu i identifica el document),

l'organització de la informació, les eines de gestió, i finalment els diferents actors i socis amb les seves diferents funcions.

Un dels punts crítics que els models de gestió del flux d'informació han de resoldre és la gestió de la informació de documents no estructurats que plantegen uns desavantatges fonamentals: la reutilització del contingut és pobre, hi ha dificultats en la recerca i recuperació d'aquest tipus d'informació, hi ha un baix nivell d'interoperabilitat entre els sistemes d'informació i la informació difícilment pot relacionar-se amb un context determinat (per exemple: degut a la falta d'estàndards).

El flux d'informació depèn, gairebé per definició, del maneig i l'ús de la informació. Així doncs, aprofitar el contingut disponible en els processos d'organització de la informació en grans sistemes d'informació s'ha identificat com una oportunitat important per al disseny del propi sistema. El flux de la informació es pot considerar com un mètode estratègic en els processos de capturar, compartir i reutilitzar el coneixement. No obstant això, els enfocaments actuals cap a l'accés al contingut i l'ús encara revelen limitacions importants (costos elevats, tecnologia i estàndards canviants, etc.). A més, molts dels projectes no han pogut tenir continuïtat per la falta de recursos principalment econòmics, malgrat les enormes inversions que s'han fet en els darrers anys en eines i sistemes per a la gestió de la informació digital, com es veurà en el darrer apartat d'aquest capítol (apartat 4). Les principals barreres que poden sorgir, en termes de flux d'informació, inclouen l'ús de la tecnologia inadequada, manca de coherència en el disseny d'informació, manca d'habilitats en TIC, actituds negatives com la territorialitat i la tecnofòbia, la resistència al canvi, les eines de gestió inadequades, la manca de recursos per al tractament de la informació i, el més important de tots, els límits de la pròpia informació (per exemple: la "no disponibilitat" de tota la informació). En poques paraules, la força de qualsevol model d'informació de gestió del flux rau en el disseny de les estratègies d'informació i l'establiment de la responsabilitat política de la informació i el seu moviment.

Des del punt de vista ecològic i amb l'enfocament en la gestió, tots els models i mètodes esmentats estan d'acord a subratllar que el flux d'informació és necessari per facilitar el flux de continguts digitals (informació, dades i documents). Aquest nou paradigma ha generat l'aparició de cinc tipus bàsics d'aplicacions en la gestió del flux d'informació: anàlisi de contingut (classificació, extracció d'informació, cartografia, significat), rutes (creació de rutes del sistema de documents rellevants per a les persones); alerta (sistemes d'alerta on es generin esdeveniments o informes enviats a les persones interessades), la mineria de dades (els usuaris poden explorar o analitzar col·leccions o fluxos) i la recuperació de la informació (ajudar als usuaris a trobar i entendre els documents pertinents).

Davenport i Prusak (1997) van suggerir ja fa molts anys que el flux d'informació útil depèn de les persones, no de la tecnologia o dels sistemes. No obstant això, en aquest estudi es proposa que el flux eficient de la informació depèn de la convergència de quatre paràmetres: els usuaris, més la gestió de la informació de qualitat, més els límits de la pròpia informació, més les eines d'innovació. Per articular la complexitat d'aquests quatre paràmetres, l'arquitectura de sistemes d'informació ha de ser canviant, flexible i adaptable, i les relacions i interaccions dels diversos components s'ha d'implementar en el seu propi entorn. L'entorn o el medi ambient que envolta els usuaris o organitzacions és la clau en el disseny i gestió de sistemes d'informació. Per tal de demostrar l'eficàcia d'aquesta hipòtesi, en els següents capítols s'analitza la gestió de la informació des de la perspectiva de l'ecologia de la informació i l'aplicació de la metàfora en la gestió de la informació, és a dir, els ecosistemes d'informació i/o ecosistemes digitals que envolten els sistemes d'informació especialitzats, i concretament, sempre que ha estat possible en l'àmbit de les ciències aquàtiques.

1.3. Sistemes distribuïts en un marc especialitzat

Els sistemes d'informació distribuïts són considerats el model de sistema més adequat en l'àmbit de sistemes d'informació especialitzats, així doncs es consideren objecte d'estudi i avaluació per contribuir en el disseny del prototipus de sistema d'informació digital de qualitat especialitzat en ciències aquàtiques i marines. D'una banda, són models de sistemes generalitzats i d'altra banda, el seu model d'ecologia de la informació i/o de comunicació engloba tots els actors i agents implicats en un sistema d'informació (usuaris, entorn, cicle de vida dels documents, gestió i organització de la informació, com a factor clau, etc.).

La majoria dels sistemes distribuïts són coneguts com a portals temàtics o passarel·les temàtiques. Al llarg de la història han estat anomenades de diverses formes, *subject-based information gateways (SBIGs)*, *subject based gateways*, *subject index gateways*, *virtual libraries*, *clearinghouse*, *subject trees*, *pathfinders*, *quality controlled subject gateways*. Concretament al llarg d'aquest estudi són concebudes com a sistemes distribuïts d'informació de qualitat. Tots els diferents tipus de sistemes d'informació distribuïda segueixen procediments similars pel que fa a tècniques de catalogació, indexació i classificació dels seus continguts, així com el seu control de qualitat de la informació, és a dir, les polítiques de gestió de la informació o de selecció de recursos.

Un objectiu específic d'aquest estudi és fer un estudi exploratori sobre la seva arquitectura de la informació (indexació i navegació creuades), i en general conèixer els seus avantatges i desavantatges i el seu ús. Aquest estudi bibliogràfic pretén estudiar també, on se centra la investigació sobre aquest

camp i veure les possibilitats que hi ha per al desenvolupament de futurs sistemes distribuïts, i així poder crear un marc conceptual per a qualsevol tipus de sistema d'informació distribuït digital i de qualitat.

En aquest capítol no s'ha posat cap limitació quant a l'abast cronològic o geogràfic. Es tracta d'analitzar la major part dels projectes d'impacte duts a terme fins l'actualitat. No obstant això, alguns d'aquests s'han detingut per falta de fons (crisi econòmica i financera), i en alguns casos sols s'ha pogut estudiar la fase de prototip. Alguns dels projectes estudiats són les “subject gateways” americanes dels EUA (The Gateway: To 21st Century Skills; and site of scientific USA: U.S. National Cancer Institute), les del Regne Unit (RDN: SOSIG, PSIGate, Humbul, Altis, Artifact, EEVL, Gesource biomass), Europa (Renardus i Desire), Alemanya (el projecte CARMEN i la passarel·la Vascoda), les australianes (Ariadne, AustLit, AustLII, etc.), el catàleg de la Biblioteca Nacional Sueca (LIBRIS), una base de dades japonesa (RIKEN base), un projecte mundial de biblioteca digital el BHL (Biodiversity Heritage Library), i el portal d'informació de patrimoni cultural europeu Europeana.

1.3.1. Tipologia de sistemes distribuïts

L'escenari que afavoreix el desenvolupament de sistemes distribuïts és l'aparició de noves disciplines, noves xarxes científiques i els nous models de sistemes de computació en xarxa. La tipologia de sistemes distribuïts presents en la literatura és la següent (Tannenbaum, et al. 2006):

1. Sistemes de computació distribuïda (“cluster computing systems” i “grid computing systems”),
2. Sistemes d'informació distribuïts (biblioteques digitals, bases de dades múltiples i passarel·les temàtiques, etc.), i
3. Sistemes distribuïts generalitzats o “pervasive” (sistemes que impliquen mobilitat i dispositius informàtics integrats).

Els objectius comuns dels tres tipus de sistemes distribuïts són: l'intercanvi de recursos, la transparència, l'escalabilitat, l'obertura, la interoperabilitat i la simultaneïtat.

Els sistemes distribuïts han estat desenvolupats gràcies als avantatges que presenten respecte a altres models de sistemes com els piramidals o els centralitzats. El principal avantatge és que garanteixen una reducció dels costos econòmics, en relació amb el nombre d'equips, recursos humans, i duplicitat de continguts d'informació. Al mateix temps, també augmenta la fiabilitat i seguretat, per exemple, davant una caiguda de l'ordinador (servidor), tot el sistema podria sobreviure. També són sistemes amb major capacitat, més velocitat i més potència per a un millor rendiment de càrrega distribuïda. Un

altre avantatge a tenir en compte és que proporcionen un creixement incremental del sistema a través de l'intercanvi de dades, l'intercanvi de recursos, la millora de la comunicació "humà-a-humà" i una major flexibilitat (com per exemple, repartir la càrrega de treball sobre diverses màquines disponibles).

En aquest capítol només s'analitza la literatura científica que fa referència al grup de sistemes d'informació distribuïts, ja que fan referència a l'òptica de la ciència de la informació, on en general el concepte de sistema d'informació distribuït es refereix a una col·lecció de sistemes d'informació independents que apareix als seus usuaris com a un únic sistema i servei d'informació coherent i de qualitat. En la literatura es troba una àmplia terminologia per descriure el mateix concepte, que és necessari conèixer i esmentar per determinar les característiques principals d'aquests tipus de sistemes d'informació, i al mateix temps poder diferenciar aquells sistemes d'informació de qualitat dels que no ho són.

1.4. Sistemes d'informació de qualitat

Els primers sistemes distribuïts de qualitat, anomenats *subject gateways* (SGs) o passarel·les temàtiques, van aparèixer a mitjans dels anys noranta. En un principi van ser concebuts com infraestructures d'informació acadèmiques i, més tard, com àrees de col·laboració entre diferents comunitats científiques. Juntament amb l'aparició de les "*subject gateways*" es trobem també d'altres sistemes d'informació digitals d'aparença similar, però que no compleixen els objectius i característiques principals que conformen les SGs: "donar accés a recursos d'informació d'alta qualitat".

Existeixen principalment dues àrees geogràfiques d'influència, Europa i Austràlia, de les quals s'han extret les definicions més adients del que són les "*subject gateways*". D'altra banda, s'ha de tenir en compte també, que en aquests dos grans blocs es troben diferents grups d'estudi, el Regne Unit i els Països Nòrdics entre altres iniciatives. En conseqüència, s'observa un àmplia terminologia utilitzada per descriure el mateix concepte:

- Subject gateway, definició de l'*Australian Subject Gateways Forum* (2005): "Mecanisme basat en el web per accedir a una col·lecció d'alta qualitat, de recursos avaluats i identificats per donar suport a la recerca en una determinada disciplina temàtica". A més a més, "és un servei on s'accedeix a través d'un portal, via protocols estàndards (com Whois++, LDAP, Z39.50, Harvest Broker).
- "Selective subject gateways", definició del projecte DESIRE (Comissió Europea) el qual és i ha estat un model a seguir en la creació de polítiques d'ús i funcionament per a la majoria de sistemes distribuïts: "*Selective subject gateways* d'Internet es caracteritza principalment pel seu control de

qualitat. Basades principalment en la selecció de recursos i en la seva descripció manual (tant per part de bibliotecaris, acadèmics com experts), unes activitats que no és fàcil automatitzar”.

Quality-controlled subject gateways (QCSGS) és un concepte nou introduït per T. Koch (2000): “*Quality-controlled subject gateways* com a serveis temàtics amb alts estàndards per al control de qualitat amb una riquesa en la descripció i estructura”. Per entendre-ho millor es descriuen les principals característiques que hauria de tenir una QCSGS:

1. **Selecció de recursos** de forma manual i seguint uns criteris de qualitat: qualitat, autoria, accessibilitat, actualització i rellevància en la matèria. També podríem afegir cobertura temàtica (exhaustivitat i especificitat).
2. **Descripció de recursos** amb estàndards de metadades. Aquest procés ha de tenir una intervenció humana per garantir l’alta qualitat dels recursos.
3. Bona **gestió de la col·lecció**. Actualització de recursos, manteniment, etc. La gestió es pot dur a terme tant de forma manual com automàtica.
4. Alta **qualitat d’accés a les matèries** per mitjà de la indexació i la classificació. És la característica més important que han de tenir les QCSGS. Per tant, és necessari l’ús de vocabularis controlats (tesaurus, encapçalaments de matèries, etc) per tal de facilitar la navegació als usuaris.
5. Oferir una ampla i profunda estructura classificatòria per la **cerca i navegació** dels usuaris.
6. Ús d’**estàndards**. L’aplicació d’estàndards oberts permet la interoperabilitat amb altres serveis demostrant així la seva qualitat com a passarel·les.
7. Característiques de valor afegit. Les passarel·les han d’incloure altres elements de valor afegit que contribueixin a la qualitat. Un exemple podrien ser els serveis de personalització, serveis d’alerta, butlletins de notícies, etc. que ja apareixen en la majoria de SG’S.

Tot i que les definicions presentades clarifiquen el concepte de “passarel·la temàtica”, probablement el terme més correcte o el que es vol adoptar com a model referent és el presentat per Koch (2000), “*Quality-controlled subject gateways*”. S’ha de tenir en compte que tant a la bibliografia com en l’anàlisi de més de 50 sistemes distribuïts que s’ha dut a terme en aquest capítol, segueix havent una clara preferència per l’ús del terme “subject gateway”. Ressaltar

però, que en l'actualitat, com s'anuncia en propers apartats, és un terme en desús. A continuació es mostra la resta de terminologia trobada que fa referència al mateix tipus de sistemes d'informació distribuïts de qualitat o QCSGS:

- "Subject-based information gateway o SBIG" (Desire Project). B àsicament és un sinònim de "subject gateway".
- "Resource discovery system". Ve a ser el terme genèric que inclouria no només les passarel·les temàtiques sinó també els directoris web, OPAC's, les bases de dades locals, etc.
- "Gateway". Fa referència a un concepte genèric on es poden incloure tant les SGs com altres conceptes tècnics d'altres àmbits que no tindrien res a veure. La xarxa RDN (Resource Discovery Network) del Regne Unit a vegades utilitzava aquest terme en comptes de "subject gateway".

Altres temes relacionats apareixen molts cops en la literatura científica sobre sistemes d'informació distribuïts:

- Portal. Similar al que s'entén per "subject gateway" però sense incloure els elements de control de qualitat. És un terme inventat per les empreses privades i sobretot pels cercadors comercials. Molts cops són simples directoris que aparentment poden tenir un cert nivell de qualitat però no solen complir les característiques de les QCSGS, mencionades en les directius de Koch (2000).
- Hub. Terme que va ser utilitzat per RDN per definir com una organització ofereix diferents serveis que inclouen una o més passarel·les de recursos.
- Information vortal. És un sistema d'informació especialitzat en serveis d'Internet. Normalment aquests sistemes es creen en dos escenaris diferents, s'utilitzen com a sistemes d'informació per alimentar o subministrar informació a diferents institucions de recerca que tenen projectes de recerca en comú, o bé són creats per orientats a la difusió de materials per a un públic més ampli, o bé, també potser usat per distribuir informació en un sistema tancat.
- Portal semàntic. El contingut del portal és emmagatzemat i gestionat per una web descentralitzada de proveïdors d'informació individuals o col·lectius (organitzacions). Les comunitats poden afegir nous sistemes de classificació i esquemes organitzacionals i ampliar l'estructura de la informació. Cerca multidimensional mitjançant ontologies de dominis. La informació és semi-estructurada i extensible" (*SWAD Europa. Semantic portals requirements specification, 2001?*).

Aquest marc conceptual ens ha permès veure l'evolució dels diferents tipus de sistemes distribuïts pel que fa a la terminologia, però també en relació amb el contingut. Perquè a partir del 2000 la pròpia literatura ja ens descriu les passarel·les com alguna cosa més que el "control de qualitat de recursos" i el seu "accés". Es pensa més en la creació de nous serveis especialment centrats en l'usuari, i es comença a parlar de portals d'informació i no de passarel·les (Vascoda, 2003).

Finalment, afegir que hi ha diferents definicions desistemes distribuïts. Però en aquest estudi exploratori, els sistemes distribuïts són considerats com un sistema de recuperació d'informació avançada que compleixen els valors ecològics com ara l'aplicació de la "indexació creuada" i "navegació creuada", processos característics en el sistema distribuït de passarel·la temàtica. I per tant, s'escull com a model a analitzar la seva evolució des dels inicis als anys noranta fins l'actualitat.

1.4.1. Context i evolució

Als anys noranta, degut a l'expansió d'Internet, va augmentar la necessitat per part de les comunitats científiques i/o universitàries d'accedir als recursos digitals d'alta qualitat i, al mateix temps, també es demanava la creació d'eines que facilitessin la seva gestió, manteniment i preservació. No és d'estranyar, doncs, que les primeres iniciatives de sistemes distribuïts de qualitat es desenvolupin en entorns acadèmics.

L'aparició de les *subject gateways* va ser a causa del naixement de noves disciplines i noves xarxes científiques, on el creixement del coneixement multidisciplinari i interdisciplinari obria la necessitat de crear noves infraestructures de gestió, organització i recuperació de la informació electrònica, sobretot, informació continguda en un ambient canviant i heterogeni com és Internet. A més a més, les SGs compleixen unes necessitats claus per a les comunitats científiques i educatives: l'intercanvi i la cooperació. En el seu creixement s'han identificat quatre fases:

1. L'esotèrica. Les SGs es consideren xarxes esotèriques o instruments de treball per a les comunitats científiques.
2. Comunitat. El seu ús se centra en les comunicacions, la investigació i els hàbits de col·laboració entre els diversos camps i la investigació acadèmica. Es comença a crear un espai social comú i també està començant a ser utilitzat com a serveis d'informació. Un exemple d'aquesta fase es converteix en l'aparença de la biblioteca virtual (VLib) per CRAD Bernners-Lee (1994), altres exemples d'aquesta fase són Aliweb i BUBL (UK, 1994).
3. Infraestructura d'informació en l'àmbit acadèmic i de recerca. Aquest tipus de xarxes integrades d'informació acadèmica i de recursos estratègics per a

la investigació i l'aprenentatge sorgeixen per la necessitat de col·laboració entre diferents organismes que participen en projectes de recerca comuns i es requereix de plataformes d'intercanvi i gestió de la informació.

4. Infraestructura de la informació pública. En aquesta etapa Internet era un espai de comercialització i privatització que començava a formar part de les infraestructures públiques.

Els primers sistemes d'informació definits com a sistemes d'informació de qualitat, precursors de les passarel·les temàtiques, provenien del Programa eLIB, desenvolupat pel *UK Electronic Libraries* al 1994 pel *Joint Information Systems Committee (JISC)* del Regne Unit. Es tractava de simples sistemes de navegació amb a través de servidors locals. Un altre projecte similar que apareix al mateix any i que sorgeix del programa eLib és Road (Resource Organization and Discovery for subject-based services, 1994), un conjunt de regles del *freeware* que permetia la creació de portals temàtics.

Així doncs, en el Regne Unit, els primers sistemes distribuïts formaven part d'una xarxa anomenada RDN (Resource Discovery Network), fundada pel JISC amb el suport del ESRC (Economic and Social Research Council), AHRC (Arts and Humanities Research Council), la Universitat de Manchester i el suport tècnic de UKOLN (Consorti d'arxius, biblioteques i museus del Regne Unit), amb l'objectiu d'estudiar el millor del web tan per a la recerca, l'aprenentatge com per a l'ensenyament. A partir de 1995, amb el Programa de biblioteques Electròniques (eLib) on s'inclouïa una àrea d'interès sobre "accés a recursos en xarxa" o ANR, van començar a aparèixer projectes de "passarel·les temàtiques" que eren finançats per l'àrea d'ANR.

Els primers projectes de sistemes distribuïts identificats com a "passarel·les temàtiques" que formaren part de RDN (Hiom, D. 2006) van ser:

- Altis: Oci, turisme i esport
- Artifact: Arts i indústries creatives
- BIOME: Salut i ciències de la vida
- EEVL: Enginyeria, matemàtiques i informàtica
- GEsources: Geografia i medi ambient
- Humbul: Humanitats
- PSIGate: Ciències físiques
- SOSIG: Ciències socials

Posteriorment, varen sorgir altres projectes relacionats que formaven part de l'ANR (Access to Network Resources), però, no tots eren identificats com a passarel·les. Aquests tenien l'objectiu de desenvolupar eines de navegació en xarxa i fer créixer les eines i servidors d'informació locals de les pròpies passarel·les temàtiques, eren:

- ADAM: Arts, disseny, arquitectura i mitjans de comunicació
- Biz/ed: Economia i empresa
- CAIN: Arxius de conflictes – era més aviat un recull de recursos que una SG.
- CATRIONA II: *Projecte sobre recursos institucionals i departamentals de universitats escoceses.*
- EEVL: Biblioteca Virtual d'Enginyeria d'Edimburg
- IHR-Info
- OMNI: Xarxa d'informació mèdica
- ROADS (*Resource Organisation and Discovery in Subject-based services*): realment era una plataforma des d'on es desenvolupava software i estàndards que poguessin suportar algunes de les passarel·les.
- RUDI: Recursos sobre disseny urbà (era més un recull de recursos que no pas una SG).
- SOSIG: Ciències socials

Pel que fa a les passarel·les com SOSIG o Humbul, ja existien molt abans de la creació d'aquestes agrupacions en xarxa dels diferents sistemes distribuïts.

Cal fer esment específic també de que l'ANR aconsellava al JISC sobre els sistemes i serveis d'informació, i es treballaven aspectes com l'aproximació a les xarxes d'informació, la importància potencial de les conductes en la recerca i l'aprenentatge de les xarxes emergents. Havien arribat alguns consensos i tenien la convicció que l'efectivitat dels usos dels recursos d'informació en l'ensenyament i la recerca dependria de crear serveis de suport per a l'accés a la informació i coneixement, la descoberta dels recursos, la cerca i la recuperació de la informació entre d'altres. També es considerava que aquests serveis haurien d'estar accessibles dins d'una mateixa xarxa i que haurien de tenir en compte l'ús d'estàndards i protocols internacionals (Gopher, HTTP, WAIS, ftp, etc.), ja que s'estava parlant d'una infraestructura molt diversa. Així mateix, tenien en compte crear serveis orientats a l'usuari, la navegabilitat i els sistemes de cerca.

Com s'explicava al principi, al transcurs d'aquest capítol, s'anirà veient que el tipus de sistema distribuït més estès al llarg dels anys han estat les passarel·les temàtiques o SGs. I per entendre la seva evolució caldrà tenir en compte dues àrees geogràfiques clau, Europa i Austràlia. Europa destaca pel seu caràcter precursor i Austràlia, per la perdurabilitat en el temps de les passarel·les temàtiques que van crear. A manera de resum, la creació i el desenvolupament dels sistemes distribuïts es va produir entre 1992 i 1995. A partir del 1995 els sistemes distribuïts comencen a evolucionar de simples sistemes de navegació de servidors locals a xarxes amb serveis eficaços per a l'aprenentatge i la investigació, que responien a la necessitat de proporcionar i gestionar les col·leccions de recursos per mitjà de la descripció dels recursos, l'accés a través de la matèria (navegació) i accés al cercador (recerca). A continuació, hi hauria un segon període (1996- 2003) en el que es van desenvolupar altres

iniciatives i projectes de “passarel·les temàtiques”, on cal fer una especial atenció al diversos projectes europeus.

1.4.2. Context europeu

El projecte DESIRE (1996-2000) va ser un gran pla europeu que es va desplegar en dues parts DESIRE I i DESIRE II. Hi havia quatre països implicats: Holanda, Noruega, Suècia i el Regne Unit. L'objectiu principal era promoure l'intercanvi d'informació entre la comunitat científica europea. Es treballava en xarxa i principalment aspectes sobre seguretat, connexions, creació d'eines i serveis web, descripció de recursos, cerca i recuperació de la informació, etc.

DESIRE I comptava amb la col·laboració de 23 institucions i va tenir una durada de tres anys (1996-1998). En aquesta primera fase es va centrar en temes de catalogació, indexació, connexions, seguretat i aprenentatge, i també es van crear algunes especificacions per a la descripció de recursos (metadades i criteris per al control de la qualitat en les SGs). En la segona fase, DESIRE II (1998-2000), es van estudiar tres àrees: connexions, accés als recursos web i als serveis. Dintre aquesta fase es va elaborar un manual per a SG i una guia per aplicar criteris de qualitat a les SGs en relació amb les metadades.

També cal fer esment al projecte Renardus (1996), que pretenia donar accés als recursos i serveis web acadèmics existents en l'àmbit europeu. Comptava inicialment amb 14 institucions (biblioteques nacionals, universitats i centres de recerca) . de Dinamarca, Finlàndia, França, Alemanya, Holanda, Suècia i també el Regne Unit (coordinat per UKOLN i que aportava la seva experiència en disseny d'arquitectura, especificacions, models de col·laboració, etc.). En aquest projecte també es treballava en models que permetessin intercanvi de metadades d'accés obert, protocols, solucions tècniques, ús d'estàndards, vocabularis controlats i sobretot poder desenvolupar un sistema pilot que donés un servei funcional als seus usuaris.

Una SG d'aquest tipus permetia a l'usuari tenir un accés integrat a la informació, a part de tenir accés a una àmplia cobertura temàtica a través d'una sola interfície i, a més a més, on s'implementen processos de *cross-searching* i *cross-browsing*, característiques principals de QCSGS.

El projecte Renardus ja venia desenvolupant-se amb iniciatives prèvies quehavia establert les bases tècniques del que eren les SGs, com per exemple els projectes ROADS, DESIRE o EELS. Amb Renardus es crea una nova iniciativa per desenvolupar un projecte europeu amb un sol punt d'accés a recursos web seleccionats i de qualitat, és a dir, es converteix en una nova infraestructura amb models d'interoperabilitat i organitzatius més forts. L'objectiu principal de Renardus era establir un marc col·laboratiu europeu de

SG per tal d'oferir un valor afegit tant als seus usuaris com a les mateixes passarel·les que integraven el sistema. Al principi del projecte va haver una tasca analítica per tal d'esbrinar quins haurien de ser els requeriments necessaris per dissenyar el sistema pilot. Des de la Universitat de Bath es va estudiar la interacció persona-ordinador (HCI) per veure problemes existents quant a la cerca d'informació. També es va estudiar la usabilitat del sistema mitjançant estudis d'usuaris on es plantejaven aspectes sobre diferències culturals dels usuaris, multilingüisme, si les tasques desenvolupades serien o no suficients i la granularitat de les tasques entre d'altres aspectes. El resultat d'aquest estudi va permetre desenvolupar models aplicables al disseny i l'avaluació de les SG i va servir com a base metodològica per avaluar la usabilitat del sistema.

Una altra de les problemàtiques amb les que es va trobar Renardus va ser la interoperabilitat. Fins aleshores, la interoperabilitat només se centrava en l'intercanvi de dades, l'ús d'estàndards de metadades i de protocols d'intercanvi d'informació, però Renardus va anar més enllà per tal d'abordar les problemàtiques d'interoperabilitat semàntica. S'ha de tenir en compte que un dels punts forts i diferenciadors d'aquesta passarel·la és l'accés a les matèries, a través del *cross-browsing* i *cross-searching* els quals permeten la interoperabilitat entre els diferents sistemes que integren la passarel·la a través de les jerarquies temàtiques (vocabularis controlats). A més a més, aquesta SG utilitzava sistemes de classificació com la DDC (Dewey Decimal Classification) per permetre la interconnexió amb la resta de directoris temàtics independentment de la llengua, en un procés que s'anomena mapeig.

El mapeig té el problema que no sempre totes les estructures jeràrquiques poden ser equivalents amb la DDC o amb qualsevol altre esquema de classificació que s'utilitzi. Conseqüentment, cada SG estableix i defineix uns nivells de rellevància per poder identificar almenys els nivells de les estructures jeràrquiques que vol tenir en compte en el mapeig (termes genèrics, específics, no acceptats, etc.).

Un altre inconvenient afegit és que les estructures del *cross-browsing* d'aquest tipus de sistemes són limitades. Per exemple, en el cas de Renardus no es pot anar directament a la descripció dels recursos recuperats ja que al darrera hi ha diferents passarel·les i, per tant, s'ha d'acabar consultant el recurs de forma individual i des d'on és enllaçat. Per la seva part, Renardus, a més de crear un grup de treball per tal d'anar solucionant aquests aspectes, oferia, a totes les SGs participants la possibilitat d'utilitzar la DDC en les seves estructures de navegació.

Avui dia el mapeig de classificacions encara segueix sent experimental tot i que s'aplica en molts altres entorns digitals. En aquest sentit, cal destacar el projecte HILT (GB), finalitzat el 2009, que estudia la problemàtica del *cross-*

searching i *browsing* de matèries a través de diferents comunitats, serveis i per tipus de recursos. Un altre projecte a destacar en aquesta línia és CARMEN, on es va desenvolupar una eina pròpia de mapeig de classificacions, tot i que avui dia aquest projecte alemany sembla totalment parat i la darrera informació que vam obtenir en la seva web és del 2002.

1.4.3. Context australià

Deixant de banda el marc europeu, s'estudia també l'evolució de les *Subject Gateways* a Austràlia, per desenvolupar al llarg del temps moltes passarel·les i per se un dels països capdavanters en aquest camp, ja que tot i seguir els models europeus havien construït les seves pròpies guies i estàndards.

Les SG australianes apareixen una mica més tard que les europees, durant el 1998 i 1999. Eren projectes finançats per l'*Australian Research Council* (ARC), un organisme independent però amb el suport de la *Commonwealth Department of Education Training and Youth Affairs*. L'objectiu dels projectes desenvolupats era establir recursos i materials d'alta qualitat de diferents disciplines específiques en una sola web. Les primeres disciplines van ser: química, agricultura, enginyeria i tecnologies de la informació. Al darrera de cadascuna de les passarel·les trobem les comunitats bibliotecària i de recerca australianes.

Com havia passat anteriorment al Regne Unit, les SG havien obert el debat a Austràlia sobre la necessitat de l'accés i el descobriment dels recursos i serveis que hi havia al web sobre informació de caràcter nacional i de recerca.

Seguint el mateix model europeu i com apunten autors com Campbell (2000) i Schmidt, et al. (2003), els australians consideren que els punts claus que ha de contemplar una SG són:

- **Creació del contingut** (metadades). Les metadades s'entenen com a contingut. Els mateixos recursos web es poden emmagatzemar com a contingut i també les seves metadades, de manera que una de les primeres coses a tenir en compte és l'ús de metadades. L'esquema utilitzat és Dublin Core, com la majoria de SG, però a més a més proposen les seves pròpies metadades o adaptar els esquemes existents segons les necessitats de les diferents SG. Quant a la identificació i descripció dels recursos web en una disciplina consideren que es requereix un alt nivell de gestió estratègica. Per tant, cal tenir en compte la previsió de gestió de duplicats i proposen la creació d'un repositori de metadades que escanegi les ocurrències. Cal recordar que un altre punt important a considerar és la multidisciplinarietat i la interdisciplinarietat, assenyalat pels australians com una dificultat afegida ja que gestionar diferents disciplines, i, més encara si estan interrelacionades, no és una tasca senzilla. Per exemple, en Medi

Ambient podem trobar altres disciplines com la Biologia, l'Agricultura, o Ecologia. En conseqüència, la cobertura temàtica és molt més àmplia i llavors ha de créixer també la col·laboració i cooperació entre diferents sistemes d'informació i també entre els diferents professionals implicats per tal de compartir experiències, decisions, coneixement, etc. Després de dos anys d'existència de SG australianes varen començar a sorgir noves passarel·les per a l'agricultura, el desenvolupament sostenible, l'educació, la literatura, el dret, la música o les arts escèniques, de manera que a l'any 2000, amb suport de la Biblioteca Nacional d' Austràlia, es crea l'*Australian Subject Gateways Forum* (ASGF), una agència de coordinació similar a la del Regne Unit (RDN).

- **Integració de l'accés als recursos** digitals i impresos. Les estratègies proposades pels australians van ser descriure els recursos web amb esquemes de metadades USMARC ja que són els utilitzats per la base de dades de la biblioteca nacional (Kinetic). Podem dir que és una economia d'esforços, que a més a més els permet utilitzar el protocol Z39.50 per compartir recursos impresos i electrònics. D'altra banda també van proposar la "transferència de metadades" entre serveis però amb el desavantatge que els recursos web són molt volàtils i això suposa tenir-ho sempre tot actualitzat.
- **Arxivament** i identificació permanent dels recursos. L'emmagatzematge és un factor clau per a la preservació de les dades. La resta de SGs no ens parlaven d'aquest punt però els australians són experts en temes de preservació de documents digitals, fins i tot tenen una SG sobre aquesta disciplina d'abast mundial. Una primera proposta és implementar processos de còpia per preservar els recursos. D'altra banda, els mateixos membres de les SG han d'assegurar la longevitat dels recursos, tot i que a vegades també dependrà molt de la disciplina en qüestió, per exemple, en informàtica molts cops hi ha documents que ja queden obsolets, etc. Una altra proposta és que s'identifiqui el recurs de forma permanent encara que aquest desaparegui, o bé, es pot redirigir la pàgina web si canvia canviat la URL o redirigir a la gent a una còpia arxivada del recurs si aquest ja no existeix. Actualment les SGs poden dur a terme la preservació dels seus recursos amb l'ajuda de la Biblioteca Nacional que posa a la seva disposició: [PANDORA: Australia's web Archive](#) (una Biblioteca que ajuda i assisteix les SG en temes de preservació dels recursos electrònics i que siguin indexats per al seu futur accés).
- **Sostenibilitat** (coordinació i cooperació). Com apunta Campbell, les SG australianes estan institucionalitzades i la seva estabilitat està prou garantida. A més a més, tenen el suport tècnic de la Biblioteca Nacional, l'econòmic dels seus patrocinis i gaudeixen de beques ofertes per

l'*Australian Research Council*. D'altra banda, la coordinació i cooperació de les SG australianes, evident des de l'esforç de l'*ASGF*, dona assistència tant en aspectes de publicació, directrius de recomanacions, especificacions, estàndards en moltes de les àrees, intentant minimitzar la duplicitat d'esforços i d'infraestructures. A més, cada SG en un context local ha d'aproximar-se al desenvolupament d'una infraestructura interoperable.

- **Interoperabilitat.** Està estretament lligada al punt 1 amb l'ús d'estàndards de metadades. Tal i com s'ha exposat anteriorment utilitzen Dublin Core i també hi ha SG que utilitzen esquemes propis o adaptats d'altres estàndards, sempre de forma complementària. També utilitzen d'altres estàndards per afavorir la interoperabilitat com XML i protocols OAI-PMH (Open Archives Initiative Metadata Harvesting) per a l'intercanvi d'informació.
- **Integració de serveis.** En aquest punt només afegir que les SG australianes, al igual que la resta, també implementen serveis i que ho consideren una forma de maximitzar els beneficis dels seus esforços. Alguns exemples de serveis són: novetats, arxius, notícies o butlletins de notícies, grups de treball, serveis de subscripcions, etc. No inclouen serveis de personalització.

1.4.4. Context internacional

En aquest apartat cal destacar IMesh (1999), finançat per la NDF/JISC *International Digital Libraries Initiative (University of Bath, UK)*, que tenia per objectiu donar suport a les SG en temes de comunicació i col·laboració. Les seves activitats eren principalment la construcció d'un fòrum per intercanviar tecnologies, promoure el desenvolupament de SG i l'organització de workshops per tal de fomentar la participació i comunicació entre SG. La seva finalitat era donar a conèixer els projectes i les experiències en un context internacional, ja que fins aleshores tots els projectes s'emmarcaven en àmbits geogràfics estatals (RDN al Regne Unit, ISAAC a Rússia i la resta de projectes a Austràlia, Holanda, Finlàndia, etc.), però, entre els diferents projectes no s'havia creat cap consens internacional. Principalment, es treballen diferents aspectes tècnics comuns i dissenys d'arquitectures de SG: formats de metadades (Dublin Core o IAFA per exemple), protocols de cerca i recuperació (Z39.50, LDAP, i Whois++), i mecanismes per a la interrogació entre SG (*The Common Indexing Protocol o CIP*). Un dels punts més importants d'estudi en IMesh és la tecnologia utilitzada en les SGs, ja que els softwares que utilitzen de forma individual són incompatibles i, per això s'estudia com fer que les SG puguin treballar juntes i de forma automàtica. Per a què això sigui possible cal que les arquitectures de les SG estiguin ben definides, i coneguin les seves especificacions, components i la seva comunicació. Amb aquesta idea IMesh construeix una

eina pilot per integrar SG i també altres eines que puguin crear metadades. El projecte té tres punts estratègics:

1. Desenvolupar un marc global per a SG les quals puguin integrar els seus esforços individuals i alhora suportar altres eines i metadades.
2. Proveir d'un marc més robust per interoperar entre SG i entre les SG i altres xarxes de serveis d'informació.
3. Crear un entorn favorable per al desenvolupament de sistemes i serveis per tot tipus de públic i amb una tecnologia basada en línea.

Tot el treball i la documentació que s'ha fet en aquest projecte és pot veure en la pàgina web d'IMesh Toolkit, on es fixen les estratègies tècniques i d'arquitectura de la informació que han de seguir les SGs però sempre des del punt de vista tecnològic. Es pot trobar informació sobre els requeriments tècnics de les SGs, revisió tecnològica sobre estàndards, protocols i software (OAI, Z39.50, WHOIS++, RDF, XML, RSS, CIP, LDAP, etc.), personalització (centrada en la personalització de serveis web), anotacions (l'ús de la nomenclatura des de la perspectiva de l'usuari), intercanvi de metadades i el llenguatge d'interrogació (principalment l'ús dels protocols (Z39.50, Whois++ and LDAP).

1.5. Estat actual dels sistemes d'informació: evolució cap a portals semàntics

En aquest apartat es revisen tots els projectes anteriorment mencionats i altres projectes de SGs amb un recorregut històric i dels que s'ha trobat informació bibliogràfica. La mostra representativa escollida són aproximadament una cinquantena de SGs, tot i que per problemes d'accessibilitat (moltes no estan actives), i per la poca informació disponible de la que es disposava tampoc s'han comptabilitzat. Hi ha altres casos on es descriminen SGs perquè no compleixen les característiques mínimes (Koch, T., 2000). Les principals fonts d'informació consultades són el directori de SGs PINAKES (*Heriot Watt University*), l'*Australian Subject Gateways Forum (National Library of Australia)* i evidentment cadascun dels projectes rellevants mencionats en l'apartat anterior com CARMEN, DESIRE, Renardus, etc. [Veure Annex 1.1].

En el següent gràfic es pot veure la representació dels diferents països estudiats on es troben els projectes de SGs més rellevants. També es veurà com el Regne Unit és el país amb una representació més forta tot i que Austràlia també hi té una significant representació:

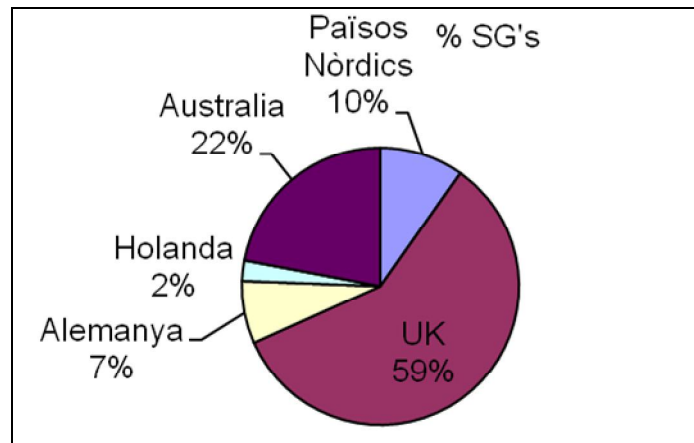


Figura 5. Representació percentual de SG's analitzades

Anteriorment, s'ha descrit l'evolució de les diferents passarel·les temàtiques o SG'S, on es dibuixen tres etapes ben diferenciades, una primera etapa de creació i creixement (1993-2002), una segona etapa amb dues fases (estancament durant el 2002-2008 i fase d'evolució del 2006-2010) i finalment, una tercera etapa de decreixement i/o cessament (2010-2011).

En la primera etapa o fase d'inici i desenvolupament de projectes de SGs, dura aproximadament des del 1993 fins al 2000 i en alguns casos fins al 2002. L'etapa de creixement, amb una evolució mínima, ha estat contínua fins al 2011, tot i que hi ha projectes on la seva etapa final esdevé al 2000-2002. Convé ressaltar la poca informació disponible al respecte, o no es troba informació bibliogràfica, o bé, no hi ha informació en la pàgina web dels projectes analitzats.

En la segona etapa d'evolució de les SGs, s'identifica una etapa de canvi amb dues fases diferenciadores. Una primera fase d'estancament de les SGs que aniria del 2002 al 2008 aproximadament, on la majoria de projectes semblen estar aturats, ja que, o no es troba informació actualitzada (ni en la bibliografia ni en les pàgines web dels projectes analitzats, o bé moltes SGs han deixat de funcionar, com és el cas dels projectes DESIRE, Renardus o CARMEN. No obstant això, molta informació d'aquests projectes aparentment acabats, segueix essent vàlida, ja que en la bibliografia actual es segueix parlant d'ells, com és el cas del projecte DESIRE i les seves directrius per a la creació de SGs. Tanmateix, en aquesta segona etapa s'identifica una fase d'evolució de projectes de SGs (2006-2010), amb una tendència a créixer, ja que s'uneixen amb xarxes cada cop més grans i de caràcter multidisciplinari, seguint el caràcter col·laboratiu i cooperatiu que caracteritza les SGs. D'una banda, trobem el cas del Regne Unit, on hi ha una transformació completa, la xarxa RDN es transforma amb una *subject gateway* anomenada INTUTE (2006-2011), fundada pel *Joint Information Systems Committee* (JISC), l'*Arts and Humanities Research Council* (AHRC). INTUTE segueix el model que ja havia construït SOSIG, però, englobant les disciplines de Ciència i Tecnologia, Arts i

Humanitats, Ciències Socials i Ciències naturals i mèdiques, en una mateixa interfície. Era "És" un dels projectes més capdavanters i competitius que "existia" a nivell mundial i que "va servir" de model a la resta de *subject gateways*, ja en els seus inicis amb SOSIG. De fet, va deixar de ser un projecte i va esdevenir un Sistema d'Informació avançat d'alta qualitat amb una ampla cobertura temàtica i una arquitectura de la informació que la caracteritzava com a *subject gateway*: personalització de serveis, navegació a través d'encapçalaments de matèria, implementa motors de cerca simple i avançat (amb tipologia de cerca facetada), utilització d'operadors booleans (AND, OR NOT), utilització de truncaments i cerca per frases, indexació manual i automàtica, utilitza CDU, etc. D'altra banda, hi ha el cas d'Alemanya on es crea el portal Vacoda (2003-2011). No obstant això, la falta de recursos, principalment econòmics, ha fet que INTUTE finalitzes els seus serveis al juliol del 2011.

En el cas d'Alemanya, on ja existia un model de treball en xarxa d'unes 40 institucions d'àmbit científic. Aquesta xarxa unia un sistema de biblioteques virtuals alemanyes i un sistema d'aliances entre portals d'alta qualitat especialitzats (MedLib en medicina, GetInfo per tecnologia i ciències interconnectades i una tercera aliança per sociologia, psicologia i educació), a més a més, la xarxa incorporava i coexistia amb catàlegs col·lectius i bases de dades bibliogràfiques d'àmbit nacional. Al 2003 es van unir aquests dos tipus de xarxes de biblioteques en un mateix sistema, formant un punt d'accés interdisciplinari amb la possibilitat de fer cerques per diferents camps temàtics, aquest sistema i portal d'informació s'anomena Vascode. Aquest model de portal va ser possible gràcies a què Alemanya constitueix un model federat que facilita i contribueix al funcionament i desenvolupament d'un sistema d'informació simultani. El portal Vascode i els portals temàtics integrats estaven configurats i mantinguts per les biblioteques universitàries i altres proveïdors d'informació amb una reputació internacional. Aquestes institucions oferien l'accés al portal gratuït i continguts d'alta qualitat, que no podien ser trobats pels motors de cerca web normals. No obstant això, per a alguns dels seus serveis i recursos es requeria una quota (pay-per-view). Una característica de Vascode, era que estava connectat a fons de les biblioteques locals i serveis d'obtenció de documents, i les seves dades integrades en el portal de ciència global WorldWideScience.

Malgrat les facilitats i avantatges existents per l'evolució i creixement òptim de Vascode, el portal va aturar la seva activitat al desembre del 2011, com comenten en el seu blog (Hohlfeld, 2011). Hi ha moltes causes relacionades amb la finalització del portal, principalment econòmiques, però, també de gestió, parlen de manca de finançament cooperatiu i sostenible dels serveis i obligacions per part dels associats, falta de suport polític, finançament i lideratge en l'administració de la infraestructura de totes les biblioteques virtuals

i portals especialitzats que envoltaven el portal, i falta de subministrament d'informació, entre d'altres.

En general, aquesta segona etapa es caracteritza per la davallada tant pel que fa a creació de nous projectes o sistemes de SGs, com per la desaparició o aturada de projectes. [Veure Figura 2]

A nivell europeu destaca la parada de projectes capdavanters com CARMEN o DESIRE, amb informació no actualitzada des del 2002. Pel que fa a Renardus, que actualment és accessible mitjançant d'una organització alemanya *SUB Göttingen*, no disposa d'accés al sistema d'informació de Renardus que abans si era consultable, només hi ha accés a informes i publicacions diverses no actualitzades (2002). Un altre canvi és la seva URL, l'accés a <http://www.renardus.org> ja no existeix, ara s'accedeix a través de la <http://renardus.sub.uni-goettingen.de/>

Pel que fa a les SGs australianes, les úniques actives en l'actualitat, segueixen el mateix model de desenvolupament que han seguit des dels seus inicis: col·laboració i cooperació des d'una perspectiva nacional. Utilitzen estàndards de metadades i llenguatges documentals (vegeu *Australian Libraries. Manuals, guidelines*), però, també adapten els seus propis esquemes (<http://www.nla.gov.au/initiatives/sg/standards.html>)⁸. El gran valor afegit de les SGs australianes és el recolzament que tenen a través de la Biblioteca Nacional Australiana, i anteriorment per l'ASGF (*Australian Subject Gateways Forum of National Library of Australia*), des del qual es facilita l'intercanvi d'informació entre les SGs existents i les de nova creació, treballaven en les necessitats comunes com en el desenvolupament tecnològic, temes d'especificacions i estàndards comuns, utilització de vocabularis controlats (tesaurus principalment), temes de qualitat, etc. i buscant sempre la interoperabilitat i sostenibilitat dels sistemes. Del 2005 al 2010 hi va haver un augment de SGs i també de passarel·les especialitzades amb disciplines com ciències del treball, ciència i tecnologia, preservació de documents electrònics (PADI, entre d'altres consultables a *Subject based gateways for archives*), etc. Actualment l'ASGF està inactiu, però, no afecta al bon funcionament de les SGs que si estan en actiu com és el cas d'EdNa, AustLit o PADI. És cert que actualment hi ha moltes menys SGs actives que en els seus inicis, però el fet que els projectes existents estiguin recolzats per la Biblioteca Nacional Australiana fa que siguin estables, i assegura la perdurabilitat almenys dels recursos que continguin aquestes passarel·les en cas que deixessin de funcionar.

Per veure una representació gràfica del que s'ha exposat, s'han valorat principalment els següents paràmetres: país de creació de la SG, any de

⁷ Nota: [inactiva des de [29 Nov 2012 03:21:47](#)]

⁸ Nota: [inactiva des de [29 Nov 2012 03:21:47](#)]

creació i de finalització o continuïtat de la SG, si segueix activa o inactiva i la cobertura temàtica.

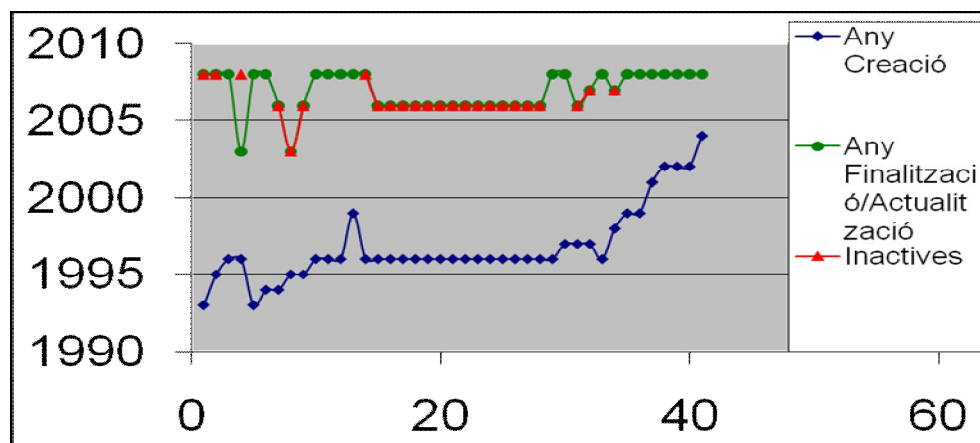


Figura 6. Evolució de les SG's des de 1993 fins el 2010

Com és veu en la gràfica el creixement de les SGs, que es pot observar amb el paràmetre d'Any de creació arriba fins el 2004, en tot l'eix podem observar 4 pujades que fan referència als països representats en la gràfica seguint l'ordre de esquerra a dreta: Països Nòrdics, Regne Unit, Alemanya i Holanda, i Austràlia.

Com es pot veure als Països Nòrdics (representats en els 4 primers punts) trobem passarel·les des de 1993 i la darrera apareix al 1996. Quant al Regne Unit (representada fins a punt 30 aproximadament) també apareixen les primeres passarel·les al 1993 i el punt més alt d'aparició de noves SGs és dona al 1998, després s'observa que la majoria de SGs del Regne Unit apareixen al 1996. La tercera pujada en l'eix de creixement o aparició de SGs correspon a Alemanya i Holanda (representades per només 3 punts) apareixen totes sobre el 1997. Amb referència al cas d'Alemanya, cap al 2005 es llança el portal Vascoda (projecte que es gestava des del 2003 aproximadament) amb la perspectiva de ser un projecte amb gran creixement (més de 40 institucions involucrades), però es paralitza al 2011. Finalment, la darrera corba on s'observa un creixement més exponencial de SGs és Austràlia on apareixen SGs des del 1998 fins el 2004, tot i que hi ha hagut alguns tancaments de projectes al llarg dels anys, tampoc s'han creat noves passarel·les. Austràlia és l'únic país on les SGs s'han mantingut actives fins l'actualitat.

El punt més destacable d'aquesta gràfica és veure la comparativa dels eixos que ens donen dades de quines SGs encara estan actives i les que no. I com s'observa en tots els països hi ha casos de SGs inactives però el cessament més fort es dona al 2006 en el Regne Unit, produït, com s'explicava, per la integració en el mateix any de la majoria de *Subject Gateways* existents al Regne Unit en una única plataforma, INTUTE. No obstant això, és difícil concretar les dades de duració o tancament ja que la majoria de bibliografia de

referència arriba fins el 2002 aproximadament, i, en molts casos, les pàgines web dels projectes ja no estan accessibles, sols en alguns casos es pot agafar de referència la data de la URI d'arxiu de la *Home Page*.

En aquesta segona fase (2006-2010), pel que feia als casos dels països destacats com Alemanya, Holanda o els Països Nòrdics entre altres països involucrats amb els projectes models com Renardus o DESIRE, tot i ser projectes paralitzats, pel que comuniquen a les seves webs, no es considerava que fossin projectes acabats ja que s'esperava el seu redisseny. Tanmateix, al 2005 Alemanya llençava el Portal Vascoda. En l'actualitat i després de 6 anys, es pot considerar que són projectes tancats.

Finalment, com ja s'ha exposat, afegir que les *subject gateways* australianes són les que han sofert menys canvis i han seguit actives des dels seus inicis.

Respecte a la tercera etapa identificada, que va des del 2010 fins l'actualitat (2012-), com ja s'ha anunciat en la segona etapa, és una època caracteritzada per ser l'etapa final de les passarel·les temàtiques.

D'una banda, al 2011 es paralitzen o es tanquen projectes de passarel·les temàtiques principalment a causade la manca de finançament descrit pels mateixos gestors de les SGs. D'altra banda, en el mateix any, l'evolució de la pròpia tecnologia ha fet que apareguin d'altres tipus de sistemes distribuïts.

Pel que fa als casos de cessament de SGs, els casos més destacables són INTUTE (UK, 2006-juliol 2011) i Vascoda (Alemanya, 2005-desembre 2011) en les seves pròpies webs (Intute Staff, 2011) expliquen que el principal motiu del tancament és la manca de recursos econòmics, però en ambdós casos deixen entreveure que poden ser projectes reactivats en un futur. Cal ressaltar que INTUTE ha estat un dels projectes més robusts, primer com a part de l'estructura de xarxa a través de RDN, on des dels seus inicis amb eLIB (vegeu Intute About), i després per la constitució d'INTUTE al 2006 que va arribar fins al 2011 (es va suspendre el seu pressupost central per part del JISC, al juliol del 2010.). A més, ha estat un sistema d'informació de gran influència sobre altres SGs (vegeu bibliografia) i del que s'augurava una gran futur pròsper.

En relació a Austràlia, el tercer país amb major representativitat de passarel·les temàtiques, es van produir diversos tancaments de SGs com WebLaw.edu.au, AVEL, Agrigate o MetaCheml, però no es disposen de dades per confirmar la seva data de finalització, tot i que el darrer cop que es va accedir va ser entre el 2009 i el 2010. Per confirmar la situació d'aquestes SGs es va contactar via e-mail amb Debbie Campbell (*Manager of Infrastructure Projects, Australia Government*), per preguntar-li sobre la situació actual i futura dels projectes de

passarel·les temàtiques. En la seva resposta, data 2009⁹, comentava la manca de recursos per part de les organitzacions en aquests tipus de projectes, la qual cosa dificultava el seu manteniment, a més, la tendència del suport governamental es centrava en l'*open access*, els repositoris institucionals i el patrimoni bibliogràfic. Per tant, les infraestructures cooperatives i col·laboratives com les *Subject Gateways* que necessitaven de molts recursos s'estaven tancant. Es va intentar contactar amb altres responsables de passarel·les temàtiques perquè corroboressin la mateixa situació, però tampoc es va obtenir resposta.

1.5.1. Tendències actuals: sistemes d'informació semàntics

El següent punt a considerar és l'evolució de la tecnologia i dels propis sistemes distribuïts. Les tendències tecnològiques actuals van cap a el web semàntic i el *linked data*. Una possible causa de que les passarel·les inactives no s'acabin reactivant potser la ràpida evolució de la tecnologia, tot i que no es pot descartar que algunes podrien integrar-se o migrar-se cap a nous sistemes distribuïts de tipus semàntic.

El terme web semàntica (*semantic web* en anglès), és un conjunt d'activitats desenvolupades pel *World Wide Web Consortium*, que es basen en la creació de tecnologies per publicar dades llegibles per aplicacions informàtiques. Aquesta idea promoguda per Tim Berners-Lee (1998), se centra en la incorporació de metadades semàntiques i ontològiques (vocabularis controlats) que descriuen el contingut, el significat i la relació entre les dades publicades al World Wide Web, i evidentment, en la seva extensió es pot aplicar a qualsevol sistema d'informació que sigui publicat al web i que vulgui incorporar aquesta semàntica.

El terme *linked data* és definit per la pròpia comunitat, com el fet de connectar dades web relacionades que no estan vinculades prèviament. No obstant això, hi ha un altre terme relacionat, el *linked open data* o LOD (Peset, 2011), concepte relacionat amb l'*open data*, que fa referència al moviment d'alliberació de dades de les organitzacions i que es fomenta amb l'*opening up Government* (2010). Aquest moviment promou que les dades de les organitzacions es facin accessibles per mitjà de formats intercanviables amb llicència per al seu ús i reutilització, la diferència de *linked open data* i el *linked data* és que es parla de dades web en obert i amb format RDF.

En aquest apartat no s'aprofundeix en la matèria sobre el web semàntic perquè es tractarà en d'altres capítols, però, evidentment un sistema d'informació és més avançat i intel·ligent, sobretot en els processos de recuperació de la

9 Entrevista realitzada pel doctorant, on es va preguntar pel desenvolupament i continuïtat dels projectes de passarel·les temàtiques a Austràlia (no es reproduceix l'entrevista per respectar la privacitat).

informació depenent si aquest sistema distribuït aplica o no tecnologia semàntica. Convé ressaltar que, tot i que les passarel·les temàtiques ja s'estudiaven la semàntica (Doerr, 2006) i la seva problemàtica com a base del sistema de recuperació de la informació dels sistemes distribuïts, Vascoda va ser l'única passarel·la temàtica que implementava en l'arquitectura de la informació formats i estàndards semàntics.

Una altra tendència dins el context Europeu és l'orientació de la major part del finançament per projectes tecnològics cap a projectes relacionats amb el patrimoni cultural europeu¹⁰. I és per això que les iniciatives més destacables del web semàntic, les biblioteques digitals i els sistemes distribuïts fan referència al patrimoni cultural, com és el cas del portal d'informació Europea (Europa, 2009) o el cas d'Hispana, el primer servei i sistema d'informació semàntic que es desenvolupa a Espanya des del Ministeri de Cultura entre altres entitats privades per formar part d'Europeana. Aquests tipus de portals gestionen múltiples col·leccions digitals d'arxius, biblioteques i museus inclosos en la iniciativa d'arxius oberts (*Open Archives Initiative* o OAI) que promou la Unió Europea. Europea com a sistema distribuït d'alta qualitat té exigents requisits de qualitat (digitalització, publicació d'accés obert i compliment d'estàndards de metadades) per als països que vulguin participar. Espanya és un dels cinc països que més recursos digitals aporta a Europea, aportant el 8'31% dels més de 20 milions d'objectes (Europeana Facts & Figure, 2012).

En el context de les biblioteques, el web semàntic potser vista com una extensió i/o evolució de les metadades i el que aquestes signifiquen en el concepte de la "biblioteca digital". En aquest context es poden trobar dues dimensions, d'una banda, el web semàntic basat en una lògica i marc comú que s'adapta al web (instruments tecnològics de normalització i estandardització), i, d'altra banda, gràcies a l'ús d'ontologies és vista com a una oportunitat d'accés a informació pública (potenciador de recursos d'informació: millora de l'accessibilitat i recuperació de la informació), facilitant la interoperabilitat entre sistemes i/o repositoris institucionals (biblioteques, institucions patrimonials i similars), i els registres personals o de l'organització existents en els sistemes (Lytras, Sicilia et al. 2005; Ferreira de Castro and Ventura Amorim da Costa Santos 2007). A més a més, a través de la representació dels recursos d'informació, aquests poden ser fàcilment vinculats amb les metadades que es descriuen en els diferents sistemes d'informació. Com apunta FOAF (*the Friend of a Friend Project*, 2000) les metadades

10 VII Programa Marc, fa referència a l'aplicació de les Recomanacions del Consell d'Europa sobre la digitalització, l'accessibilitat en línia de material cultural y conservació digital, publicades en el Diari Oficial de la Unió Europea de 24 d'agost del 2006.

emprades amb el web semàntic poden descriure els recursos amb molt detall, com per exemple, trets característics de les persones com són les preferències personals individuals o de grup, els objectius, els vincles i totes les relacions personals que puguin haver, etc.

Per veure l'evolució de les biblioteques en el web semàntic s'ha fet un anàlisi de l'estat de l'art sobre la bibliografia publicada i els projectes que s'han desenvolupat al llarg dels anys.

En la revisió bibliogràfica, s'ha consultat les dues bases de dades més importants en l'àmbit de les ciències de la informació i la documentació: LISA (*Library and Information Science Abstracts*) i *Social Sciences Citation Index (Web of Science)*, entre d'altres recursos d'Informació com repositoris online, indexats en el cercador especialitzat *Google Scholar*.

Com és sabut, el concepte de web semàntic va començar a tenir força al 1998 amb *Tim Berners-Lee*, i en el camp de les biblioteques o les ciències de la informació aviat es va començar a tenir en compte. Al 1999 (*Tudhope and Cunliffe 1999*) ja apareixen articles respecte les perspectives del web semàntic com a sistema d'indexació i facilitador de l'accés a la informació web, i, com juntament amb les metadades bibliotecàries podien ser claus per a l'arquitectura de la informació dels hipertextos i el WWW. Fins al 2003, els estudis se centraven amb la potencialitat del web semàntic per millorar la cerca web i el descobriment de recursos d'informació a la xarxa, l'estandardització com a element d'èxit i les oportunitats i reptes futurs del web semàntic en noves tecnologies web.

A partir del 2003 ja apareixen iniciatives concretes del web semàntic en les biblioteques (*Goossens 2003*) des de diferents perspectives. D'una banda, iniciatives de les biblioteques nacionals d'Anglaterra, Itàlia i Dinamarca on el web semàntic és percebut com una gran biblioteca universal per ser usada en diversos contextos. Aquestes iniciatives fan referència a intercanvi de dades universal, amb la conversió de formats bibliogràfics i/o controls d'autoritats (tesaurus) continguts en els seus sistemes d'informació (catàlegs o biblioteques digitals), cap a formats semàntics (RDF o XML). D'altra banda, apareixen estudis sobre com les ontologies (llenguatges que forment part del web semàntic), juguen un paper clau en la recuperació de la informació dels sistemes d'informació (biblioteques, arxius, museus, etc.). Per exemple, a través dels controls d'autoritats tradicionals, o amb les ontologies, que permeten relacions semàntiques i sintàctiques, a més a més de poder assignar facetes descriptives dels conceptes relacionats, enlloc de les estructures jeràrquiques tradicionals. És a dir, el web semàntic fa que els sistemes d'informació siguin escalables i extensibles, dues característiques importants en el desenvolupament futur de les biblioteques digitals (*Franklin 2003*). La Biblioteca Nacional de França (BNF) va presentar un exemple d'ontologia en

l'ELAG 2002 (Gossens 2003), i al 2006 apareix el projecte VIAF (*Okalova 2006*), projecte comú entre la Biblioteca Nacional Alemanya, la *Library of Congress* i la OCLC, els quals van començar a unir i vincular els registres d'autoritat de noms de persona entre la *Die Deutsche Bibliothek* i la *Library of Congress*.

Posteriorment, la majoria de treballs tècnics i científics sobre biblioteques i el web semàntic que destaquen en la literatura fan referència als aspectes tècnics i funcionals requerits per a desenvolupar l'arquitectura de la informació de biblioteques digitals semàntiques, ja sigui per accedir als recursos, recuperar informació i/o per a descobrir nous recursos relacionats amb els documents d'una biblioteca concreta. Cal destacar també, que a partir del 2006 comencen a materialitzar-se projectes de biblioteques digitals semàntics o infraestructures d'informació especialitzada i de qualitat gestionades des de biblioteques, etc., que s'exposaran en la secció següent. Al mateix temps, al 2006, es començà a parlar de serveis semàntics i la seva aplicació en biblioteques digitals (Zhiping and Qinglai 2006; Weibel 2007; Peis, Herrera-Viedma et al. 2008). Al 2007 va haver una major producció científica sobre el tema, però no és fins al 2009 quan realment s'experimenta una clara evolució tan en la literatura científica com en el desenvolupament tecnològic i aparició de projectes i iniciatives relacionades amb les biblioteques i el web semàntic. A continuació, es presenta un gràfic de l'evolució de la literatura científica sobre biblioteques i el web semàntic, on s'han analitzat un centenar d'articles extrets de les fonts d'informació esmentades. Cal assenyalar que la informació analitzada fa referència a nous projectes de sistemes d'informació o portals semàntics de biblioteques que apliquen semàntica i que han tingut un fort creixement entre el 2009 i el 2011. Com s'observa en el gràfic inferior, en el 2012 cap endavant ja no hi ha una fase tant de desenvolupament de nous sistemes sinó que és més una fase d'evolució dels sistemes desenvolupats (noves eines semàntiques, millores de tecnologia existent, etc.).

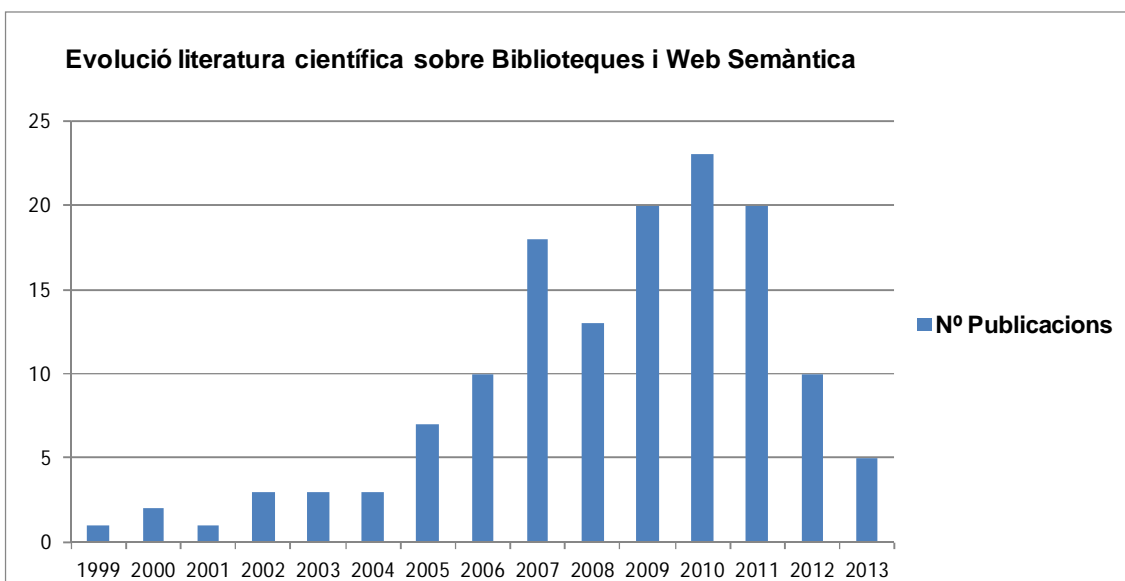


Figura 7. Evolució de la literatura científica sobre biblioteques i web semàntica

Al llarg de la literatura científica es parla de les oportunitats, reptes i tendències del web semàntic (WS) en les biblioteques digitals, però hi destaquen tres tendències generalitzades (*Marshall and Shipman 2003*):

1. El web semàntic entesa com una biblioteca universal: col·lecció de recursos que són fàcilment accessibles i utilitzats pels éssers humans on la informació potser emprada en una varietat de contextos. Es pot entendre com un enfocament cap al moviment de *Linked Open Data* (LOD, setembre 2007-). Un exemple, és BDPedia (creada per la Universitat de Leipzig, Universitat Lliure de Berlin i l'empresa OpenLinkSoftware), projecte per a l'extracció de dades de Wikipedia que és una proposta a una versió del web semàntic. Aquesta està interconnectada amb Geonames, MusicBrainz, CIA World Factbook, Projecte Gutenberg i Eurostat entre altres recursos d'informació. La base de dades BDPedia descriu més de 3.640.000 entitats, entre elles almenys 416.000 persones, 526.000 llocs, 106.000 àlbums de música i 60.000 pel·lícules i conté 2.724.000 enllaços a imatges, 6.300.000 enllaços a pàgines externes, 6.200.000 enllaços a *datasets* externs i 740,000 categories Wikipedia.
2. L'aplicació de tecnologies del web semàntic com a base del treball de fons dels agents computacionals sofisticats per completar les seves activitat. Un bon exemple d'això és la gestió de la informació d'una forma més eficient, facilitar l'accés necessari als recursos pertinents per als usuaris, etc. En concret, hi ha el cas de les consultes via API a repositoris RDF (Jena, Sesame, etc.), o bé, la consulta web amb un punt d'entrada SPARQL (RDQL o Virtuoso), on amb el llenguatge de consulta per a dades emmagatzemades en RDF (SPARQL) es pot extreure

informació en forma de URIs, literals o grafs que permeten consultar i integrar informació de múltiples col·leccions.

3. Finalment, hi ha una tercera tendència, on el web semàntic és entès com a un mètode per a la federació de bases de dades o d'una base del coneixement en concret. És a dir, el web semàntic permet identificar i relacionar coneixement de diverses fonts d'informació i sistemes, no obstant, el coneixement ha d'estar codificat amb els formats del web semàntic (RDF, etc.). Aquest és el cas de projectes de sistemes d'informació, com és el cas d'Europeana (2005-), projecte de biblioteca virtual europea per fer accessible a tothom el patrimoni cultural europeu (Purday 2009).

En aquest sentit, el present treball només agafarà la tercera tendència com a objecte d'estudi, ja que se centra en els sistemes d'informació (biblioteques, centres de documentació, arxius, etc.) i com aquests es transformen i evolucionen amb el web semàntic.

Per veure l'evolució d'aquests tipus de sistemes s'analitzaran diferents projectes relacionats amb les biblioteques digitals en tota la seva extensió, és a dir, sistemes d'informació i documentació gestionats per biblioteques, arxius, etc. que apliquen tecnologia semàntica.

Les principals fonts d'informació consultades han estat bases de dades científiques que permeten fer una revisió de la literatura científica al respecte (projectes sobre biblioteques digitals i temes relacionats) i el portal del W3C (Consorti internacional del *World Wide Web*), concretament el grup especialitzat en biblioteques (*W3C Library Linked Data Incubator Group*), ja que és el principal organisme que contribueix a augmentar la interoperabilitat global de les dades web contingudes en biblioteques. A més de tenir l'objectiu de reunir a les persones involucrades en activitats del web semàntic, principalment relacionades amb el *linked data* i la comunitat bibliotecària, el *Library Linked Data Incubator Group* també vetlla per les iniciatives existents i crea les bases per a la col·laboració futura entre els diferents professionals implicats.

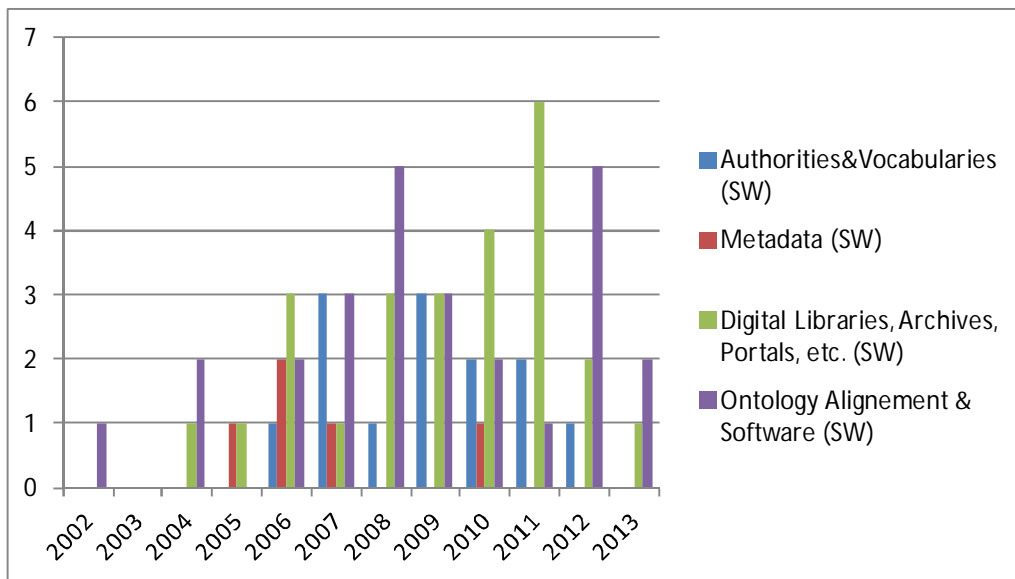


Figura 8. Evolució de projectes relacionats amb el web semàntic i les biblioteques

En l'anàlisi de la literatura científica, s'observen tres períodes temporals destacats (1999?-2005, 2006-2009, 2010-2012) que caracteritzen l'evolució històrica del web semàntic i com aquest s'ha anat aplicant en el món de les biblioteques i altres centres de gestió d'informació i coneixement.

El primer període, es caracteritza per un període més tecnològic, és una etapa d'experimentació on la recerca sobre el web semàntic fa referència bàsicament al llenguatge, estàndards i formats d'ontologies, i al software per a la creació, gestió i alineament d'ontologies. El segon període, 2006-2009, és un dels períodes més forts per a el web semàntic en general i, també, per l'aparició dels primers projectes referents a l'aplicació del web semàntic en l'àmbit de les ciències de la informació (projectes exemplars: AGROVOC, INTUTE o VASCODA). El final d'aquest període destaca per la necessitat de compartir i difondre el coneixement científic (hi ha un gran volum de congressos però cap centrat en el món bibliotecari). Les línies de recerca d'aquest segon període se centren en aparició dels primers prototipus de sistemes d'informació semàntics, aparició de nou software d'ontologies, la interoperabilitat entre sistemes, i l'adaptació de formats, estàndards i vocabularis (món del web semàntic i món 'bibliotecari' entre d'altres). Aquest període destaca també pel creixement i evolució de la recerca del web semàntic en el context internacional, en canvi, a Europa es focalitza amb el centre i nord. La tercera etapa d'estudi, 2010-2012, és l'etapa d'extensió i explosió del web semàntic en relació als projectes centrats amb biblioteques, és el període més madur i per conseqüència d'extensió de la recerca a tota Europa.

1.5.2. Projectes i casos d'ús de biblioteques i el web semàntic

Els casos d'ús registrats en el Grup Incubador W3C *Library Linked Data*, fan referència a vuit grups tipificats segons el tipus de projecte de *linked data*: dades bibliogràfiques, dades d'autoritat, alineament de vocabularis (tesaurus, etc.), dades heterogènies i arxius, citacions, objectes digitals i col·leccions, usos nous i socials.

Els projectes relacionats amb dades bibliogràfiques (Annex 1.2) fan referència a l'estandardització semàntica dels elements bibliogràfics (normalització de registres), duplicació i unificació dels registres (combinar registres duplicats per un mateix recurs en un registre únic mestre), etiquetatge dels recursos web amb termes bibliogràfics estandarditzats, integració d'interfícies de cerca de metadades a través de diversos proveïdors i agregació d'informació (refinar resultats de cerca i ampliar-los amb recursos externs relacionats, identificació de nous recursos i difondre'ls, i accedir al text complet dels recursos mitjançant els enllaços accessibles des del registre bibliogràfic del recurs).

Els projectes relacionats amb dades d'autoritats i *linked data* (Annex 2.3), fan referència a la millora de les metadades pels documents pujats pels usuaris (Si l'usuari vol incloure metadades sobre un autor, el títol de l'obra, paraules clau, etc, el sistema pot suggerir possibilitats recuperats dels fitxers d'autoritats, tesaurus, o vocabularis controlats. El valor suggerit s'identifica pel seu URI. D'aquesta manera, els sistemes de suport a la recuperació precisa, millorar la usabilitat i mitiga problemes com ara la duplicació de registres. A través del *linked data*, les autoritats també poden millorar les funcions de cerca dels repositoris, i obtenir uns resultats més complets, ja que els sistemes fan ús dels registres d'autoritat associats per reconèixer les diferents formes dels noms i dirigir als usuaris finals a tots els registres relacionats amb la forma autoritzada de l'entitat, com ara recerques en els seus registres bibliogràfics relacionats. Un altra característica de la publicació de dades d'autoritat en *linked data* és que en l'agregació d'autoritats cada registre d'autoritat s'identifica pel seu URI (*Uniforme Resource Identifier*). Hi ha diversos sistemes i iniciatives (com és el cas de VIAF), que utilitzen les autoritats per poder crear enllaços semàntics entre diferents registres relacionats, alhora, referits a les mateixes entitats, com es veurà al llarg d'aquesta treball, passa el mateix amb els casos d'alineació de vocabularis.

Els projectes sobre **alineació de vocabularis** s'utilitzen amb quatre objectius o aplicacions bàsiques (Annex 2.4):

1. Reindexar col·leccions. Suport a la indexació de documents amb un Vocabulari "B" sobre la base de la indexació existent amb un Vocabulari A, o viceversa).
2. Cerca de conceptes basats en els vocabularis que indexen col·leccions heterogènies. Suport a la recuperació de documents indexats amb un

“Vocabulari A” que venen de consultes que utilitzen conceptes d’un Vocabulari B, o viceversa.

3. Navegació a través de vocabularis. El suport a l'exploració dels espais conceptuals amb vocabularis, per accedir a als elements de la col·lecció indexada amb els conceptes seleccionats.
4. Fusió de Vocabularis. Donar suport a la construcció d'un nou vocabulari que abasti tan el “Vocabulari A” com el “Vocabulari B”, O bé, ' integrar un vocabulari en un altre (com una extensió o com a satèl·lit de l'altre vocabulari).

Al mateix temps, s'identifiquen tres tipus de casos d'ús segons el W3C:

1. Enriquiment i descobriment de casos relacionats amb l'ús. Aquests casos d'ús es centren en les col·leccions que han sol·licitat vocabularis d'origen o destinació que són part dels esforços d'alineació.
2. Millora d'un vocabulari existents o reutilització d'un vocabulari tant per estendre altres vocabularis de valor com per a servir de base per a la creació de nous vocabularis.
3. Publicació, descobriment i manteniment de les eines o serveis d'alineació de vocabularis.

Projectes relacionats amb dades heterogènies i arxius (Annex 2.5): els arxius difereixen de les biblioteques perquè tenen col·leccions úniques de manera que no han tingut la mateixa necessitat que les biblioteques de compartir informació i/o crear consorcis basats en necessitats com la catalogació de documents. No obstant això, sí hi han hagut reptes similars quant a la necessitat d'agregar material heterogeni, per exemple integració de col·leccions de diferents institucions (arxius, biblioteques i museus). Així doncs els casos d'ús més generalitzats que es poden donar són:

1. Connexions semàntiques. Un grup d'arxius que vol compartir informació per explotar-la. És a dir, tenen catàlegs independents i aquests catàlegs no necessàriament utilitzen els mateixos formats de dades. Exportar i compartir les seves dades en format *linked data* els permet connectar entre les col·leccions utilitzant temes, noms, noms de llocs, i altra informació continguda en les metadades.
2. Descobriments fortuïts. Un arxiu vol oferir un millor servei per descobrir nous recursos web per als seus usuaris. Els mètodes tradicionals de bases de dades no deixa als usuaris seguir les connexions que es poden descobrir en les descripcions dels materials d'un arxiu comprimit. Com que és difícil predir quins mètodes ha d'utilitzar un cercador i quina informació li podrà ser útil, a través del *linked data*, els usuaris poden

seguir els camins previstos pels punts de dades en les metadades de l'arxiu.

3. **Convergència.** Un arxiu vol obtenir una major visibilitat mitjançant la vinculació dels recursos web als seus materials. Això és possible amb mitjançant la creació i exportació de les seves metadades en format *linked data*, i afegint les dades al núvol de *linked data*. Aquest escenari hauria de facilitar la creació de vincles semàntics.
4. **Millora de la gestió de les dades:** construir una xarxa d'institucions que utilitzen metadades similars per descriure les accions de preservació i intercanvi d'informació recollida. S'utilitzen tecnologies del web semàntic per facilitar i millorar la interoperabilitat entre dades heterogènies descrites utilitzant diversos formats de metadades. Augmentar l'ús dels materials conservats digitalment per una base d'usuaris més àmplia. Per exemple el cas del projecte FOAF (2000-) va començar com un projecte experimental d'informació vinculada, creat per *Dan Brickley* i *Miller Libby*, i avui dia és un dels projectes més estesos del web semàntic del W3C, on proporcionen descripcions de persones amb RDF (control d'autoritats), les relacions entre ells i la relació amb tot el que els envolta (les coses que creen i fan o han fet i pel qual són coneguts).

Projectes relacionats amb citacions bibliogràfiques (Annex 2.6): a través del web semàntic és pot millorar també la representació de la informació, ja que des d'una publicació es pot accedir de forma directa a una referència, des de la cita o cites que es troben en la publicació. O bé, es pot millorar la navegació, un usuari pot *clicar* directament en una cita que el porti a la ubicació de referència (URI o Open URL). Una altra utilitat és l'avaluació automàtica de publicacions, la valoració d'un recurs es pot dur a terme de forma fàcil i automàtica, mitjançant l'anàlisi del contingut de les cites contingudes. I finalment, una altre avantatge de l'aplicabilitat del web semàntic potser la recuperació en el context de la citació, és a dir, a través del *linked data* es poden trobar altres publicacions basades en el mateix recurs citat, que es podran utilitzar per citar treball relacionats.

Projectes referents a objectes digitals i col·leccions (Annex 2.7): els projectes referents a objectes digitals estan orientats a la personalització de serveis, i es poden dividir amb quatre grups segons la seva funció:

1. **Agrupament.** Permetre que siguin els usuaris finals que defineixin grups de recursos del web per algun objectiu o raó, ja que molts cops aquestes relacions no existeixen o no es produeixen (no té perquè haver un control de 'grups de col·leccions' per part d'una institució).

2. Enriquiment. Projectes que permeten que els usuaris finals vinculin recursos junts, per exemple, descripcions de persones o temes que puguin estar relacionades.
3. Navegació. L'usuari pot navegar a través de recursos i grups de recursos, a través d'interenllaços, permetent a l'usuari explorar les connexions entre els recursos.
4. Reutilització. Els usuaris tenen la capacitat de tornar a utilitzar la totalitat o parts d'una col·lecció, amb la totalitat o part de les seves metadades en d'altres llocs web disponibles a la xarxa (linked web).

Projectes referents a les col·leccions (Annex 2.8): es refereixen a l'habitual interacció entre un usuari individual en front el servei de biblioteca quan intenta accedir a recursos específics. En aquest cas, les metadades són necessàries per a què es produeixin aquestes interaccions, i, principalment, aquelles metadades que s'apliquen a la col·lecció que en general ja són obligatòries en la descripció de la col·lecció. Així doncs, els casos d'exemple que es troben a la web fan referència al nivell de descripció de les col·leccions que permetran recuperar informació sobre col·leccions contingudes en diferents col·leccions, o bé, sobre innovacions referents a com descobrir/identificar nous recursos i col·leccions basats en el nivell de descripció de la col·lecció. Finalment, es poden trobar serveis d'informació referents per identificar comunitats web específiques (per identificar i classificar col·leccions d'interès).

Casos nous i d'ús social (Annex 2.9): hi ha molts casos d'ús en relació a casos d'ús del web semàntic en relació a aplicacions socials. Principalment aquest casos fan referència a publicació de continguts web i/o a sistemes de col·laboració social.

Quant als casos de publicació de continguts, s'anomenen casos de publicació distribuïda i/o infraestructura d'agregació, que porten informació sobre els usuaris d'un sistema d'informació, o bé són els propis usuaris que volen agregar informació sobre una col·lecció/informació determinada. Aquest tipus d'informació social és informació basada amb el propi usuari, ja sigui informació que aporta l'usuari (comentaris, dades noves, referències bibliogràfiques, etc.), com informació sobre l'activitat de l'usuari en un sistema, creada per les pròpies accions dels usuaris durant l'ús de les aplicacions (compres de llibres, navegació de pàgines web, etc.) que poden aportar informació d'un usuari en relació a un ítem consultat (com les seves preferències lectores, autors, temes d'interès, etc.).

Referent als casos de col·laboració social (*Crowdsourcing / Collaboration / Curation*), fan referència a les contribucions d'un individu en una comunitat. Hi ha diferents tipus de col·laboració social, la tradicional o "crowdsourcing", que permet a moltes persones enviar contribucions a un servidor compartit, la

“col·laborativa” on diverses persones tenen un espai compartit on hi aporten les seves contribucions, i el sistema de curació o “curation” (DH Curation Guide, 2012), una comunitat determinada que gestiona un servei o col·lecció (vegeu cas d'ús W3C d'un Servei d'Informació Comunitari).

Ambdues tipologies de casos d'ús presentades, tenen un interès comú anomenat la participació social, però tenen objectius diferenciats. En el primer cas d'agregació de continguts, l'objectiu és la publicació social, i en el darrer cas, l'objectiu se centra en la col·laboració social.

Els tipus de casos més estesos fan referència a serveis i/o sistemes de gestió de dades bibliogràfiques. Però encara hi ha molta feina a fer, perquè els camps de dades necessaris són a voltes insuficients, manca la gestió d'identificadors d'entitats i enllaços, i en general, falta una norma d'intercanvi de dades bibliogràfiques vinculades. En aquesta línia existeix BibJSON, un estàndard senzill d'intercanvi de dades adequat i compatible amb els principis de dades enllaçades, i influenciat per l'èxit de BibTeX i el format de metadades acadèmiques RePEc. Aquesta norma és de fàcil maneig i compresa per gestors típics de serveis de dades de comunitats, fins i tot sense la necessitat d'emprar eines de programari avançades.

Finalment els nous casos d'ús social, que apareixen són diferents tipus de ‘dades llegibles per màquina’ (poden comptar com a casos d'ús social, perquè són dades compartides per comunitats d'interès). Moltes aplicacions requereixen noves dades llegibles per màquina, que són publicades en diferents formats, incloent *linked data*, RDFa i microformats que, a més són distribuïdes a través d'interfícies de programació d'aplicacions (API).

1.6. Conclusions

L'ecologia de la informació és l'eina i la metodologia que dona accés a informació de qualitat, tractant de vetllar per la supressió de la infoxicació o el soroll documental.

Aquesta és una de les premisses fonamentals de tot aquest treball i al mateix temps és un dels objectius principals dels sistemes distribuïts especialitzats, que estan essent dissenyats com a eines d'accés a la informació de qualitat (provinents d'una o més institucions).

L'engranatge de tot aquest sistema d'informació de qualitat són els processos i eines que permeten la navegació i indexació creuada com a eines ecològiques (Capurro, 1990), per tal d'accedir a la informació sense cap tipus de barreres socials o lingüístiques. Però, en la nova era del web semàntic, l'usuari també juga un paper important, ja que els sistemes s'adapten als usuaris i permeten la

seva participació activa en la creació de coneixement. Els casos d'exemple són els sistemes de navegació creuada, dissenyats per a què els usuaris puguin accedir a la informació d'una forma senzilla, mitjançant la cerca per temes (autonomia de l'usuari per trobar informació), a part dels tradicionals sistemes de cerca simple i avançada que permet autoritzen a l'usuari introduir text lliure per realitzar cerques. Al mateix temps, els sistemes actuals donen la possibilitat de buscar en vocabularis controlats o el que és per al usuari un diccionari de sinònims (ontologies, tesaurus, sistemes de classificacions, i d'altres), tot deixant a l'usuari refinar la seva cerca la seva cerca i arribar a altres col·leccions web fàcilment, a més a més d'interrelacionar dades (*linked data*). En aquests sistemes semàntics, també es pot emprar el llenguatge natural amb la combinació de vocabularis controlats per facilitar i millorar així la qualitat de la recuperació de la informació.

Un dels grans avantatges del web semàntic és poder crear una infraestructura mundial de xarxa de metadades que integrin i interpreti les dades a la web, gràcies als principals elements del web semàntic (RDF, OWL) i els esforços de normalització relacionats (W3C).

A més a més, les biblioteques juguen un paper important tant en la web com en llocs on es concentra la catalogació i classificació del coneixement. Els seus catàlegs inclouen informació semàntica codificada, que fa que la informació sigui accessible a tothom, i amb una major riquesa d'informació, ja que després de la codificació adequada per a la seva adaptació en el web semàntic, el coneixement i les bases de dades que el contenen, queden obertes per a altres aplicacions web i al mateix temps es possible una retroalimentació entre diferents sistemes facilitant així tant l'intercanvi d'informació com la riquesa del creixement. L'usuari final no només té al davant un registre únic d'un recurs, sinó que té accés a enllaços de múltiples registres provinents de diversos recursos.

Un altre avantatge del *linked data* és la interrelació de diferents tipus de dades. Fins ara, les biblioteques i les dades de representacions (per exemple, dades científiques) s'han mantingut separades, però, amb l'ús de *linked data*, que facilita l'intercanvi d'informació i la classificació del coneixement, la integració de diferent tipus de coneixement i sistemes d'informació és possible, un cas exemplar, és les dades enllaçades d'agències europees de medi ambient (Rüter et al., 2010), o el cas de la UBA, l'Agència Federal de Medi Ambient d'Alemanya (vegeu Annex 9).

Malgrat tots els avantatges exposats, el web semàntic té alguns inconvenients associats. Per exemple, a nivell econòmic, social i legal, els agents del sector (editors comercials, universitats, societats científiques, etc.) són encara conservadors i es neguen a alliberar les seves metadades amb una llicència

oberta, enriquint d'aquesta manera els recursos disponibles en el web semàntic.

Referent al desenvolupament tecnològic, també hi ha certes dificultats encara no resoltes, per exemple, existeix gran varietat de tipus de dades enllaçades (persones, llocs, regions, etc.), la coherència dels tipus de categories i llengües (per exemple, noms de carrer en anglès, castellà o xinès), l'heterogeneïtat dels recursos, la falta d'integració respecte a la tecnologia (encreuament de tecnologies, per exemple, RDF basat en XML), etc. Un altre problema important i, encara no resolt, és l'ús del format SKOS, ja que encara cal fer les assignacions entre un vocabulari pre-coordinat i un vocabulari post-coordinat. Aquest tipus de mapeig és necessari per poder utilitzar (per exemple, un post-coordinat) paraules de vocabulari per recuperar els recursos que han estat anotats amb un altre vocabulari (per exemple, pre-coordinat). En general, poden haver també problemàtiques d'inconsistència causades per la herència de les dades existents en un sistema (per exemple ús de MARCXML i evolució cap a RDF, OWL, SKOS).

En general, els obstacles tècnics giren al voltant de la convergència de normes àmplies que s'adaptin a l'intercanvi de metadades bibliogràfiques adients, que serveixin per a l'intercanvi d'informació d'ús comunitari. És cert, que tots els esforços van cap aquesta línia però, encara falta una major universalització dels formats i estàndards del web semàntic, i/o que els sistemes de gestió d'informació tinguin fluxos de treball senzills en referència a les dades enllaçades.

Per promoure el "Linked Data" i estendre el seu ús és necessari també proporcionar interfícies d'usuari per a la gestió, adaptació i manteniment de la informació per professionals i/o usuaris finals del servei. És també imprescindible tenir el suport dels CMS necessaris sobre els quals poden funcionar les interfícies d'usuari perquè l'usuari final pugui gestionar les dades fàcilment (com per exemple *Needlebase*, sistema promesa que proporciona una interfície d'usuari adequada per gestionar dades de tipus gràfic).

És ben cert que els tipus de sistemes ecològics i de qualitat més estesos són els que s'orienten a un públic especialitzat. Actualment, els projectes de nova creació i capdavanters en tecnologia (relacionats amb les biblioteques i el web semàntic) se centren en un altre tipus d'usuaris i col·leccions més universals, s'orienten cap a la gestió de informació sobre patrimoni cultural (per exemple, cas Europea). Tot i que hi ha molts grups especialitzats que treballen amb el web semàntic (com ara la FAO), la seva presència és inferior, i, en molts casos, quan apareixen projectes especialitzats en l'àmbit científic que integren diferents tipus de dades (per exemple, les bibliogràfiques i les científiques) no hi ha la integració de perfils professionals de l'àmbit de les ciències de la informació, o almenys no són visibles en la coordinació d'aquests projectes. Un

clar exemple és INSPIRE, infraestructura d'informació per gestionar dades espacials ambientals que ha de permetre l'intercanvi d'informació espacial ambiental entre organitzacions del sector públic i facilitar l'accés públic a la informació espacial a Europa.

CAPÍTOL 2. COMPARATIVA DE MODELS DE SISTEMES D'INFORMACIÓ DISTRIBUÏTS

2.1. Introducció

El present capítol pretén estudiar els sistemes d'informació distribuïts de qualitat a través de l'anàlisi de casos reals i representatius dels dos tipus de sistemes d'informació considerats models de qualitat: les passarel·les temàtiques i els sistemes d'informació semàntics.

Per comparar els sistemes es durà a terme una recerca qualitativa i quantitativa que analitzarà els sistemes des de tres perspectives: la recuperació de la informació del sistema, els serveis oferts als usuaris i la col·lecció. En cadascuna d'aquestes tres dimensions s'identifiquen uns atributs o paràmetres de qualitat (vegeu Annex 2.1 Glossari d'atributs d'IQ), que s'utilitzaran com a instruments d'anàlisi per realitzar l'estudi comparatiu i que han estat extrets del model d'avaluació proposat en el capítol d'IQ.

En l'avaluació qualitativa, s'utilitzen les tècniques d'observació en la dinàmica del comportament del sistema respecte a l'accés de la informació; és a dir, s'observen bàsicament les dimensions de la recuperació de la informació i de la col·lecció i s'utilitzen unes fitxes descriptives per recollir les observacions. Després, es fa una avaluació quantitativa per tal d'objectivar els resultats obtinguts en el procés d'observació dels sistemes i donar-los una major solidesa. En l'anàlisi quantitativa, s'utilitza una escala d'avaluació on es puntuen diferents indicadors considerats imprescindibles per avaluar la IQ de cada sistema (vegeu Annex 2.2).

L'anàlisi de les dimensions de les col·leccions o de l'usuari no és possible, ja que s'estudien diferents sistemes, i l'especialització i temàtica és diferent. Quant a la col·lecció, se n'analitza l'adequació al tipus d'usuari/audiència al qual es dirigeix (recerca, educació, etc.) i el grau d'especialització (científicotècnic, educatiu, etc.) en relació als usuaris potencials per si s'adapta a les seves necessitats. Així doncs, es proposa avaluar els serveis i el sistema com a factors clau que afecten directament l'avaluació de la IQ des de la perspectiva de l'usuari, i només es farà referència a la col·lecció des de la perspectiva de la interrogació del sistema, per avaluar la disponibilitat de la informació, la completesa d'aquesta (metadades) i l'adequació al nivell de l'usuari de cadascun dels sistemes, entre d'altres paràmetres (vegeu Annex 2.3).

Finalment, caldrà tenir en compte que no tota la informació dels sistemes està disponible o és accessible als usuaris (com per exemple, els serveis interns de pagament) i que, per tant, l'avaluació del sistema no pot ser completa. No obstant això, els resultats d'aquest estudi serviran, d'una banda, per conèixer les característiques significatives dels diferents models de sistemes

d'informació estudiats i, d'altra banda, per crear una guia d'avaluació de sistemes d'informació des del punt de vista de la qualitat de la informació o IQ i donar-li suport (aquesta part s'amplia en el capítol d'IQ).

2.2. Casos d'estudi

Els casos d'estudi que s'han analitzat són Intute, Vaskoda i AustLit com a passarel·les temàtiques, i Europeana i LIBRIS com a sistemes distribuïts semàntics.

2.2.1. Cas d'estudi 1. INTUTE (2006 – juliol 2011)

Url: <http://www.intute.ac.uk/>

2.2.1.1. Descripció. Aquesta passarel·la temàtica prové de la xarxa RDN (Resource Discovery Network) creada el 1995 que és una evolució de la passarel·la temàtica SOSIG (passarel·la especialitzada amb ciències socials), que va evolucionar cap a una passarel·la multidisciplinària, tal com s'ha explicat en els capítols previs. Intute és un portal temàtic que ofereix, a la comunitat educativa i científica, serveis gratuïts en línia per cercar recursos web. Facilita l'accés a milions de recursos d'informació d'alta qualitat disponibles a Internet. Tot i la extensa bibliografia que es pot consultar dels sistemes SOSIG i RDN, no es coneixen tots els detalls sobre l'arquitectura de la informació ni sobre els canvis esdevenuts per la pròpia evolució. Pel que fa a la indexació, navegació i recuperació de la informació, només se'n coneix informació descriptiva referent als criteris de catalogació i metadades. Però sí que se sap que l'evolució mostrava uns canvis en paral·lel a l'evolució de la pròpia web i de les noves necessitats dels usuaris; els canvis significatius més visibles eren l'aplicació de nous serveis basats en la web 2.0, i en la darrera etapa s'evolucionava cap a la web 3.0.

2.2.1.2. Responsabilitat (autoria). L'origen d'Intute es remunta al 1996 amb el programa de biblioteques electròniques eLib, quan un grup de bibliotecaris i investigadors van obtenir finançament del JISC (Joint Information Systems Committee) per desenvolupar un servei nou de passarel·la de recursos d'informació d'Internet; més tard aquest servei va esdevenir el Resource Discovery Network (RDN). Intute va ser creat per un consorci de set universitats, en col·laboració amb una gran quantitat de socis: la Universitat de Birmingham, la Universitat de Bristol, la Universitat Heriot-Watt, de la Universitat de Manchester, la Universitat Metropolitana de Manchester, la Universitat de Nottingham i de la Universitat d'Oxford. També es va tenir el suport de la Universitat de Manchester i Mima (la unió dels centres del Regne Unit, JISC i ESRC).

2.2.1.3. Cobertura: Tot i que el país d'origen és el Regne Unit, l'abast geogràfic no està limitat, ja que té un abast internacional. Quant a la cobertura

temàtica, és una passarel·la multidisciplinar que inclou els àmbits següents: Arts i Humanitats; Salut i Ciències de la vida; Ciència, Enginyeria i Tecnologia, i Ciències Socials.

2.2.1.4. Arquitectura de la informació: Pel que fa a la interoperabilitat en passarel·les temàtiques com a sistemes d'alta qualitat, es fa referència al nivell de desenvolupament de la col·lecció. En aquest cas, el desenvolupament de la col·lecció d'Intute deriva del model de projecte RDN, que es basa en les polítiques de col·laboració i cooperació, tot i que les polítiques específiques es poden trobar en cadascuna de les quatre disciplines temàtiques que Intute cobreix (anteriorment passarel·les temàtiques individuals). Aquesta arquitectura de la informació (IA) és compatible amb el procés de selecció manual que duen a terme els catalogadors coordinadors i els experts en les matèries, d'acord amb les polítiques específiques següents:

La **selecció dels recursos** es basa en un procés manual que duen a terme els catalogadors i coordinadors experts en les matèries d'Intute, els quals segueixen les polítiques indicades:

- Recursos útils per a la recerca, l'ensenyament i/o l'aprenentatge en l'àmbit del Regne Unit.
- Rellevància dels recursos segons les necessitats de la comunitat del Regne Unit a qui serveix Intute.
- Recursos segons el propòsit o l'audiència.
- Recursos segons l'autoritat i la reputació de la font.
- Recursos segons la presentació acurada de la informació o coneixement.
- Evidència dels recursos utilitzats i citats.
- Recursos segons l'originalitat en el contingut i la presentació.
- Recursos segons la cobertura i profunditat de la informació.
- Evidència de les citacions externes (enllaços o llocs web).
- Objectivitat dels recursos.
- Recursos segons l'estructura i la forma: aparença física (claredat en la presentació i navegació), navegabilitat del recurs, accessibilitat (sobretot la seva regularitat), disseny i etiquetatge, termes i condicions d'ús, etc.

- El manteniment del sistema: actualitat i estabilitat o durabilitat del contingut, potencial longevitat del recurs, administració i manteniment del sistema, ús de tecnologies i reconeixements d'estàndards.
- La **gestió i el manteniment** de la **col·lecció** (selecció i avaluació) és responsabilitat de cada grup temàtic que integra Intute (Ciència i Tecnologia, Arts i Humanitats, Salut i Ciències de la Vida) i estan basats en polítiques preestablertes d'alta qualitat (vegeu *Health and Life Sciences Evaluation Guidelines* d'Intute) de selecció de recursos. El sistema d'organització dels continguts es caracteritza per la classificació en categories temàtiques establertes prèviament; sota d'aquestes categories genèriques es troben diferents vocabularis controlats (tesaurus o llistes d'encapçalaments de matèries, entre d'altres vocabularis) que serveixen per organitzar els recursos d'informació. Aquestes categories temàtiques corresponen a les disciplines dels grups temàtics, que alhora són navegables i mostren els termes específics de cada àrea i els vocabularis controlats que s'han utilitzat per indexar els continguts; a més també hi ha categories que permeten navegar en un tercer nivell (subdivisions específiques). La indexació dels recursos és també responsabilitat de cada grup temàtic i dels experts de les àrees que s'inclouen de la passarel·la temàtica, però aquesta selecció ha de complir uns criteris de qualitat estàndards:
 - Han de complir el criteri de context, és a dir, han de ser útils per a la recerca, ensenyament i/o aprenentatge en l'àmbit de l'educació del Regne Unit, i han de ser rellevants per a la comunitat d'usuaris d'Intute.
 - El contingut intel·lectual d'un recurs, és a dir, la informació que conté el lloc web, s'avalua i es prioritza la forma en què es mostra tenint en compte:
 - declaració d'intencions i audiència
 - autoritat i reputació
 - exactitud de la informació i del contingut presentat
 - evidències de les fonts usades i citades
 - originalitat del contingut
 - exhaustivitat i profunditat
 - evidències de citacions externes (enllaços a altres llocs web)
 - Avaluació de l'aparença física i navegabilitat del recurs:
 - claredat de la presentació i navegació
 - facilitat d'ús
 - accessibilitat (recursos que segueixin la normativa web Accessibility Initiative)
 - disseny i composició

- guies d'usuari o documentació relacionada
- termes i condicions d'ús del recurs, etc.
- Avaluació de l'accessibilitat i permanència en el temps dels llocs:
 - actualitat i estabilitat del contingut
 - potencial durabilitat del recurs
 - fiabilitat del sistema (administració i manteniment)
 - ús de tecnologia i estàndards reconeguts
- Criteris bàsics de selecció i avaluació dels recursos d'Intute

Context	Contingut	Funcionalitats	Característiques del sistema
Procedència	Abast / Cobertura	Accessibilitat	Integritat (disponibilitat, etc.)
Autoritat	Precisió	Navegabilitat	Fiabilitat
Singularitat	Actualitat	Termes i condicions i ús	Estàndards i tecnologies apropiades
Relació amb altres recursos	Substantivitat	Drets i legitimitat	
Audiència	Integritat	Disseny, disposició i estètica	
	Composició i organització	User suport and documents	

Taula 2. Criteris de selecció i avaluació dels recursos d'Intute

- **Emmagatzematge.** Intute ajuda a emmagatzemar, compartir i abordar sistemàticament el desenvolupament de la col·lecció amb un servei que recompta els recursos més rellevants per a una disciplina o àrea anomenat "recursos interdisciplinaris", però no explica com aquests recomptes es duen a terme.
- **Actualització.** Intute permet revisar els registres i el manteniment d'enllaços dels recursos a través d'una eina de vigilància de llocs web (Website Watcher).
- **Polítiques d'interoperabilitat:** Intute garanteix la interoperabilitat amb les institucions a través de les següents polítiques:
 - Ús de metadades Dublin Core i RLLOMAP-Cataloguing list i Anglo-American Cataloguing Rules (AACR2) i Resource Description Access (RDA),

- Garantir o treballar per tal d'assegurar la interoperabilitat sobre aspectes tècnics,
 - Fer accessibles els vocabularis controlats temàtics a través dels membres que hi participen,
 - Intercanvi d'experiències i desenvolupament de procediments per catalogar, metadades i vocabularis.
 - Ús de protocols d'intercanvi d'informació generalitzats: Z39.50 (retrieval protocol client), Common Indexing Protocol (CIP), WHOIS++ (directory access protocol).
- **Cerca** (cerca simple i avançada). Es contemplen els dos tipus de cerca que admeten operadors booleans, comodins i cerca per frase, a més a més la recerca es pot refinar a través de 19 encapçalaments de matèria definits per Intute. La diferència entre aquests dos tipus de cerca és que la recerca avançada permet la limitació a certes àrees, temes o tipus de recursos; a més a més, els resultats es poden ordenar per rellevància o per ordre alfabètic, i els resultats també es poden mostrar per títol o per la descripció completa (*all files*). En totes dues opcions, si no es recuperen bons resultats, es pot utilitzar un motor de cerca de Google; per tant, es fa una cerca de forma externa en altres bases de dades fora del catàleg nucli d'Intute.
 - **Navegació.** L'opció de *web resources option* s'escull per cercar o navegar per contingut temàtic a través de matèries determinades prèviament (llenguatge controlat del sistema). Aquesta opció permet navegar amb 19 epígrafs seleccionats per Intute: per sota d'aquests conceptes principals hi ha un desplegament jeràrquic de fins a quatre nivells per sota de l'encapçalament principal.
 - **Cerca i navegació creuades.** Tant la cerca com la navegació inclouen una arquitectura de cerca i navegació creuades, és a dir que des d'una sola interfície i sistema s'accedeix a diferents recursos i a bases de dades distribuïdes. Des de la cerca creuada, el procés consisteix a mapejar la descripció dels recursos (processos de mapeig de metadades o *crosswalk*). Des de la navegació creuada, s'elabora un menú temàtic navegable i jeràrquic (conceptes o termes que provenen de vocabularis controlats) que enllaça amb els recursos indexats en el sistema. La representació d'aquests conceptes descriptors es du a terme a través del format RDF (Resource Description Format), que esdevé un format marc comú entre els diferents sistemes per tal d'intercanviar la informació web i perquè sigui interpretable. A més, s'usa l'estàndard XML per representar la informació de forma textual. Ambdós formats estàndards (XML/RDF) faciliten la interoperabilitat del sistema tant pel

que fa a l'emmagatzematge de dades com per a la interrogació i intercanvi d'informació dels vocabularis controlats emprats.

2.2.1.5. Desplegament de serveis. Actualment no són accessibles. Tanmateix, tenint en compte el grup temàtic hi havia disponibles uns serveis o altres. Tots els grups temàtics, a part dels serveis de cerca, ajuda, contacte, etc., incloïen els serveis següents:

- *Behind the Headlines.* Servei històric de notícies d'interès.
- *Blogs.* Per cadascun dels grups temàtics que formaven la passarel·la
- *Browse Intute.* Navegació temàtica per encapçalaments i subencapçalaments de matèries.
- *ePrints UK.* De moment és un projecte en fase de desenvolupament; es tracta de fer un repositori digital d'àmbit nacional.
- *Harvester.* És una base de dades més potent que permet indexar moltes més pàgines que la base de dades principal d'Intute; és un sistema d'indexació automàtica, per tant, no assegura la qualitat del servei al 100 %.
- *INFORMS.* És un servei només accessible per a estudiants que permet el disseny d'eines basades en la cerca d'informació, que, a més, poden ser personalitzades.
- *Intute-Channel-Lite.* És un servei de RSS que permet instal·lar Intute en un lloc web. Utilitza la tecnologia RSSxpress JavaScript.
- *Intute-Include.* Es tracta d'un *software* que permet descarregar un cercador Intute que es pot instal·lar en un lloc web, i realitzar les cerques des d'allço.
- *Intute-Lite.* És un servei que permet afegir un cercador d'Intute en un lloc web. És una eina més sofisticada que l'anterior però es necessiten coneixements d'HTML per usar-la.
- *Jobs search.* És un servei d'alerta sobre llocs de treball.
- *MyIntute.* És un servei de personalització que permet enviar alertes dels continguts desitjats per correu electrònic des d'Intute, construir/costumitzar col·leccions de recursos descrits per Intute i enllaçar-los en un lloc web, enviar/exportar registres d'Intute per correu electrònic i crear llistes de recursos en línia (*reading lists*).
- *News.* És un servei de notícies unit als blogs que ofereix Intute i al servei de *Job search*. No és un servei propi de la plataforma d'Intute.

- *OpenSearch plugin*. És una aplicació que permet tenir un cercador en la barra de navegació. Requereix Firefox 2.0 o Internet Explorer 7 per instal·lar-la.
- *Research support*. És un projecte d'un servei que serà interoperable amb diverses institucions per tal que la comunitat científica pugui avaluar els recursos electrònics a través de les àrees de coneixement que contempla Intute. Actualment hi ha un servei similar que s'anomena *Suggest a site*.
- *RSS news channels*. És un servei que permet descarregar i guardar els continguts d'Intute (a través de la selecció dels canals temàtics que ofereix Intute) a un lloc web; és similar a un servei d'alertes.
- *View the range of RSS channels*. Es tracta d'un servei gratuït d'Intute per llegir notícies RSS; es pot utilitzar a través del web d'un usuari o servei, o bé es pot descarregar a l'ordinador a través de la versió per a l'escriptori.
- *Search Intute*. Es tracta del servei propi de cerca d'Intute. Intute disposa d'una cerca simple i una cerca avançada. A més a més, Intute ofereix la possibilitat de cercar recursos a través de matèries i tipus de recurs, entre d'altres (navegació).
- *Seminars and events*. Són llistes d'esdeveniments i seminaris.
- *Subject booklets*. Ofereix la visualització, descàrrega o sol·licitud de guies temàtiques.
- *Suggest a site*. És un servei de suggeriments per incloure nous recursos d'informació a Intute.
- *Virtual Training Suite*. Són tutorials d'ajuda per millorar les habilitats en tecnologies de la informació (TI) i en la cerca de recursos a Internet per als usuaris (investigadors, estudiants, etc.).
- *Z39.50 target*. Intute inclou una opció de cerca creuada a través del protocol Z39.50; és a dir, es poden realitzar cerques simultànies a bases de dades diferents que tinguin el protocol d'intercanvi d'informació Z39.50.

2.2.2. Cas d'estudi 2 – Vascoda (2003 - gener 2011)

Url: www.vascoda.de (inactiu) i <http://vascoda.wordpress.com/>; (blog actiu a: <http://wikis.sub.uni-hamburg.de/webis/index.php/Hauptseite>)

2.2.2.1. Descripció. És un portal d'informació multidisciplinar de recursos d'Internet, gratuït i que conté informació científica i acadèmica. Igual que Intute, va deixar de funcionar l'any 2011; com ja s'ha explicat en el capítol de l'evolució del sistema d'informació de qualitat, la manca de pressupost i l'inversió amb altres tipus de sistemes ha fet que moltes passarel·les temàtiques hagin finalitzat els seus projectes. No obstant això, segueix essent un dels grans exemples de sistemes d'informació distribuïts i de qualitat.

Vascoda és un projecte alemany que integra catàlegs nacionals i recopila bases de dades bibliogràfiques d'un total de 40 institucions científiques. Aquest model és possible perquè Alemanya és un sistema federal, característica que fa que disposi d'una estructura bibliotecària compacta. A Vascoda s'integren el sistema de biblioteques virtuals d'Alemanya i un sistema d'aliances entre portals d'informació especialitzats d'alta qualitat alemanys (MedLib per a medicina, GetInfo per a la ciència i la tecnologia, i una tercera aliança per a la sociologia, la psicologia i l'educació, etc.); coexisteixen, donc, diferents tipus de sistemes, com catàlegs col·lectius i bases de dades bibliogràfiques d'àmbit nacional. Aquest sistema d'aliances va facilitar desenvolupament simultani d'aquests dos tipus de xarxes de biblioteques i va contribuir que el 2003 esdevinguessin un sistema únic, Vascoda. L'objectiu de Vascoda era ser un punt d'accés interdisciplinari per cercar informació en àrees temàtiques diferents i de forma simultània, i oferir accés a informació i a textos complets multidisciplinaris amb una interfície integradora.

No hi ha literatura científica sobre les polítiques de selecció, gestió i organització del coneixement, però sí que se sap que el portal Vascoda i els portals temàtics integrats estan configurats i mantinguts per les biblioteques universitàries alemanyes i per altres proveïdors d'informació de reputació internacional. Aquestes institucions ofereixen continguts d'alta qualitat, molts dels quals es poden trobar a través dels motors de cerca web normals. L'accés al portal és gratuït; això no obstant, per a alguns dels serveis i recursos cal pagar una quota de subscriptor (*pay-per-view*). D'altra banda, Vascoda està connectat a fons de les biblioteques locals i de serveis de lliurament de documents. Les dades de Vascoda estan integrades en el portal de ciència global WorldWideScience.

2.2.2.2. Responsabilitat. A través de la col·lecció s'identifiquen tres institucions membres: Electronic Journals Library (EZB), Virtual Libraries i Information Alliance.

Aquesta coalició permet accedir a una gran tipologia documental (dissertacions, bases de dades, bibliografies, revistes, documents a text complet, etc.), així amb un sol sistema d'informació es pot arribar a tota la comunitat científica i acadèmica d'una manera més senzilla.

2.2.2.3. Cobertura. Aquesta passarel·la temàtica tot i ser d'àmbit nacional inclou també altres recursos a nivell internacional.. Quant als temes, Vascoda abasta les disciplines següents: Cultura Angloamericana, Negocis i Economia, Art Contemporani, Cultura Holandesa, Enginyeria, Etnologia, Geociències, Història, Llatinoamèrica, Dret, Medicina, Orient Mitjà, Farmàcia, Física, Política i Pau, Psicologia, Llengües Romàniques, Ciències Socials, Medicina Veterinària i Tecnologia de la Fusta.

2.2.2.4. Arquitectura de la informació

El model de qualitat del portal de Vascoda es basa en dues premisses fonamentals: coordinació i cooperació. Pel que fa a la coordinació, Vascoda, situat a la Biblioteca Nacional Alemanya de Ciència i Tecnologia, funciona com a plataforma d'intercanvi de coneixements en els portals temàtics alemanys. A més a més, coordina activitats de màrqueting i ofereix serveis com llistes de distribució, llistes de correu, blogs, paquets de notícies i tallers oberts als membres Vascoda, etc. Quant a la cooperació, Vascoda, com a principal portal d'informació a Alemanya, esdevé també un punt d'informació científica global, i realitza activitats internacionals i de cooperació; així els associats internacionals que estiguin integrats a Vascoda són visibles mundialment.

Pel que fa a la col·lecció, Vascoda no proporciona informació sobre la gestió i/o les polítiques de la col·lecció per seleccionar recursos, etc. Quant a la col·lecció de recursos digitals (selecció i accés), Vascoda disposa dels serveis combinats de la Biblioteca de Revistes Electròniques (EZB), dels fons de les biblioteques virtuals alemanyes i de l'Aliança d'Informació, i ofereix un accés directe a molts tipus de documents digitals diferents, com ara tesis, bases de dades, bibliografies, revistes, textos complets, etc., que han estat prèviament avaluats per cadascuna de les institucions de les quals provenen. Per tant, s'entén que, en conjunt, cada sistema integrat a Vascoda disposa d'unes polítiques de gestió i de qualitat de la col·lecció pròpies; és a dir, es tracta d'un model de sistema d'informació distribuït però amb un model centralitzat quant a la selecció i a l'accés a la col·lecció. Per a l'usuari, només hi ha un punt d'accés a la informació des del qual s'hi pot accedir de forma fàcil i convenient (integració de tots tipus de documents i àmbits acadèmics); aquest és l'objectiu més important d'aquest projecte. Perquè això sigui possible, els serveis de cerca individuals dels socis integrats a Vascoda són normalitzats per tal de proporcionar un servei de cerca i navegació creuades. Principalment es té en compte el tipus de document indexat, idioma, regles de catalogació, ús d'estàndards de metadades i vocabularis controlats (tesaurus, sistemes de classificació, etc.). Com a objectiu principal, el qual afavoreix la interoperabilitat de les dades dels diferents sistemes, s'utilitzen les directrius d'aplicació del Dublin Core, com a estàndard de referència per representar les dades, tenint en consideració l'anàlisi de tots els elements de metadades utilitzats per la

resta de socis, que, o bé s'han adaptat al model Dublin Core Metadata Element, o bé han realitzat processos de mapeig de Dublin Core a altres formats (pICA3, Pica+¹¹, MAB2¹² i MARC 21¹³) seguint les directrius següents:

- Definició semàntica de cada element (s'hi inclouen millores i esquemes de codificació).
- Definició sintàctica de cada element (normes, codis, etc.).
- Regles de catalogació (per exemple, d'autor, descripció, etc.)
- Repetibilitat d'elements.
- Tipus d'obligació de cada element (per exemple, obligatori, recomanat o opcional).
- Qualificador idioma (pel títol, la descripció, tema) per informar sobre la llengua de les metadades.
- Elements administratius necessaris per a la realització tècnica d'un servei comú (per exemple acrònim i l'URL del servei de cada soci, etc.)

Finalment, la integració de la informació és possible gràcies als punts d'entrada d'informació, sobretot els punts d'accés temàtic (navegació) del sistema, que són components fonamentals d'aquest portal d'informació interdisciplinari.

- **Interfície de cerca** (estàndard i amigable):
 - **Cerca simple.** Des de la pàgina principal ja es pot accedir a la cerca simple amb una opció de navegació per àrees temàtiques diferents que es poden seleccionar/anul·lar per limitar o filtrar la cerca. La tecnologia del motor de cerca permet una cerca ràpida i d'alta estructura a través d'un gran nombre de fonts (75 milions de bases de dades), com bases de dades temàtiques específiques, bases de dades bibliogràfiques, i de directoris de fonts d'Internet, a més, permet cercar simultàniament a un màxim de 100 bases de dades i catàlegs de gairebé totes les matèries.
 - **Cerca avançada:** Des de la pàgina principal, també es podia accedir a l'opció de cerca avançada, amb l'opció de navegació temàtica (per disciplines). Aquest desplegament per tema és un desplegament jeràrquic de la classificació Decimal Dewey (DDC), que s'estava desenvolupant quan la plataforma funcionava.
- **Interfície de *subject gateways*.** Permet accedir a la col·lecció a través de la navegació per matèries. Les passarel·les temàtiques es divideixen jeràrquicament en grups, matèries/regions i portals especialitzats (bases de dades i catàlegs). Els portals temàtics es classifiquen en un màxim de tres matèries. En el nivell superior, la navegabilitat es dona a partir de 5

¹¹ Pica+: The German National Library's internal format

¹² MAB2: Maschinelles Austauschformat für Bibliotheken, format alemany per a l'intercanvi de dades.

¹³ Marc21: format per a dades bibliogràfiques, produït i editat per la Library of Congress. Versió actualitzada abril 2013. Url: <http://www.loc.gov/marc/bibliographic/>

grups temàtics o de matèries (Enginyeria i Ciències Naturals; Medicina i Ciències de la Vida; Ciències Jurídiques, Econòmiques i Socials; Humanitats i Estudis Culturals, i Regions i Àrees Culturals), i en un segon nivell es pot navegar a través de 22 subtemes i 11 regions o àmbits geogràfics. En cadascun dels 5 grups hi ha tres opcions: classificació, navegació de temes específics i filtratge de resultats, i els resultats es poden ordenar per data, rellevància o camps (any de publicació, tema, autor/creador, idioma, etc.). A més a més, si és possible, es pot veure el text complet del document, se'n pot comprovar la disponibilitat, es poden desfer les cerques, exportar els resultats, imprimir, gestionar els registres, etc.

- **Sistema d'Indexació.** No s'especifica com a tal, però se sap que la recuperació de la informació es du a terme a través de metadades i es va estudiar la possibilitat d'utilitzar la DDC com a sistema de mapeig. La metadada 'descriptor' és un element clau del sistema, ja que durant la recuperació d'informació el sistema permet a l'usuari la possibilitat d'utilitzar sinònims i descriptors indexats; la majoria de termes estan amb anglès i alemany.
- **Sistema de cerca (metacerca).** Es fa servir un metacercador basat en l'estructura dels serveis web. S'usa WSDL, format XML, per descriure serveis de xarxa, juntament amb protocols de comunicació SOAP 1.1, HTTP GET / POST i MIME. El conjunt d'aquests elements fa possible l'accés a les bases de dades, a OPAC i a altres recursos d'informació:
 - Es fa servir un sistema híbrid de metacerca i de tecnologia de cerca, un sistema tecnològic intel·ligent que permet a l'usuari accedir ràpidament tant als processos de cerca com als serveis. A més a més, el sistema ordena cronològicament els resultats en ordre ascendent o descendent i, alhora, també es poden classificar per rellevància. Finalment, també es poden refinar les cerques per any de publicació, tema, autor, tipus de publicació o llengua.
 - A través dels resultats, l'usuari pot emmagatzemar els documents més interessants, com una mena de favorits o marcadors, i pot usar un historial de cerques desades. No obstant això, per utilitzar aquests serveis cal ser membre de la plataforma, és a dir, cal registrar-se a la secció My VASCODA, i així poder emmagatzemar els resultats de les cerques realitzades.
 - Un altre punt interessant és que en cada cerca apareixen els proveïdors d'informació (centres o sistemes d'informació, com per exemple, OPAC de bases de dades bibliogràfiques) des d'on es recupera el registre consultat.

- **Cerca creuada.** És un procés de cerca a través de metadades (mapeig/*crasswalk* de metadades provinents de diferents bases de dades): com a exemple, segueixen el projecte MetaLib que es basava en el model Dublin Core (amb refinaments, qualificadors i vocabularis controlats inclosos) com a estàndard de metadades per a l'intercanvi de formats entre les metadades locals (biblioteques que intengren Vascoda) i el format de metadades dissenyat per Vascoda. Aquest sistema estava en fase de desenvolupament, però s'havien fet proves, i l'objectiu final era assegurar i garantir l'estandardització i la interoperabilitat entre els diferents proveïdors d'informació.
- **Navegació creuada.** No s'ha aplicat el model navegació creuada o navegació per matèries, però era un dels objectius i s'esperava que el desenvolupament fos a través de l'ús del sistema de classificació de Dewey (DDC). Els darrers anys, es va estudiar i contemplar la possibilitat d'usar ontologies com a models per estandarditzar col·leccions heterogènies dels recursos d'informació; tampoc es descartava l'ús simultani de diversos mètodes. En aquesta línia s'investigaven dos enfocaments:
 - **Crossmatching o mapeig creuat.** Es tracta del mapeig entre diferents terminologies, tesaurus i sistemes de classificació. No és l'elaboració de cap metatesaurus, sinó que es tracta d'una tècnica extesa per normalitzar la terminologia a tot el món (web 3.0). A diferència de les concordances creuades (indexació creuada i navegació creuada), representa un model que només conté un vocabulari amb les connexions semàntiques genèriques entre terminologies que es donen en els diferents sistemes integrats en una passarel·la o portal d'informació. En aquests sistemes, molts termes poden no estar relacionats, fet que es diferencia de l'enfocament de les ontologies (relació dels termes a través del significat). Es pot dir que les concordances creuades només cobreixen el problema estàtic de la transferència de dades. Pel que fa al cas de Vascoda, havia desenvolupat 18 mapejos de concordances creuades.
 - **Enfocament estadístic o quantitatiu.** La transferència pot ser considerada com una situació confusa entre els llenguatges de descripció del contingut (metadades). Així doncs, se suggereixen diversos mètodes per millorar la recuperació d'informació entre la terminologia de l'usuari i el contingut de la base de dades. El projecte Vascoda, en aquest sentit, seguia el model probabilístic proposat per Zhang (2004), que és un mètode aplicable i conegut per transferir mòduls bilaterals de sistemes d'organització del

coneixement (KOS). Els primers estudis empírics es troben en el treball de Marx (2005).

2.2.2.5. Desplegament de serveis

Es donen a través de My VASCODA. No s'hi pot donar d'alta si no es forma part de les xarxes de biblioteques alemanyes, però com a usuaris externs es pot accedir a una part dels serveis d'emmagatzematge de cerques (*search history*) i selecció de recursos preferits (*folder*).

Entre els serveis hi ha els serveis de cerca ja exposat i serveis web 2.0: *blog*, *newsletter*, *Twitter*, *RSS-Feed* i *FriendFeed*¹⁴.

2.2.3. Cas d'estudi 3. AustLII (1994? -)

Australasian Legal Information Institute. Url: <http://www.AustLII.edu.au/>

2.2.3.1. Descripció. És un portal d'informació legal de l'Institut d'Australàsia, amb el sistema d'accés lliure més popular d'Austràlia i Oceania per obtenir informació legal. Rep més de 900.000 visites diàries, així que atén a un gran nombre de peticions d'informació dels usuaris. Aquest sistema contempla un model distribuït (diferents institucions aporten continguts al sistema, però segueixen un model base per gestionar la informació).

El principal objectiu d'AustLII és facilitar l'accés a la justícia a de la informació. Actualment, es considera una de les majors fonts de materials legals a la xarxa, ja que disposa de més de quatre milions de documents consultables. AustLII publica informació jurídica pública, és a dir, materials legals primaris (lleis, tractats i decisions de les corts i tribunals), materials legals secundaris creats pels organismes públics i que són d'accés obert (per exemple, reformes legislatives i informes reals de les comissions) i una col·lecció important de revistes de dret.

La col·lecció AustLII conté 554 bases de dades d'aspectes legals, entre les quals s'inclouen bases de dades de text complet de legislació i decisions de la Court and Tribunal Court. Actualment també es pot accedir a les bases de dades de legislació de la Commonwealth, Australian Capital Territory, New South Wales, Northern Territory, Queensland South Australian, Tasmanian, Victorian and Western Australian; a la legislació i les regulacions de la majoria dels tribunals federals (High Court, Federal Court, Family Court, AAT, etc.), i a la majoria dels tribunals estatals. AustLII també inclou un nombre de bases de dades especialitzades, amb que inclouen matèries específiques, una col·lecció extensa de reformes legislatives, una col·lecció de revistes de lleis, així com l'índex més ampli de legislació australiana a la xarxa.

¹⁴ Friend Feed: eina 2.0 que proporciona informació de referències de recursos d'informació web d'altres usuaris, similar a un sistema social de recursos de referència online.

2.2.3.2. Responsabilitat. AustLII és un sistema posat en funcionament conjuntament per la Facultat de Dret de University of Technology, Sydney (UTS) i la University of New South Wales (UNSW). AustLII depèn econòmicament de la generositat dels seus contribuents; aquestes contribucions gaudeixen de deduccions d'impostos.

2.2.3.3. Cobertura. La cobertura geogràfica és principalment Australàsia (Oceania sud: Austràlia Melanèsia i Nova Zelanda). Quant a cobertura temàtica, només inclou la disciplina de Dret.

2.2.3.4. Arquitectura de la informació

Es tracta d'una plataforma col·laborativa que permet diferents nivells de coordinació, per tal de que les bases de dades dipositades en diferents servidors puguin interactuar entre elles, amb l'objectiu de donar suport a la recerca sobre aspectes legals. Això és possible gràcies a l'ús d'un motor d'inferència (YSH) i a la seva interfície web (wysht, Allen, R., 2000); a través d'aquest motor d'inferència es fan vincles als enllaços d'hipertext automatitzats al sistema (AustLII) i a les cerques de text vinculades (cerques de l'usuari). Aquest sistema són programaris col·laboratius que permeten vincular-se amb d'altres programaris web amb la sentència INCLUDE. Això permet la col·laboració amb un grau de coordinació baix i no és necessari un desenvolupament compartit de cap pàgina web o l'ús conjunt d'altres recursos web; no es descarten d'altres models de col·laboració, però tampoc es disposa de més detalls tècnics. Quant a la col·lecció AustLII, defensa 6 requeriments bàsics perquè les dades siguin d'accés lliure i es puguin republicar:

1. Disposició en forma completa, amb informació addicional preferiblement. amb la font detallada (per exemple, consolidació).
2. Disposició tenint en compte l'autor, incloent-hi citacions.
3. Disposició tenint en compte la facilitació de la difusió.
4. Disposició de qualsevol republicació en un cost marginal-base.
5. Subministrament sense restriccions de reutilització o de drets de llicència.
6. La preservació d'una còpia de l'autoritat pública.

Hi ha disponible una wiki de codi obert (wiki-wiki AustLII) per compartir informació i desenvolupar eines i tècniques que ajuden arribar a consensos i desenvolupaments comuns per garantir la interoperabilitat dels sistemes.

- **Cerca.** Es fa servir un motor de cerca propi: SINO (vegeu documentació en SINO AUSTLII help); és un motor de cerca normal que permet la cerca amb operadors booleans (AND, OR i NOT) i de proximitat (near, w/n i pre/n), també permet truncaments (*), literals ("term"), etc. A més a més, incorpora un rànquing de rellevància tenint en compte: a) quants

cops apareix el terme de cerca en cada document, i b) el nombre de vegades que apareix cada terme de cerca en el document, amb l'ajust d'un coeficient corrector aplicat a la ocurrència d'un terme, per exemple, que aparegui en el camp títol. AustLII proporciona diferents formes de cerca:

- **Cerca simple.** Cerca bàsica estàndard, que apareix a la pàgina frontal. Els termes de cerca i els operadors s'introdueixen a la finestra de cerca i s'inicia la cerca a través de l'opció "Search".
 - **Cerca avançada.** També s'anomena cerca completa. Permet combinar les equacions de cerques per a la cerca simultània a diferents bases de dades. És a dir, les bases de dades són una opció per limitar la cerca per tipus d'institució o base de dades. D'una banda, si no se selecciona cap recurs, la cerca es fa sobre la legislació, la jurisprudència i els materials secundaris de totes les bases de dades al mateix temps. D'altra banda, si se selecciona la opció "Select Database", es limita l'abast de la cerca a una sola base de dades, o a un grup predefinit de bases de dades ('Totes les bases de dades de legislació', 'Totes les bases de dades de jurisprudència' i 'Totes les bases de dades de materials secundaris'). Les bases de dades d'AustLII només tenen 2 camps cercables, títol i text; en la cerca avançada hi ha una altra opció de filtratge que es pot utilitzar per concretar la cerca "Search Scope", que permet fer una equació de cerca en diferents opcions: 'qualsevol d'aquestes paraules', 'totes aquestes paraules', 'aquesta frase', 'nom de legislació', 'nom de cas', 'títol del document', 'equació booleana' i ús d'autocerca.**Cerca LawCite:** És un sistema de cites automàtiques de casos jurídics internacionals. És a la pàgina principal i es tracta d'un servei de cerca addicional gratuït. Permet als usuaris cercar per cites; a més les cites són avaluades (valoració amb estrelles). LawCite és una part d'un projecte de tres anys finançat pel Consell Australià de Recerca en cooperació amb tribunals australians i altres organitzacions comercials i no comercials. Es desenvolupa a AustLII en col·laboració amb altres membres del Moviment de la Llei de Lliure Accés com: BAILII, HKLII, PacLII, SAFLII, NZLII i CYLAW. Hi ha indexats prop de 18.000 informes sobre lleis i series de revistes, i uns 3,8 milions de casos i articles de revistes especialitzades de tot el món.
- **Navegació.** La navegació del web és bastant alta. Hi ha diferents seccions i apartats que faciliten l'accés a les col·leccions i a les bases de dades accessibles des de AustLII. A la pàgina principal, s'anuncien les bases de dades incorporades recentment, amb enllaços directes a les

interfícies de cerca. La navegació en relació amb la col·lecció és bastant baixa (no hi ha un desplegament jeràrquic de navegació per matèries, tot i que s'hi treballa per incorporar-lo). La navegació per les bases de dades es pot du a terme a través de tres seccions situades a la columna esquerra de la pàgina principal, que són el punt d'accés a les col·leccions de les diferents bases de dades. La primera secció és "Casos i legislació", aquí hi ha enllaços per accedir a "Totes les bases de dades" o bé a les bases de dades per tipus d'institució que es vulguin seleccionar. Després hi ha una secció de biblioteques amb navegació és temàtica (només amb un nivell de jerarquia); finalment, la darrera secció és l'accés als catàlegs, on es pot navegar per jurisdicció, per matèria o per àmbit geogràfic.

- **Cerca i navegació dintre les bases de dades.** En cada base de dades hi ha un menú navegable amb set opcions per fer cerques en legislació recuperada:
 - **Table:** accés a la taula de continguts
 - **Notes:** mostra les notes associades, esmenes, etc.
 - **Noteup:** cerca automàtica a tots els materials d'AustLII tenint en compte la secció on es troba l'usuari
 - **Previous:** opció de navegació cap a la secció anterior, "pas enrere" a través de les seccions.
 - **Next:** opció de navegació per passar a la següent secció; pas a pas entre seccions
 - **Download:** opció de descàrrega de còpies de la legislació general (no una sola secció); es pot escollir entre els formats ASCII o RTF (si són disponibles), i es poden imprimir o guardar amb comandes normals de navegació web.
 - **Related items:** si estan disponibles, enumera els articles relacionats (exposició de motius, declaracions, notes, etc.)

2.2.3.5. Desplegament de serveis

Actualment el desplegament de serveis no està gaire desenvolupat, si es compara amb altres portals d'informació com Intute. Disposa dels serveis bàsics, com accés als serveis de cerca (accés a aproximadament 430 bases de dades de text complet a la jurisdicció d'Australàsia), contacte, guia de l'usuari, serveis d'instal·lacions de programari (ha desenvolupat aplicacions gratuïtes per a Google Android i dispositius iOS d'Apple per ajudar els usuaris a obtenir el màxim profit dels recursos del web), *toolbar* (eina de navegació instal·lable al navegador del usuari), servei de subscripció de notícies (notícies de lectura en línia o per correu d'electrònic a través de la subscripció a la distribució), serveis d'esdeveniments i seminaris i Twitter, entre d'altres serveis futurs amb els quals es treballa:

- Serveis d'alerta. Aplicació de diversos serveis, com l'ús de serveis d'alerta, incloent els canals RSS, per identificar contingut nou, com projectes de llei nous o actes, o esmenes a la legislació existent, lleis noves i articles de revistes recents que apareixen a AustLII. Però l'ús d'aquests serveis és complicat, ja que amb la jurisprudència hi ha un problema afegit, que és la preservació dels canals RSS que s'utilitzen per alimentar la republicació de continguts, ja que és difícil controlar els continguts a causa del canvis que s'hi produeixen i dels recursos d'on s'extreuen: les lleis s'envien directament des dels tribunals i dels editors que reben dades directament dels tribunals, etc.
- Minería de dades. Creació de matèries útils en biblioteques específiques per àrees relacionades amb les lleis australianes (Llei indígena, legislació industrial i fiscalitat) a través de mitjans parcialment o totalment automatitzables, que actualitzen automàticament els continguts de qualsevol base de dades AustLII. L'enfocament inclou 10 mètodes d'identificació i aïllament en les bases de dades de contingut general sobre matèries concretes, mitjançant processos automatitzats de cerca.
- Generació de continguts per part de l'usuari. No obstant això, a causa de la naturalesa autoritària dels continguts jurídics, hi ha limitacions importants quant a la interacció de l'usuari amb el contingut. Una de les propostes és que hi hagi un servei de suggeriments de cites addicionals per a documents i citacions no reconegudes en els documents (per millorar el servei de LawCite), o comentaris, etc.

2.2.4. Cas d'estudi 4. LIBRIS (2003? -): Linked data interface to the Swedish LIBRIS union catalog.

Url: <http://libris.kb.se/?language=en>

2.2.4.1. Descripció

LIBRIS és un servei de cerca nacional que proporciona informació sobre els ítems que hi ha a les biblioteques universitàries i de recerca de Suècia, i en una vintena de biblioteques públiques. Es pot accedir a diferents biblioteques simultàniament. Aquest sistema d'informació distribuït s'actualitza diàriament i dóna accés a documents de tot tipus (llibres, revistes, articles, mapes, cartells, música impresa, recursos electrònics, etc.).

LIBRIS és el nom en suec de 'catàleg col·lectiu', però pel que fa al sistema es contemplen les mateixes premisses que les passarel·les temàtiques (cooperació, coordinació i sostenibilitat). Es pot dir que LIBRIS és el catàleg col·lectiu de les biblioteques acadèmiques i de recerca de Suècia. Així doncs, des de LIBRIS, quan es fa una cerca, es poden veure les biblioteques que

tenen el títol que interessa a l'usuari, si el títol està disponible per al préstec o no, i si és un document d'Internet disponible a text complet. Els resultats mostren també informació de les metadades bàsiques de títol, autor, any de publicació, etc., i sovint es proporciona informació addicional, com taules de contingut, resums, il·lustracions i enllaços a documents a text complet, etc. Les biblioteques que formen part de LIBRIS proporcionen la catalogació dels documents i contribueixen conjuntament al desenvolupament dels continguts de la base de dades.

2.2.4.2. Responsabilitat

El Departament de la Biblioteca Nacional de Suècia és responsable del funcionament i desenvolupament de LIBRIS.

2.2.4.3. Cobertura

LIBRIS conté 6,5 milions de títols provinents de 300 biblioteques. Comprèn la totalitat de la Bibliografia Nacional de Suècia i una selecció de bibliografia especialitzada de diverses àrees. Les biblioteques de LIBRIS produeixen totes les bibliografies, que poden ser registrades com a subconjunts de la base de dades de LIBRIS, consultables en la opció de cerca "Subdatabases", que apareix a la pantalla d'inici. Cada bibliografia és responsable de seleccionar el material que cobrirà (articles de revistes, llibres, enregistraments de so, etc.). Les bibliografies que es poden buscar com a subconjunts de la base de dades de LIBRIS a "Subdatabases" compren bibliografies nacionals, especialitzades o temàtiques, bibliografies locals o regionals, així com bibliografies d'autors/persones. A diferència de les passarel·les temàtiques, aquesta col·lecció no es basa en els recursos d'Internet i no hi ha un servei per seleccionar els recursos d'alta qualitat repartits al web, tot i que no s'ha de descartar aquesta possibilitat a causa del moviment global de dades obertes vinculades *Linked Open Data*, tecnologia que també aplica LIBRIS.

2.2.4.4. Arquitectura de la informació

Quant aquesta secció, no hi ha informació disponible en el web del sistema, però sí que se'n troba en la literatura científica. Aproximadament 170 organitzacions bibliotecàries cataloguen a LIBRIS; per aquesta raó diversos vocabularis controlats s'indexen i es mapegen en el mateix sistema per tal de ser interoperables; tot i això no ha estat possible obtenir-ne informació completa. Pel que fa als aspectes tècnics se sap que s'apliquen dades enllaçades per assegurar la interoperabilitat del sistema, però tampoc hi ha informació disponible sobre l'ús dels processos de cerca, indexació i navegació creuades. Les característiques d'interoperabilitat són:

- Sistema: hi ha molts mètodes diferents amb l'ús de protocols SRU i Z39.50 que donen suport a l'accés a les dades de LIBRIS.
- Sintaxi: s'utilitza l'eina XSearch (eina de cerca clàssica però amb més funcions de cerca addicionals), que és una versió lleugera de l'API, per accedir a les dades de LIBRIS a través d'HTTP d'una manera simple i directa. En l'actualitat, es proporciona suport per als formats RDF, MARCXML, Dublin Core, Json, RIS i MODS.
- Estructura: ús de formats estàndards del web semàntic, RDF, SKOS i XML.
- Semàntica: en l'opció de cerca avançada (estesa), en el camp de "matèria", apareixen diversos vocabularis controlats indexats. Això és un indicador del grau alt d'interoperabilitat semàntica, perquè coexisteixen diversos vocabularis controlats (majoritàriament encapçalaments de matèria). Conseqüentment, en el procés de recuperació de la informació, quan se selecciona un registre, també es poden visualitzar el nombre de biblioteques que disposen del document, així com el nombre de descriptors i esquemes de classificació que han utilitzat diferents biblioteques o institucions (similar als processos de cerca i indexació creuada de les passarel·les temàtiques). També hi ha indicis d'ús de navegació creuada, ja que la navegació es pot du a terme a través dels temes i esquemes de classificació utilitzats per indexar els documents recuperats que provenen de les diferents institucions que integren LIBRIS. Altres recursos externs per tal de millorar la interoperabilitat semàntica (amb l'ús d'URI) són: dbpedia.org, id.loc.gov (autoritats de la *Library of Congress*) i lcssubjects.org (encapçalaments de matèria de la Biblioteca del Congrés). Els formats i dades estructurals emprats per tal de facilitar la interoperabilitat són metadades bibliogràfiques (MARCXML, MODS, Dublin Core Elements i Dublin Core) i ontologies com FOAF (Friend of a Friend), BIBO (Biblioontology, Bibliographic Ontology) i GEO (Geonames, base de dades geogràfica gratuïta i accessible via web i integrada al web semàntic¹⁵).

En primer lloc, el portal d'informació té diversos punts d'accés a la informació, resultat d'un alt nivell d'interoperabilitat, tal i com es mostra en les diferents opcions de cerca i navegació disponibles:

- **Cerca simple.** Amb l'opció "Cerca simple", es realitza una cerca ràpida amb termes de text lliure. No obstant això, l'usuari sempre pot canviar els resultats de la cerca a través del menú de la dreta de la llista de resultats, amb l'opció "Filtra el conjunt de resultats". En aquesta opció,

¹⁵ Integració Semàntica de GeoNames: els seus recursos disposen d'identificador URI (identificador únic de recursos), descripció de dades amb els llenguatges ontològics OWL SKOS, i les dades incloses a GeoNames es relacionen amb dades Dbpedia i altres dades RDF

es poden utilitzar els qualificadors (només en suec), opcions de truncament, màscares (operadors que permeten trobar diferents derivats i ortografia), una opció de frase exacta (per exemple, "Biologia molecular") i l'ús de la cerca booleana, tot i que l'operador per defecte en la cerca és l'AND.

- **Cerca avançada** o cerca estesa. Opció de cerca per camps, com títol, autor o persona responsable pel contingut, tipus de materials, subjecte (es pot triar entre diferents vocabularis: encapçalaments de matèria suecs, MesSH, Swe.MeSH, AGROVOC , KVINNSAM, LCSH i Fills SH) o a través d'esquemes de classificació (SAB, DDC i UDK). La cerca aleshore es limitarà a aquelles parts del registre.
- **Cerca Booleana.** És un tipus de cerca avançada on s'han de formular equacions de cerca amb l'ús d'operadors booleans. És una cerca per a usuaris experimentats amb els llenguatges de cerca i la recuperació de la informació. Amb l'ajuda dels operadors i qualificadors, proporciona més flexibilitat en la formulació de les consultes de cerca.
- **Cerca per Índex** (Índex d'A-Z). És una cerca per camps dins la informació emmagatzemada. Ofereix diferents possibilitats de navegació: navegació per llistes alfabètiques d'autor, tema (títols sobre un tema o una determinada persona, entitat corporativa, lloc, etc.), títol del document (llibres, mapes, articles de revistes, etc.), títol de revista, títol de publicacions seriades, organismes (títols publicats per una determinada institució, etc.), conferències, publicacions d'acollida (articles/capítols publicats en una determinada revista o llibre, etc.) i treballs revisats.
- **Navegació per matèries** (només en suec). Permet obtenir una visió general de la literatura en un tema específic i navegar jeràrquicament en una estructura de matèries (*cross browsing* sistema).

2.2.4.5 Desplegament de serveis

A part dels serveis de cerca i navegació esmentats en l'apartat anterior, el serveis oferts són, en la gran majoria, bàsics i estàndards:

- Servei d'ajuda: amb diferents llocs web que expliquen els serveis que ofereixen als usuaris, documentació relativa i informació sobre el funcionament general i detallat del sistema (per exemple, tipus de cerques, etc.).
- Servei *about*: informació sobre com s'autodifineix LIBRIS, informació sobre la responsabilitat/autoria del sistema, els continguts, l'accessibilitat i les polítiques de privacitat i drets d'autor.

- Servei de contacte: per oferir suport i ajuda a l'usuari a través del correu electrònic, telèfon o xat.
- *Feedback*: llista de distribució oberta a tots els usuaris per rebre suggerències, comentaris, etc.
- Guies d'usuari: a través del servei "Explore", l'usuari troba informació resumida sobre el sistema, on s'expliquen les possibilitats del sistema per accedir a les col·leccions. Conté també una guia d'usuari en PDF (*Libris Quick Guide*).
- Customització: personalització d'algunes parts de la interfície que s'adapten a les necessitats dels usuaris. Per exemple, si s'escull del menú superior l'opció "Preferences", es pot canviar el contrast i la mida de les pestanyes o planes web, l'idioma (només anglès i suec), i també es poden canviar les opcions de visualització dels registres (llista estesa o resumida dels resultats, tipus d'ordre en què es volen visualitzar els resultats i nombre de registres per pàgina).
- *My Libraries*: és una opció més de personalització de les cerques. Aquesta opció permet desar en una llista de favorits les biblioteques, organitzacions o regions que l'usuari prefereixi per cercar la informació. És una opció més de filtratge de cerca o de preferències de cerca.
- Informació de novetats: interfície principal on apareix informació de les novetats bibliogràfiques disponibles i/o activitats que es realitzen. Té un format de taulell d'anuncis virtual; no existeix la possibilitat de veure l'històric d'activitats.
- Serveis de préstec (amb servei de préstec interbibliotecari inclòs): no es poden demanar en préstec llibres des de LIBRIS, però LIBRIS ajuda a trobar una biblioteca que pot prestar un llibre determinat. A través de LIBRIS també es pot veure si el document en qüestió està prestat o disponible en algunes de les biblioteques. També hi ha un enllaç directe al catàleg local de la biblioteca del document que es selecciona i on es poden reservar documents, veure els préstecs, etc. Quant al préstec interbibliotecari, una trentena de biblioteques permeten fer comandes directament als usuaris mitjançant un formulari disponible a LIBRIS (LIBRIS Customer Orders), i d'altres biblioteques faciliten informació als usuaris a través de LIBRIS, amb les instruccions sobre el préstec interbibliotecari.
- Eines diverses:
 - Crear referències. Crear una o més referències bibliogràfiques a través d'una llista de resultats des d'un interval de 10 a 100

registres; aquestes referències es poden exportar guardant-les en un arxiu o de forma manual (seleccionar, tallar i enganxar referències) des de la interfície. Es poden escollir diferents formats de referència: text pla, Harvard, Oxford, BibText, RefWorks, RIS (format d'etiquetatge bibliogràfic acceptat per sistemes de gestió bibliogràfica reconeguts com EndNote, ProCite i RefManager), APA Style, Vancouver, Bokref (referències per Wikipedia) i format MARC-XML.

- Servei de notícies via *feeds* (RSS). A través de les cerques realitzades es pot crear un canal RSS per rebre informació sobre les novetats bibliogràfiques d'una cerca determinada
- Correu electrònic. Es pot enviar un o més registres d'una llista de resultats per correu-e. A més a més, es pot enviar la mateixa informació fins a 5 usuaris diferents.
- Enllaç. Opció d'enllaç permanent (*permanent link*) amb el qual es pot accedir a un registre concret o bé a tota una llista de resultats.
- *Plugins* de cerca. Aplicació per incorporar un quadre de cerca en el lloc web o aplicació d'un usuari; se'n poden personalitzar elements de disseny, com, per exemple, el color (modificació dels fulls d'estil). Hi ha l'opció d'instal·lar un *toolbar* o quadre de cerca al navegador.
- *Search facilities*. Enllaços externs a bases de dades relacionades amb LIBRIS o bases de dades rellevants en el context regional i nacional (Suècia)

2.2.5. Cas d'estudi 5. Europeana (novembre 2008 -)

Url: <http://www.europeana.eu/>

2.2.5.1. Descripció

Europeana es va posar en marxa el 2008, amb l'objectiu de fer accessible a tot el món el patrimoni cultural i científic d'Europa.

Europeana és un projecte enorme: inclou aproximadament 2.200 organitzacions europees, i té l'objectiu principal de reunir totes les col·leccions sobre patrimoni europeu i fer-les accessibles globalment (intercanvi i reutilització de dades a través del *Linking Open Data*). Europeana ha de proporcionar accés a continguts reutilitzables i d'alta qualitat, eines i serveis que permetin la creativitat i la innovació.

És un portal d'informació format per 13 altres portals i projectes europeus que cobreixen diverses àrees temàtiques en l'àmbit del patrimoni cultural europeu.

Europeana treballa també per millorar les funcionalitats i característiques del portal per tal de cobrir les necessitats dels usuaris. Europeana està en evolució constant, un exemple n'és el canvi d'interfície el 2012; a més a més, contempla diferents versions o fases de prototipatge, per exemple, amb Europeana v1.0 (gener 2009-setembre 2011) s'inclouen millores importants en el servei com interfícies de programació d'aplicacions (API), exposicions virtuals o la millora de les capacitats de cerca del sistema. En els darrers anys, la tendència és desenvolupar serveis web per als usuaris, tal i com es mostra a Europeana v2.0 (octubre 2011- febrer 2014); s'observen esforços per millorar-ne el contingut, incrementar i facilitar la reutilització de continguts, millorar el model de prestació, desenvolupar un finançament sostenible, crear repositoris centralitzats de recursos lingüístics i noves versions, i mantenir el servei i la seva API, desenvolupar funcions i funcionalitats, millorar l'experiència de l'usuari mitjançant la distribució de continguts en entorns d'usuaris i coordinar la xarxa d'organitzacions que hi contribueixen.

El criteri mínim per ser una institució participant a Europeana és que la col·lecció o obres digitals s'exposin preferiblement en el format XML i que estiguin recollides en un servidor OAI-PMH (protocol per recollir dades), com a alternativa, les metadades poden posar-se en un servidor FTP. A més a més, els registres han de tenir un identificador persistent per tal que l'usuari final trobi recursos estables i consistents. Per organitzar les dades, Europeana utilitza un model de dades complex anomenat Europeana Data Model (EDM). Si una institució vol proporcionar dades a Europeana directament com a proveïdor de dades o indirectament a través d'un agregador, l'ideal seria que les metadades s'adaptessin al Europeana Data Model (EDM), que incorpora les metadades Dublin Core basades en l'Europeana Semantic Elements (ESE). No obstant això, s'accepten metadades originals que s'acabaran convertint a EDM, sempre que se subministrin els vocabularis adequats.

2.2.5.2. Responsabilitat

És un projecte finançat per la Comissió Europea, impulsat i coordinat per la Biblioteca Nacional dels Països Baixos, la Koninklijke Bibliotheek. A més a més hi participen unes 180 organitzacions de diversos països d'Europa (biblioteques, arxius i museus), experts en tecnologies de la informació de tota Europa i diversos projectes europeus implicats en la gestió del coneixement i el patrimoni cultural que contribueixen a millorar la tecnologia i els continguts d'Europeana.

2.2.5.3. Cobertura

El portal dona accés, a través d'una interfície multilingüe, a milions d'objectes digitals de més de 1.000 col·leccions patrimonials (24 milions d'objectes el 2012) de tots els membres de la Unió Europea, que hi contribueixen, La

cobertura geogràfica és Europa, no hi ha límits temporals i quant a la cobertura temàtica, que depèn de les institucions i col·leccions que integren el portal, s'inclouen tots els aspectes referents a la cultura i al patrimoni d'Europa.

2.2.5.4. Arquitectura de la informació

L'arquitectura de la informació d'Europeana es basa amb el moviment del *linked open data*, implementat en els diferents portals i dominis que l'integren; a més a més també es basa amb tecnologia *cloud*¹⁶. S'utilitzen identificadors web únics (URI) per identificar els objectes que formen part de les col·leccions integrades en el sistema d'informació, i aquestes són descrites de forma comuna a través del format estàndard RDF (Resource Description Format). A més, aquests recursos es vinculen a altres recursos existents al web, com ara DBpedia (projecte per la extracció de dades de Vikipèdia; és una proposta de web semàntica) o Geonames (també citat en el projecte LIBRIS). En l'actualitat, la col·lecció completa del Museu d'Amsterdam està disponible a Europeana *Linked Open Data*, amb més de 70.000 objectes del patrimoni cultural com a resultat de les dades relacionades amb el tesaurus d'art i arquitectura holandesa i (AATNed) i la Getty's Union List of Artists Names (ULAN), conjuntament amb Geonames i DBpedia.

Quant a la interoperabilitat semàntica, s'està fent un gran esforç per desenvolupar sistemes de recuperació d'informació d'alta qualitat, com ara la navegació i la cerca semàntica. El prototip està disponible en <http://eculture.cs.vu.nl/europeana/session/search>

La tecnologia d'aquest motor de cerca semàntic és compatible amb els sistemes d'organització del coneixement (sistemes de classificació i vocabularis controlats, contemplant també controls d'autoritat), igual que l'esquema de metadades utilitzat per a cada participant (biblioteques, arxius i museus). EuropeanaConnect (May 2009 - October 2011) és el grup de treball que s'ha encarregat de la creació de les capes de dades semàntiques, segons el pla de treball publicat a <http://www.europeanaconnect.eu/workplan.php>. L'objectiu d'EuropeanaConnect és proporcionar els components bàsics essencials per realitzar Europeana, és a dir, assegurar que com a biblioteca digital europea sigui un sistema i servei totalment interoperable, multilingüe i orientat a l'usuari (ciutadans europeus). Al web del grup de treball, a la secció "Results and resources" es poden trobar les característiques d'aquest sistema.

En termes d'ingestió de dades i flux de treball de processament de la informació per a Europeana implica els passos següents:

¹⁶ Cloud Computing: les dades i equips ja no depenen dels sistemes operatius sinò de la xarxa, del núvol o cloud.

- *Skosificació*. Procés en el qual s'han de transformar les entitats de fitxers d'autoritats en recursos per al web semàntic, expressats en termes dels SKOS estàndard; per tant, han de tenir una URI (cf. Miles & Bechhofer (2009) i redirigir enllaços a aquestes URI, per ser utilitzats posteriorment com a dades obertes enllaçats (LOD).
- *Matching* (coincidència). Els recursos de contextualització semàntics subministrats (LOD o fitxers d'autoritat) molts cops són redundants en diferents proveïdors de dades (persones idèntiques o recursos de conceptes diversos utilitzats en els àmbits de treball respectius). Aquests casos han de ser detectats de manera sistemàtica, idealment, per tal de reunir totes les entitats a les quals pertany el recurs d'un concepte determinat.
- *Mapping/Merging* (mapeig/fusió). Basat en operacions de mapeig de recursos d'un concepte determinat que pot subseqüentment fusionar (en cas que es puguin controlar tots els recursos necessaris). El resultat dóna lloc a una nova entitat SKOS amb un terme preferit, amb enllaços al concepte antic (ara fusionat); les entitats SKOS hauran de ser redirigides. El cas contrari són les dades LOD, en les quals Europeana, per definició, no té control; les entitats fusionades s'establirien i s'implementarien de forma que s'obtidria un resultat funcionalment similar a la realitat dels recursos fusionats.
- Contextualització automàtica per a les representacions d'objectes. S'utilitza quan les metadades proporcionades pels proveïdors no estan prou contextualitzades per encaixar-les amb el model funcional *Europeana Data Model* (EDM). Aquestes metadades s'han de contextualitzar mitjançant processos automàtics, amb la creació d'enllaços a recursos que ho puguin fer. Per aquest motiu, els valors literals d'atributs poden utilitzar-se en molts casos, i aquests poden ser mapejats amb èxit per elements SKOS existents, com valors *prefLabel*. En la contextualització també se solen aplicar algorismes basats en coocurrència. L'objectiu final és crear un context semàntic relativament homogeni per a les representacions d'objectes inclosos a Europeana, i crear, també, mitjans per posicionar automàticament representacions d'objectes en aquest context. Integració de dades obertes vinculades (*linked open data*). Amb totes les casuístiques presentades ja es mostra un context molt ambiciós i complex, però amb *linked open data* la situació és encara més complexa, ja que LOD creix massivament. En aquesta fase, l'objectiu és integrar les capes de dades que proporciona el context semàntic de les representacions d'objectes d'Europeana, per tal que encaixin en un model d'arquitectura LOD. Es renuncia a una certa autonomia (o una part del control de les dades), però es guanya en funcionalitat i riquesa del context de les dades/objectes i, a més, es

construeix un sistema d'interoperabilitat que fins ara no havia existit arreu del món. Evidentment hi ha problemes tècnics, com la escalabilitat, que s'hauran d'anar solucionant.

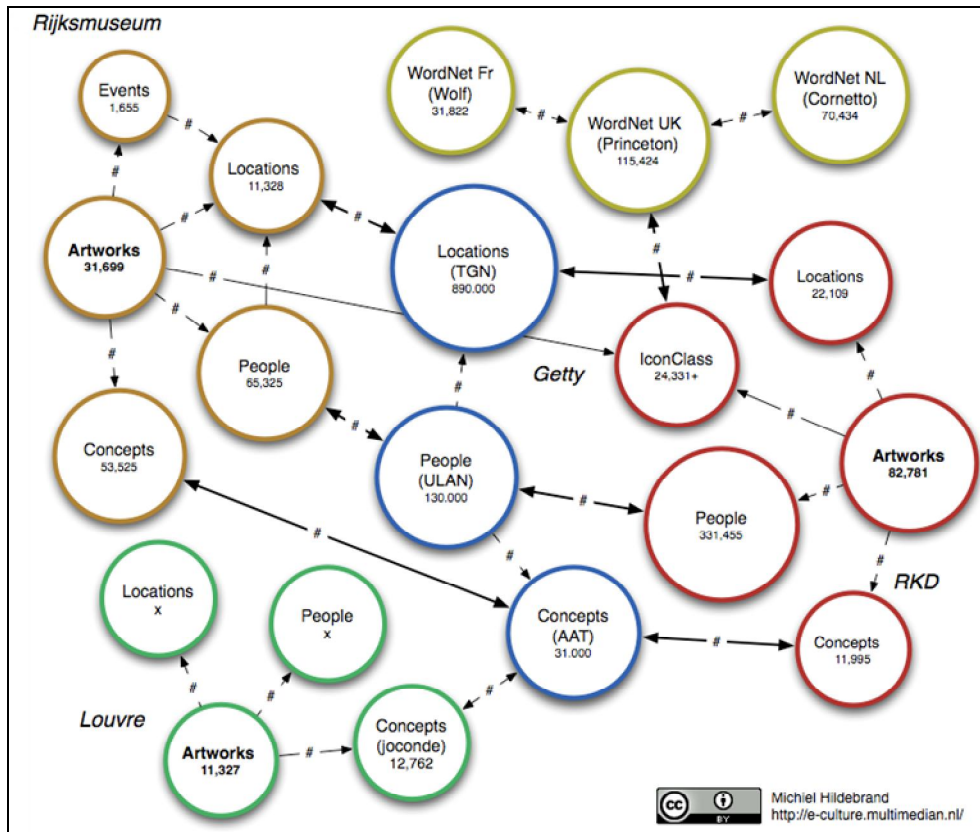


Figura 9. Arquitectura Cloud (Font: Europeana Thought Lab)

Quant a sistema i sintaxi, no se'n coneixen gaires detalls, només se sap que s'utilitza el protocol OAI-PMH per a la transmissió de dades d'Internet. Aquest protocol té una arquitectura basada en el model client-servidor, en el qual els arxius (client) es posen a disposició al públic en format Dublin Core per tal que puguin ser recuperats pel servidor. La comunicació es realitza mitjançant el protocol http i les respostes es codifiquen en format XML.

Pel que fa a l'estructura, l'arquitectura d'Europeana es basa plenament en els estàndards del W3C, i tota la informació disponible està accessible via tripletes RDF. També s'usen la resta de formats estàndards del web semàntic: XML, OWL, SKOS i SARQL.

2.2.5.5. Desplegament de Serveis

Aquesta fase encara s'està desenvolupant actualment, de fet va començar a desplegar-se amb la versió 2.0 d'Europeana a l'octubre del 2011 i es preveu acabar-la el febrer del 2014. Els serveis actius que es poden trobar a d'Europeana són:

- Sistema de cerca simple i avançada. Fins al 2011 aquestes opcions estan separades, però actualment estan integrades en una sola interfície de cerca. A la pàgina principal hi ha un motor de cerca simple, però amb les funcionalitats d'un sistema de cerca avançat, com ara les cerques per camps (vegeu Annex 2.8). Abans de posar un terme de cerca a la interfície, a través de l'opció "Search" es pot seleccionar prèviament el camp on realitzar les cerques: tots els camps, camp títol, creador, matèria, períodes/dates i llocs. Després, quan s'introdueix el terme de cerca, el motor té una funcionalitat d'autorealització (Europeana suggereix termes a l'inici de la cerca, vegeu Annex 2.8), que a la vegada permet fer un primer filtre pels camps d'índexs (títols, matèries, creadors, períodes, etc.). El motor de cerca permet cercar per frases ("ús de cometes altes", l'ús del operador booleà AND és implícit) i fer una cerca booleana d'exclusió de paraules (operador NOT). Finalment, es mostren els resultats i, al mateix temps, apareix un menú a l'esquerra de la interfície amb diferents opcions de filtre de cerca: afegir més paraules clau, seleccionar tipus de mitjà/suport (imatge, so, text i vídeo), llengua de descripció, any, país proveïdor, *copyright*, proveïdor i incloure o excloure d'altres continguts proporcionats pels usuaris. A més a més, s'inclou una opció per compartir els resultats, "Share", a través de diferents recursos (Facebook, Twitter, correu electrònic, Print, *Gmail*, Stumble upon, *favourites*, entre altres eines web 2.0).
- Cerca semàntica. De moment, el prototip de cerca semàntica dóna accés a tres col·leccions: Rijksmuseum, RKD i Louvre, i a tres idiomes (anglès, francès i holandès). El cercador integra diversos vocabularis controlats, com Joconde, IconClass, RKD, AAT, ULAN, TGN i WordNet, una gran base de dades lèxica amb anglès i Cornetto, una base de dades semàntica holandesa. El funcionament d'aquest sistema és el mapeig de vocabularis controlats, la tecnologia de mapeig fa que s'obtinguin alineacions entre diferents sistemes d'organització, així l'usuari accedeix a diferents col·leccions a la mateixa interfície de forma simultània i de forma interrelacionada (enriquiment de les col·leccions: els recursos d'informació poden estar relacionats amb altres recursos externs del web). El motor de cerca, en fase d'experimentació, funciona com el cercador simple de la pàgina principal, però aporta semàntica, és a dir, ús de dades enllaçades. Aquesta funcionalitat s'aplica ja quan s'introdueix el terme de cerca en el motor de cerca. Per exemple, si s'hi introdueix el terme 'Espanya', abans de mostrar els resultats, apareixen diferents conceptes vinculats amb el terme. Per exemple, el terme apareix com a part d'un títol d'un objecte, associat a una regió o concepte determinat, com per exemple 'castells d'Espanya', o bé el concepte pot estar vinculat al nom d'una persona; totes aquestes opcions permeten realitzar la cerca per filtrar els resultats de la forma

més òptima possible (Annex 2.8). No obstant això, el sistema encara està desenvolupant-se i, com ja s'ha esmentat, està limitat a les dades de les tres col·leccions: Rijksmuseum Amsterdam, el Museu del Louvre de París, i la Rijksbureau voor Kunsthistorische Documentatie - Institut Holandès d'Història de l'Art, a l'Haia.

- Navegació (*Exploring Europeana*). Si se selecciona l'enllaç 'Exposicions' a peu de pàgina, es pot navegar per les exposicions virtuals disponibles, organitzades temàticament (*Buildings, Treasures, Expeditions, Travelling Through History*, etc.); s'han de seleccionar individualment per accedir als continguts. Un cop seleccionat un tema, es pot navegar a través dels continguts específics de la col·lecció; tot i que cada col·lecció té diferents possibilitats de navegació, generalment es pot navegar per temes o subtemes, períodes, països i proveïdors.
- Servei beta de carrega d'arxius a Wikimedia Commons (base de dades de 18.119.409 arxius multimèdia de lliure ús). En el moment que es va estudiar aquest servei no era accessible, es coneix que va començar a implementar-se a començar el setembre del 2013 i finalitzar a final del 2013.
- Serveis web 2.0 disponibles. Comptes a Facebook, Twitter, Pinterest, Google+.
- Newsletter. Subscripció per correu electrònic de les novetats a Europeana amb una regularitat bimensual.
- About us. Secció que inclou informació sobre Europeana; a més a més, ofereix serveis relacionats amb informació sobre Europeana.
 - *New content*. Subscripció a canals RSS que proporcionen a l'usuari informació sobre continguts nous publicats pels col·laboradors d'Europeana (museus, arxius, biblioteques i col·leccions audiovisuals). Hi ha l'opció de subscripció al canal/s d'una de les institucions participants. Url: <http://europeana.eu/portal/newcontent.html>.
 - *All providers*. Llista d'institucions proveïdores d'objectes digitals a Europeana, amb dades sobre el nombre d'ítems accessible i enllaç directe a les col·leccions a Europeana. Url: <http://europeana.eu/portal/europeana-providers.html>.
 - *Press center*. Secció sobre mitjans de comunicació sobre Europeana.eu, la biblioteca digital europea, museu i arxiu. Informació disponible sobre comunicats de premsa, imatges de premsa, logos i directrius. Hi ha llistes de distribució per estar

actualitzat sobre tot el que es publica sobre Europeana (Twitter, Blog, butlletí de notícies bimensual eNews i notes de premsa via canal RSS).

- *Jobs*. Informació relacionada sobre ofertes de treball disponibles, tipus de contractacions o altres tipus de relacions laborals (estades de pràctiques, etc.) en relació a Europeana. També hi ha informació sobre el pla estratègic (2011-2015). Url: http://www.pro.europeana.eu/about/jobs?utm_source=portalmenu&utm_medium=portal&utm_campaign=Portal%2Bmenu.
- *Europeana Sites*:
 - Exhibitions. El servei d'exposicions virtuals són aparadors de continguts disponibles a Europeana. La informació està en anglès però disposa de traduccions a d'altres idiomes.
 - Blog. Inclou informació sobre arxius de contingut obert d'Europeana, informació d'esdeveniments i altres temes amb els quals s'està treballant. Els blogs proporcionen *feedback* amb l'usuari i també ofereixen la possibilitat que usuaris externs participin amb comentaris sobre els continguts que es hi ha al blog.
 - 1914-1918. És una secció que intenta recollir tot el material disponible a Europa sobre la Primera Guerra Mundial: fotografies, cartes, postals, records o articles publicats entre 1914-1918 sobre la Primera Guerra Mundial. A més a més, els usuaris poden participar-hi de forma oral o escrita a través d'una col·lecció de vídeos amb relats de persones que tenen alguna història o anècdota per explicar sobre les persones involucrades o afectades; s'anomenen "contes en línea".
 - Remix. És una plataforma interactiva basada en la pel·lícula *Otto & Bernard*. Aquest curtmetratge va ser produït pel projecte Europeana Erster Weltkrieg; es van encoratjar les famílies a compartir fotos, cartes i records de la Primera Guerra Mundial. La iniciativa Remix combina la tecnologia d'avantguarda (Popcorn.js, i HTML5 vídeo framework) i una varietat de recursos d'Europeana a través del web, per tal que els usuaris externs puguin compartir a través d'aquesta plataforma idees, històries, comentaris i enllaços.
 - Professional. Plataforma professional on comparteixen experiències i coneixements professionals bibliotecaris, arxivers i conservadors, etc.

- *Help service*. Llocs es web amb informació sobre diferents seccions de cerca i navegació d'Europeana (*Searching Europeana*, *Exploring Europeana* i *Results in Europeana*), a més a més hi ha l'espai de personalització de l'usuari "Using My Europeana" (registre a través de correu electrònic i accés per guardar cerques, etiquetar favorits i ús d'API com a desenvolupadors) i la secció de "What's new" (plana web amb informació de les darreres novetats sobre les aplicacions d'Europeana que poden afectar els usuaris, com per exemple, canvis en la interfície, etc.).
- Serveis futurs. Es treballa projectes per crear serveis on l'usuari pugui crear el seu propi contingut (UGC)¹⁷ i eines orientades cap a la generació de continguts per a l'usuari (*UGC-oriented tools*).

2.3. Comparativa entre els diferents sistemes d'informació de qualitat

Aquesta secció es divideix en dos apartats: l'estudi de les passarel·les temàtiques com a sistemes distribuïts de qualitat i l'estudi dels sistemes d'informació semàntics com a evolució de sistemes d'informació distribuïts i models futurs de sistemes d'informació de qualitat. Finalment, a partir de l'anàlisi de 5 estudis representatius, es durà a terme una comparativa entre els dos tipus de sistemes. Aquests sistemes són els ja descrits en l'apartat anterior: Intute, Vascoda, AustLII, LIBRIS i Europeana

En l'anàlisi comparatiu dels sistemes d'informació exposats se segueix la metodologia d'anàlisi qualitativa dels sistemes (ús de fitxes descriptives i tècniques quantitatives per qualificar i objectivitzar els resultats), des d'on s'examinen les dimensions de sistema, serveis i col·lecció. Quant a la col·lecció, no se n'ha avaluat la mesura o la mesura ideal, ja que els casos presentats tenen cobertures temàtiques diferents.

La metodologia que s'ha seguit és l'elaboració d'una fitxa descriptiva per a cada sistema d'informació seleccionat, on es recullen les dades principals que afecten els serveis que ofereix cadascun dels sistemes analitzats i es diferencia entre serveis bàsics i addicionals. A posteriori s'avalua la col·lecció i el sistema dels casos presentats, a través d'exemples d'interrogació i navegació dels sistemes. No s'han realitzat estudis d'usuaris perquè no eren objecte d'estudi i el fet que cada sistema té audiències diferents fa inviable l'estudi. Així doncs, per complementar l'avaluació dels estudis de casos i aportar objectivitat a l'estudi, s'han elaborat qüestionaris (Annex 2.3) amb preguntes que descriuen i defineixen cadascuna de les dimensions avaluades (sistema, serveis i col·lecció).

¹⁷ UGC: contingut generat per l'usuari.

2.3.1. Comparativa entre sistemes distribuïts de tipus passarel·la temàtica

2.3.1.1. Comparativa de serveis: Intute, Vascoda i AustLII En general, Intute és el sistema més complert quant a serveis. No només contempla tots els serveis bàsics¹⁸ que ha de tenir qualsevol sistema d'informació, sinó que també inclou molts més serveis addicionals. Això es deu principalment, d'una banda, a la seva trajectòria, ja que Intute està construïda a partir d'una de les passarel·les temàtiques més antigues, SOSIG (1998?). Aquesta característica fa que hagin pogut desenvolupar un major nombre de serveis al llarg dels anys. D'altra banda, també és un sistema basat en unes polítiques de col·laboració i cooperació sòlides.

AustLII, econòmicament, gaudeix d'un menor suport institucional que la resta de passarel·les. Els membres hi col·laboren de forma totalment voluntària, per tant, el desenvolupament de serveis és molt més lent i es prioritza l'accés a la col·lecció davant d'altres tipus de serveis.

Pel que fa a Vascoda, tot i que s'aproxima molt més al model d'Intute que al d'AustLII (polítiques, sistema, IQ, etc.), no està tan desenvolupat ni pel que fa als serveis ni pel que fa al sistema, ja que és un projecte més recent que la resta. No obstant això, Intute i Vascoda —abans del seu tancament— evolucionaven cap al web semàntic, o almenys treballaven per a la introducció de vocabularis controlats adaptats als estàndards del web semàntic en els sistemes de cerca i navegació.

A continuació es presenta la fitxa descriptiva de serveis oferts per les passarel·les temàtiques seleccionades.

SERVEIS¹⁹	INTUTE	VASCODA	AUSTLII
Tipologia Serveis	Avançats	Bàsics	Bàsics
Cerca Simple	No serveix per realitzar equacions de cerca complexes, tot i que admet operadors boolens, de	No serveix per equacions de cerca complexes (no admet més de 20 paraules de cerca). Sí admet operadors boolens, de	No serveix per equacions de cerca complexes, tot i que admet operadors boolens, de proximitat, etc.

¹⁸ Serveis mínims: s'interpreten serveis mínims els motors de cerca i navegació, serveis de notícies i arxiu (sigui quin sigui el seu mitjà o suport), serveis que permetin guardar continguts/cerques, exportar resultats, etc., i servei d'alertes sobre el contingut (independentment del mitjà i suport: e-mail, RSS, etc.), Serveis d'ajuda i contacte, i mapa del web (valor d'accessibilitat).

¹⁹ Puntuació: en alguns casos el camp avaluat es valora segons la puntuació Alta (inclou elements mínims/imprescindibles i elements addicionals), Mitja (inclou els elements mínims, però hi ha mancances importants o mancances d'elements mínims) i Baixa (falten elements mínims imprescindibles, per exemple: tenir un mapa web per tal de donar una major navegabilitat web, etc.).

	proximitat, etc.	proximitat, etc.	
Cerca Avançada	Es poden desenvolupar equacions de cerca molt extenses amb una sola casella de cerca. Límits per temes, tipologia documental i país. Ordre per rellevància o alfabètic.	Es poden fer combinacions però no suficientment complexes (ús de 4 nivells de cerca). Com a valor afegit, t'extreu sinònims de paraules clau utilitzades (tesaurus)	Opció a multicerca de bases de dades, però limitant la equació de cerca a un sol camp. Aquest camp a més obliga a l'usuari a incloure/excloure el tipus d'equació (p.ex. booleana, per frase, per títol, etc.)
Navegació (web)	Sí (Alta)	Sí (Alta)	Sí (Mitja-Baixa: no hi ha un mapa del web per navegar)
Navegació (referent a una estructura de navegació per matèries: encapçalaments i subencapçalaments)	Sí, Navegació Alta i flexible (ús de diferents Vocabularis Controlats, i combinació de cerca/filtre per metadades). Estructura jeràrquica de navegació per matèries de 2 nivells	Sí, (En vies de desenvolupament : navegabilitat Mitja, la navegació per una estructura temàtica amb desplegament de jerarquia, està accessible només a la cerca/mòdul de <i>Subject Gateways</i> , hi ha pendent l'aplicació de l'estructura de classificació DDC en la indexació i	No, Navegabilitat Mitja-Baixa: és limitada i insuficient (navegació per proveïdors o bases de dades, la navegació temàtica només existeix a la cerca del catàleg WorldLII ²⁰ , i té 2 nivells de navegació)

²⁰ WorldLII: Catàleg mundial que enllaça a més de 15.000 llocs web relacionats amb les lleis de tots els països del món. El WebSearch de WorldLII fa fer cerques al text complet a la major part d'aquests llocs amb un sistema de *web-spider*.

		navegació)	
Indexació	Alta: Metadades i vocabularis controlats (tesaurus, llistes d'encapçalaments, etc.)	Bàsica: Metadades (pendent ús de sistema de Classificació Decimal Dewey)	Bàsica: Metadades
Serveis per descarregar i emmagatzematge de cerques	Sí	No (disponible d'accés privat)	No
<i>Help/FAQ</i>	No visible a la pàgina principal	Visible a la pàgina principal	Visible a la pàgina principal
<i>SiteMap</i>	Sí	No n'hi ha	No n'hi ha
<i>Contact/Feedback</i>	Sí (pàgina principal)	Sí (pàgina principal)	Sí (pàgina principal)
<i>About</i>	Pàgina principal	Pàgina principal	
Servei històric de notícies d'interès	Sí	Sí	Sí
Guies d'usuari/manuals, etc.	Sí (textuals i tutorials virtuals)	Sí (bàsics: textuals)	Sí (bàsics: textuals)
Serveis addicionals			
MyVascoda / MyServices (servei de personalització)	Públic: crear alertes, guardar cerques, classificar continguts, valorar-los i fer comentaris	Privat: subscripció limitada a les institucions col·laboradores o d'alemanya (no hi ha disponibles les funcionalitats)	No existeix
Altres serveis de personalització	Sí	No	No

Serveis d'instal·lació (<i>Search-Plugin, Search field, toolbar</i>)	Sí	Sí	Sí
Serveis API (<i>Application Programming Interface</i>)	No	No	Sí
Blog	Sí	Sí	No
<i>Newsletter</i>	Sí	Sí	Sí (només accés online o subscripció e-mail)
<i>Twitter</i>	Sí	Sí	Sí
<i>RSS-Feed</i>	Sí (en els blogs i canals temàtics RSS específics en el portal web)	Sí (només a nivell de blog)	No
<i>FriendFeed</i> (servei web 2.0 que permet descobrir i intercanviar opinions sobre el material d'interessant sobre que els teus amics troben a la xarxa)	No	Si	No
Serveis de Seminaris i esdeveniments	Sí	No	Si
Serveis de col·laboració per part dels usuaris a nivell de col·lecció (etiquetatge social, etc.)	Sí (només a nivell d'aportacions de suggeriments de pàgines web, etc.)	No	No (però estan treballant)

Taula 3. Fitxa descriptiva: Intute, Vascosa i AustLII

2.3.1.2. Comparativa dels sistemes a través de la RI

Pel que a la recuperació de la informació, els tres sistemes tenen sistemes de cerca i navegació diferents, i, per tant, el funcionament també varia.

Tant Intute com Vascode contemplen sistemes de cerca i navegació creuats en l'arquitectura de la informació. L'arquitectura de la informació d'Intute, quant a la RI, es basa en la cerca creuada, és a dir, el mapeig de metadades de descripció de recursos (MARC21 i DC principalment). La navegació i indexació creuades es realitzen a través de vocabularis controlats (establiment de concordances creuades o mapeig de vocabularis). Aquestes característiques fan que la RI sigui més precisa, i també faciliten l'organització del coneixement i navegabilitat web de l'usuari i en general fan tot el sistema d'informació sigui més fàcil d'utilitzar. En l'Annex 2.4 (cas Intute) es pot veure ja a la interfície principal hi ha diferents punts d'accés a la informació (cerca simple, cerca avançada, navegació per matèries, etc.). En la interfície anterior d'Intute, hi havia també l'opció de navegació per matèries de forma alfabètica ("Subject A-Z", etc.). De fet, de tots els casos estudiats, Intute és el sistema d'informació que presenta l'organització del coneixement amb un major desenvolupament, ja que disposa de diferents opcions de cerca i navegació no contemplades en la resta de sistemes, o bé contemplades en un menor desplegament (per exemple, l'estructura jeràrquica de navegació).

La cerca creuada de Vascode es realitza a través de mapeig de metadades, amb Dublin Core com a estàndard de metadades central. La navegació creuada encara no s'aplicava, però sí que se n'havia previst el desenvolupament mitjançant l'ús de la classificació decimal de Dewey (DDC). No obstant això, s'havien establert ja 18 mapejos de concordances creuades per representar un model de vocabulari que pogués establir connexions semàntiques genèriques entre les terminologies del sistema. També es contemplava l'ús de mapejos creuats de diferents tipus de vocabularis controlats (tesaurus, classificacions, etc.) adaptats al web semàntic, que permetien cercar relacions entre els termes i enriquir així la indexació dels continguts de la col·lecció, però tampoc es va arribar a desenvolupar (vegeu Annex 2.5).

Intute és més complet pel que fa a l'arquitectura de la informació que Vascode a causa d'una major precisió de la RI; això suposa un augment en la rellevància i en la pertinença dels resultats. D'una banda, les equacions de cerca (Annex 2.5.1) mostren com Intute permet un major nombre de paraules clau de cerca en les interfícies de cerca que Vascode. D'altra banda, en els resultats d'una cerca similar (Annex 2.5.2), es mostra una major cobertura en Intute que en Vascode, on no s'han recuperat resultats. Aquesta diferència

també es deu a l'ús de vocabularis controlats en la cerca i en la navegació per part d'Intute enfront Vascoda. Quant a la cobertura temàtica, cal afegir que també hi ha factors externs que poden influir-hi, com que Intute és un sistema més antic que Vascoda, i, per tant, les mides de les col·leccions no són equiparables.

Contràriament, AusLII segueix un model de sistema d'informació distribuït tradicional amb ús de protocols d'intercanvi d'informació i tecnologia de cerca més estàtica. Tot i que utilitza processos de *crosswalk* (mapeig de metadades), no utilitza metadades estàndards i la representació de la informació és pobre —no està detallada suficientment—, i això es manifesta en la interfície de resultats, on només es pot visualitzar el títol i l'autoria/responsable del recurs (Annex 2.6.3). Aquest sistema mostra aspectes negatius tant en IQ com en aspectes relacionats amb la satisfacció de l'usuari en els processos de recuperació de la informació (principalment sobre la rellevància i la precisió dels resultats). A més a més, la informació no es relaciona entre si a través del llenguatge, la semàntica o els vocabularis controlats, per tant, la recuperació de la informació és de nivell més baix tant en la interacció sistema-usuari com en l'obtenció de resultats precisos i rellevants per a l'usuari. Les úniques relacions que s'estableixen entre un recurs i la col·lecció són a través de la responsabilitat del recurs, és a dir, a través de l'autoria d'un registre si es poden consultar altres documents que fan referència a aquella autoria. Pel que fa a la recuperació de la informació, es basa principalment en algorismes de cerca i recuperació de la informació tradicionals, com els de ponderació, o de cerca de concurrència de paraules o frases que apareixen al contingut (vegeu documentació del sistema de cerca SINO), i cal afegir que tampoc s'integren vocabularis controlats per indexar o classificar el contingut per tal de millorar la RI.

En l'apartat següent, d'avaluació dels sistemes d'informació des de la perspectiva de la informació de qualitat i els qüestionaris relacionats, s'ampliarà la informació sobre els sistemes de recuperació de la informació de cadascun dels casos d'estudi avaluats.

2.3.2. Comparativa entre sistemes semàntics

En aquest apartat s'inclou el sistema d'informació LIBRIS, que és un catàleg col·lectiu nacional d'universitats i centres de recerca de Suècia, que aplica tecnologia semàntica per facilitar l'accés a diferents bases de dades o catàlegs distribuïts. I Europeana, portal d'objectes digitals o —com també tradicionalment s'anomena— d'informació i documentació (inclou col·leccions de so, vídeo, text, imatge, etc.) sobre patrimoni cultural i científic d'Europa.

2.3.2.1. Comparativa a nivell de serveis

En general, pel que fa als serveis, els dos sistemes no estan tan desenvolupats com les passarel·les temàtiques, ja que l'objectiu principal era desenvolupar l'arquitectura de la informació semàntica per tal de millorar l'intercanvi de dades i la interoperabilitat dels sistemes que formen part del sistema d'informació distribuït. No obstant això, els serveis bàsics o estàndards cobreixen les expectatives de tots els sistemes d'informació de qualitat. Principalment, hi ha els serveis centrats en l'usuari per tal que pugui interactuar amb el sistema (serveis de cerca, d'ajuda, *feedback*, serveis socials o 2.0, etc.). Pel que fa a xarxes socials, Europeana té un desenvolupament molt més estès que cap altre sistema estudiat, per contra, LIBRIS no està integrat a les xarxes socials.

A diferència de la resta de sistemes de passarel·la temàtica, els tipus de col·leccions són molt diverses, ja que les passarel·les no només feien referència a col·leccions dels proveïdors d'informació sinó que l'objectiu era descobrir nous continguts *online* i proveir el sistema d'informació de qualitat per tal d'augmentar les col·leccions. En canvi, els sistemes semàntics presentats basen les col·leccions en col·leccions existents dins d'institucions molt concretes (biblioteques, arxius i museus).

Els serveis futurs previstos en els sistemes semàntics se centren en la participació de l'usuari en la millora i en l'enriquiment de les col·leccions. És a dir, que pugui aportar-hi els seus coneixements o bé en la indexació de la col·lecció (per exemple, en l'etiquetatge social) o bé ajudant a entrellaçar les dades, és a dir, que es puguin relacionar diferents col·leccions de diferents proveïdors d'informació disponibles a la xarxa amb dades significatives del documents continguts (autoritats, matèries, dates, llocs, etc.) i fomentar, així, el moviment *Linked Open Data*. A continuació es presenta la fitxa descriptiva de serveis oferts.

SERVEIS ²¹	LIBRIS	EUROPEANA
Cerca Simple	Sí	Sí
Cerca Avançada	Sí	Sí
Navegació (web)	Sí	Sí
Navegació (referent a una estructura de navegació per matèries: encapçalaments i subencapçalaments)	Sí (només en suec)	No

²¹Puntuació: en alguns casos el camp avaluat es valora segons la puntuació Alta (inclou elements mínims/imprescindibles i elements addicionals), Mitja (inclou els elements mínims, però hi ha mancances importants o mancances d'elements mínims) i Baixa (falten elements mínims imprescindibles, per exemple: tenir un mapa web per tal de donar una major navegabilitat web, etc.).

Indexació (vocabularis controlats)	Sí	Sí
Serveis per descarrega i emmagatzematge de cerques	Sí	No (però permet fer referències bibliogràfiques)
Servei d'Ajuda (Help/FAQ)	Sí	Sí
SiteMap	No	Sí
Contact/Feedback	Sí	Sí
About	Sí	Sí
Servei històric de notícies d'interès	No	Sí
Guies d'usuari/manuals, etc.	Sí	Sí
Serveis Addicionals		
Serveis de Cerca Semàntics	Sí	Sí (una part de la col·lecció)
MyLibraries (servei de personalització de cerques, etc.)	Sí	Sí
Altres serveis de personalització	Sí	No (a excepció de l'idioma)
Serveis d'instal·lació (Search-Plugin, Search field, toolbar, etc.)	Sí	No
Serveis API(Application Programming Interface)	No	Sí
Serveis per dispositiu Mòvil (visualització, etc.)	No	Sí
Blog	No	Sí
Newsletter	No	Sí
Twitter	No	Sí

RSS-Feed	Sí	Sí
Eines 2.0 o Serveis 2.0 (a part de RSS)	No	Sí
Serveis de Seminaris, esdeveniments, novetats, etc.	Sí	Sí
Serveis de col·laboració per part dels usuaris a nivell de millorar la col·lecció (etiquetatge social, associació amb altra informació relacionada a la xarxa, etc.)	No	No (però hi estan treballant)
Altres serveis	No	Sí (col·leccions específiques creades pels usuaris)

Taula 4. Fitxa descriptiva: Libris i Europeana

2.3.2.2. Comparativa dels sistemes a través de la RI

Són dos sistemes d'informació completament diferents, ja que tenen objectius, audiències i cobertura molt diversos, i això es reflecteix tant en els serveis com en la recuperació de la informació. D'entrada, cal destacar que Europeana és un projecte molt més gran, ja que la cobertura és Europa, en canvi, LIBRIS té una cobertura nacional (Suècia). Els objectius també són molt diferents: LIBRIS funciona com un catàleg col·lectiu de biblioteques i centres de recerca; ells es descriuen com un servei de cerca nacional que proporciona informació sobre els títols existents en les biblioteques de recerca i universitàries de Suècia, etc. I Europeana és un portal d'informació sobre patrimoni cultural d'Europa, format per institucions líders a europees, com biblioteques, arxius, museus i galeries, i contempla diferents tipus d'objectes continguts en aquestes institucions (imatges, text, so i vídeo).

No obstant això, els dos sistemes apliquen tecnologies semàntiques. Per tant, en aquest apartat es volen mostrar només les característiques d'aquests tipus de sistemes, com el funcionament d'aquesta tecnologia de tendència i els avantatges i inconvenients enfront la resta de sistemes analitzats.

La diferència principal que hi ha entre els dos sistemes és que LIBRIS ofereix moltes més opcions de cerca a l'usuari que Europeana, atès que Europeana només ofereix accés a una cerca simple i a un prototipus de cerca semàntica. Aquesta característica fa que la cerca de l'usuari no estigui tan guiada i limita les possibilitats de cerca de l'usuari; això no obstant, quan es realitza la cerca apareix una opció d'autocompletar amb una llista de termes relacionats amb el

concepte de cerca; en aquesta llista es visualitzen els termes en idiomes diferents (cerca multilingüe) i també apareix al costat del terme el lloc del document on apareix la paraula de cerca (títol, autor, matèria, llocs, etc.). Així doncs, es mostra clarament que són termes indexats en la base de dades (vegeu Figura 9). En canvi, la funció d'autocompletar és insuficient perquè mostra ambigüitat semàntica. Per exemple, amb el terme de cerca en anglès *plant* (vegeu Figura 9), el cercador dona diverses opcions, però no sempre s'identifica fàcilment a què pot correspondre el concepte: *plant* (fàbrica), *plant* (maquinària), *plant* (vegetal), *plant* (agricultura, per exemple: plantar una llavor).

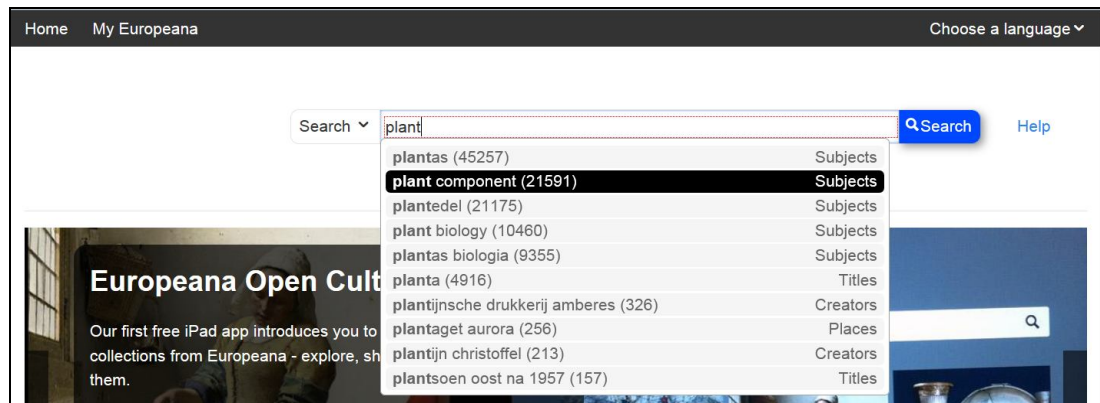


Figura 10. Exemple de cerca simple (Font: Europeana)

En el prototipus de cerca semàntica, en fase beta des del 2009, es millora la consulta de la cerca perquè el terme està associat a un context o descriptor que el defineix. D'una banda, el terme s'agrupa en diferents tipus de descriptors que representen "tipus de dades" (esdeveniments, persones, localització, conceptes, artefactes, etc.). D'altra banda, si se selecciona l'agrupació de "conceptes" els termes s'associen amb facetes²² per tal d'evitar l'ambigüitat que sorgeix en el cercador no semàntic; per exemple, si se cerca el terme *plant* apareixen associats els següents termes, *plant* (*flora*), *plant* (*building*) o *plant* (*lay*).

²² Descriptors que agrupen els termes d'indexació per camps (temes) o classes de termes.



Figura 11. Exemple de cerca semàntica (Font: Europeana)

Tot i que la cerca semàntica millora la cerca simple encara hi ha un cert nivell d'ambigüitat que cal resoldre. Per exemple, en la resta d'agrupacions (*other*, *persin event*) no hi ha prou informació per a l'usuari per poder seleccionar un terme o un altre, ja que no s'identifica el tipus de relació que hi ha amb el terme (vegeu-ne més exemples a l'Annex 2.8). Tampoc hi ha la possibilitat de navegar per les matèries (o vocabularis controlats associats), com sí que es pot fer amb LIBRIS. No obstant això, s'està treballant per desenvolupar noves tecnologies per millorar la cerca i la navegació; es poden consultar els prototipus a Thought Lab.

Finalment, cal destacar que els sistemes mostren la interoperabilitat que hi ha entre diferents col·leccions i llenguatges documentals associats (tesaurus, sistemes de classificació, encapçalament de matèries, etc.), ja que funcionen amb un model similar d'interoperabilitat, establint d'equivalències o mapeig entre diferents vocabularis controlats i multilingües (amb diferents estructures, llengües, nivells de profunditat, etc.), que permeten accedir de forma simultània a diverses col·leccions. Aquest procés s'observa en dels registres resultants de

la cerca (vegeu Annex 2.8.3), on es permet la navegació de diferents camps en les col·leccions que provenen de diferents organitzacions i sistemes (proveïdors d'informació, tipus documental, llengua, any, matèria, autor, etc.), i les relacions d'alguns camps claus (autor, data, matèries) amb dades externes. Per exemple, en el camp "Matèries" (Annex 2.8.3) hi ha un enllaç (amb identificador únic) cap a un vocabulari controlat (per exemple, GEMET) d'on s'ha extret el descriptor utilitzat en la indexació del registre. A més, els sistemes semàntics permeten establir relacions entre conceptes i dominis diferents, per tant, augmenta l'exhaustivitat de la cerca. Com s'observa en la Figura 11, el terme *plant* pot estar associat a un esdeveniment, a una persona, a un concepte, etc. Alhora, el sistema permet realitzar cerques precises; en la Figura 11 el concepte *plant* està associat a àmbits que indiquen que el concepte "planta" pot referir-se a un vegetal o a un edifici entre d'altres conceptes.

Tant les passarel·les temàtiques com els sistemes semàntics utilitzen la metodologia de mapeig o d'establiment de concordances entre els termes que provenen d'una cerca (terme de cerca d'un usuari) i els termes que provenen dels vocabularis controlats (multilingües, en la majoria dels casos), on s'estableixen relacions d'equivalència (mapeig). La diferència principal dels sistemes semàntics és l'ús d'ontologies com a vocabularis controlats i alhora ús de tecnologia *Linked Data*, que permet establir o representar un conjunt de conceptes (ontologies) en un domini i les relacions entre els conceptes (esdeveniments, agents, lloc/localització, objecte físic, concepte i temporalitat). Un exemple es pot veure en el registre obtingut en la cerca del terme "Gioconda" o "La Joconde", registre de l'obra d'art que es troba al museu del Louvre (Annex 2.8.3) on es hi ha diferents tipus de relacions entre els conceptes (jeràrquiques o associatives):

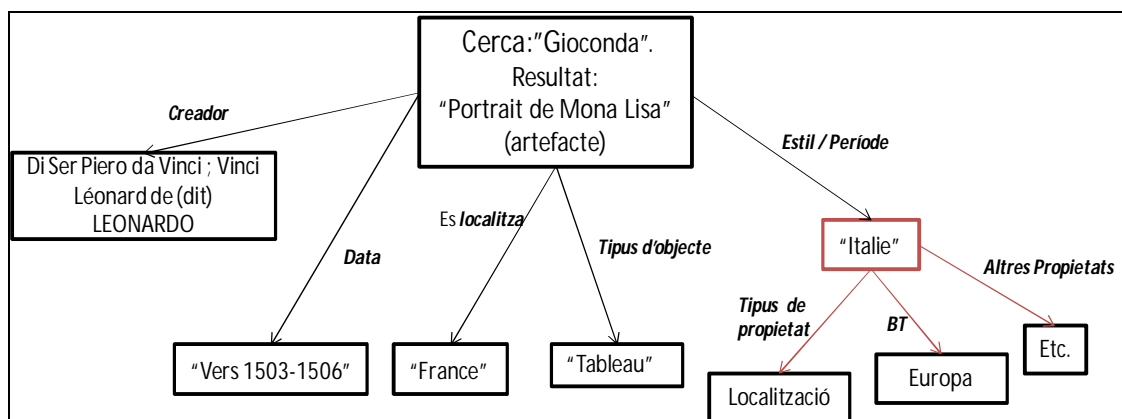


Figura 12. Exemple de relacions semàntiques d'un registre d'Europeana (Font: Elaboració pròpia)

2.4. Avaluacions dels sistemes d'informació de qualitat des de la perspectiva d'IQ

Com s'explicava al principi del capítol, les avaluacions de cada sistema es duen a terme des de tres dimensions que són essencials per avaluar els sistemes d'informació digitals (sistema, serveis i col·lecció), per tal de mesurar el nivell de qualitat de la informació. Cal destacar que els sistemes avaluats ja són d'entrada sistemes d'informació de qualitat; per tant, en aquest estudi es posa en pràctica la metodologia per identificar-los i avaluar-los.

Pel que fa al sistema o a la recuperació de la informació del sistema, s'han valorat aspectes com la representació de les dades, l'accessibilitat i/o la disponibilitat dels resultats (no s'avaluen aspectes d'accessibilitat web), l'autoria (la reputació de la informació recuperada i la integritat de les dades) i la consistència de les dades (actualització, fiabilitat, etc.). Pel que fa a la dimensió de serveis, principalment es prioritza l'eficiència (temps de resposta), l'efectivitat dels serveis (precisió), la reutilització dels serveis, la credibilitat i la facilitat d'ús (accessibilitat/disponibilitat). Finalment, quant a la col·lecció —no evaluable pel que fa a cobertura i exhaustivitat— s'avalua des de la perspectiva del context, per veure si els continguts s'adapten a l'audiència, i també es contempla la completesa de la informació, és a dir, el nivell de descripció dels resultats i si el sistema utilitza metadades estàndards per descriure els registres que formen la col·lecció. Aquest darrer paràmetre afecta directament els processos de recuperació de la informació duts a terme pels usuaris.

2.4.1. Metodologia

La metodologia aplicada en l'avaluació dels sistemes és una metodologia quantitativa realitzada a través de qüestionaris que descriuen cadascuna de les dimensions definides prèviament (sistema, serveis i col·lecció). Per a cada cas s'assignen preguntes diferents que corresponen a paràmetres de qualitat. Aquests paràmetres són els elements que permeten fer una valoració objectiva del nivell d'IQ de cada sistema d'informació. Cada paràmetre s'avalua i es puntua segons dues escales de valors. D'una banda, s'ha utilitzat una escala de valor sumatòria (Annex 2.2.1) amb 3 punts per indicar el valor de cada atribut seleccionat com a indicador d'IQ, definits per 1, nivell mínim d'IQ, 2, nivell mitjà, i 3, nivell alt. D'altra banda, per puntuar alguns d'aquests paràmetres de qualitat de la informació (per exemple, existència o no d'estàndards de metadades per descriure la informació) s'ha utilitzat una escala de distància (Annex 2.2.2) per valorar l'absència o presència d'alguns atributs de qualitat que admeten preguntes indirectes i que ajudaran a avaluar el nivell d'IQ de cada sistema d'informació (escala de dos punts: Sí=1, No=0). En general, aquestes escales de valor determinen el tipus de nivell de qualitat del sistema (alt, mitjà o baix). Un nivell alt d'IQ es representa amb un 5, i indica que el sistema aporta més del 70% d'elements IQ; un nivell mitjà o 3 indica que el

sistema aporta entre el 40% i 70% dels elements d'IQ, i finalment un nivell baix d'IQ, representat per 1, indica que s'aporta menys del 40% d'elements IQ.

Finalment, cal afegir que per tal de completar els qüestionaris s'ha tingut en compte la informació qualitativa recollida anteriorment en les fitxes descriptives elaborades per cadascun dels casos d'estudi.

2.4.2. Nivell d'IQ dels casos d'estudi

Tots els casos d'estudi avaluats, tant passarel·les temàtiques (Intute, Vascoda i AustLII) com sistemes semàntics (LIBRIS i Europeana), mostren un nivell d'IQ molt alt, per sobre del 85%, característica que confirma l'autodefinició com a sistemes de qualitat i/o amb alt nivell d'interoperabilitat.

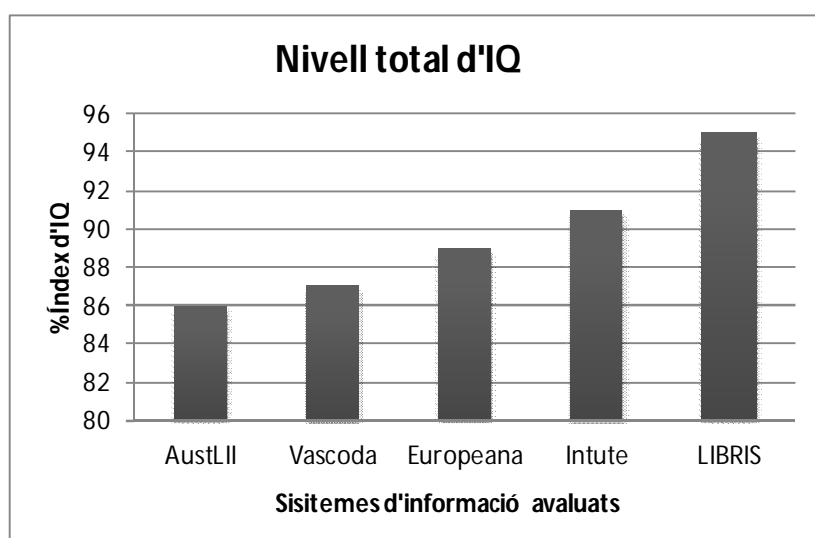


Figura 13. Nivell d'IQ dels sistemes d'informació avaluats

Les dades obtingudes en els qüestionaris d'avaluació donen uns resultats molt similars pel que fa a l'IQ global de cada sistema i analitzen les dimensions de forma individual, tal com es mostrava en l'apartat anterior, i, en aquest cas, sí que s'observen diferències significatives entre els sistemes.

Les diferències més significatives es mostren en la dimensió del sistema i la recuperació de la informació (RI). Tot i que els sistemes tenen un desenvolupament tecnològic molt diferent a causa de la diversitat de tipologia d'usuaris i de cobertura, el que es valora realment en aquesta dimensió (RI) és la integritat de les dades i la disponibilitat. En aquest sentit, només hi ha un sistema que cobreix la majoria dels aspectes d'IQ, LIBRIS, que ahora supera el 90% del nivell d'IQ pel que fa a la dimensió de sistema. En general, la resta de sistemes amb un nivell també alt d'IQ tenen un nivell més baix de disponibilitat i accessibilitat de la informació (Europeana, Intute), principalment perquè el nivell de completesa de la informació i integritat de les dades és més baix. Aquest és el cas de Vascoda i AustLII, que mostren un cop més que són sistemes que estaven o estan en vies de desenvolupament i millora, sobretot pel que fa a la

tecnologia. Un altre element que cal destaca és l'estructura i la consistència en la representació de la informació; Intute és l'únic sistema que té implementada tant la navegació i la indexació com la cerca creuada, i l'estructura organitzativa del coneixement que integra simplifica molt més l'accés a la informació per part de l'usuari, a més a més de facilitar l'ús del sistema d'RI.

Pel que fa a l'autoria, cap dels sistemes aporta informació addicional sobre la responsabilitat del document/objecte recuperat, a diferència de les bases de dades bibliogràfiques científiques on sí que sol aparèixer l'adreça electrònica o postal de l'autor principal de l'obra o *corresponding autor*. En la majoria dels casos aquest element ha fet que el nivell d'IQ no arribés al 100%.

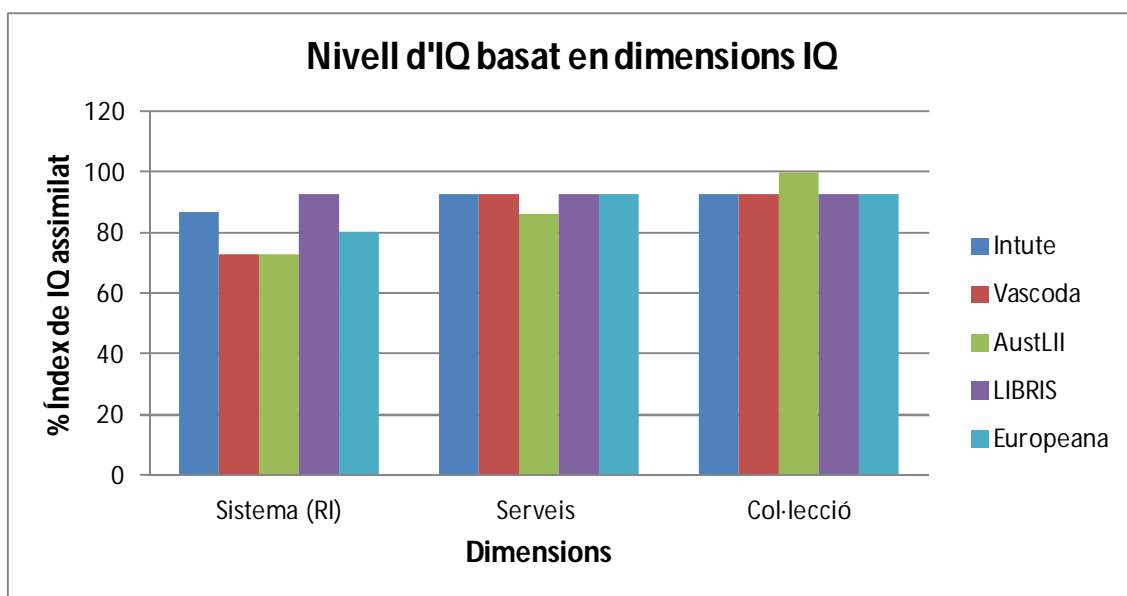


Figura 14. Nivell d'IQ individual per dimensions

Quant als serveis, com s'aprecia a la imatge, tots els sistemes tenen un nivell alt d'IQ, això és perquè tots inclouen els serveis bàsics²³ que ha de tenir qualsevol sistema d'informació digital considerat de qualitat. A excepció d'AustLII, tots tenen un nivell global d'IQ alt. Com ja s'ha explicat, AustLII està molt poc desenvolupat pel que fa als serveis, ja que la seva prioritat és el desenvolupament de la col·lecció. Pel que fa als serveis, Intute és el sistema amb un major nombre de serveis desenvolupats (un 84% són serveis addicionals) i també inclou serveis de personalització a l'usuari que no han estat implementats per cap altre sistema (Annex 2.9). Pel que fa a la personalització de serveis, les tendències dels sistemes semàntics estan orientats també cap al *Linked Data*. La col·lecció, com s'ha esmentat anteriorment, només s'avalua parcialment, ja que en aquesta investigació no s'han contemplat estudis d'usuaris, per tant, l'avaluació s'ha centrat en

²³ Serveis bàsics o mínims: s'interpreten serveis mínims els motors de cerca i navegació, serveis de notícies i arxiu (sigui quin sigui el seu mitjà o suport), serveis que permetin guardar continguts/cerques, exportar resultats, etc., i servei d'alertes sobre el contingut (independentment del mitjà i suport: e-mail, RSS, etc.), i serveis com "About us" (informació sobre la responsabilitat del sistema) o "Help" (ajuda).

aspectes que afecten l'usuari de forma global i no individual (criteris subjectius). És a dir, el nivell d'accessibilitat o disponibilitat per accedir a la col·lecció (accés al text complet o referències al text complet, serveis que facilitin l'obtenció de documents, etc.), l'actualització de la col·lecció o l'existència de polítiques d'actualització (dades sobre la freqüència d'actualització, referències a les versions dels continguts, data, perdurabilitat de la informació en el temps o la sensibilitat de la informació en relació a la freqüència de la informació, etc.). Els dos sistemes destacables són AustLII i Europeana, ja que les col·leccions es basen en contingut d'accés obert o *Open Access*. La resta de col·leccions, tot i que no sempre donen accés obert als seus continguts, ofereixen serveis d'obtenció de documents (SOD) o similars.

Seguint amb la dimensió de la col·lecció, també s'avalua l'autoria o reputació de les fonts d'informació de cadascun dels continguts (dades sobre els autors, referències o cites bibliogràfiques, etc.). Tot i que no sempre apareixen dades addicionals de les responsabilitats dels continguts (adreça, referències, etc.), sí que són fonts fiables tenint en compte la reputació, perquè és un criteri d'avaluació per part dels gestors/catalogadors dels sistemes abans d'introduir un ítem a la col·lecció.

Un altre aspecte important en la col·lecció és l'avaluació de la completesa, on s'estudia el nivell de presentació de les dades (sobretot ús de metadades estàndards) i el nivell de granularitat o especificitat (ús de vocabularis controlats en la descripció de la col·lecció i nivell jeràrquic amb què es representa). En aquest sentit, tots els sistemes cobreixen entre el 70% i el 90% d'IQ, a excepció d'AustLII que no incorpora metadades estàndards per representar la informació, ni vocabularis controlats per indexar la col·lecció. En canvi, pel que fa a la completesa de la cobertura, que no és objecte d'aquest estudi (per la manca d'estudis d'usuaris, etc.), AustLII és l'únic sistema que ofereix una cobertura exhaustiva, ja que afirma incorporar a la col·lecció totes les dades públiques de totes les fonts legals possibles d'Australàsia, i aquestes dades són proveïdes pels propis creadors.

Finalment, es tenen en compte també els paràmetres de precisió i rellevància de la col·lecció. S'avalua, d'una banda, la precisió de la informació en la presentació i la verificació (ús d'estàndards, inclusió de referències bibliogràfiques, ús de permaenllaços²⁴, etc.), i, d'altra banda, la rellevància des del punt de vista de l'aplicabilitat i la utilitat de la col·lecció tenint en compte l'audiència a la qual es dirigeix. També es contempen les eines addicionals que fan que la rellevància sigui possible en la recuperació de la informació, com l'ús de matèries en la navegació i en la indexació dels sistemes. Els sistemes amb un nivell d'IQ més baix són AustLII i Vascoda (43% i 50% d'IQ, respectivament). El primer no fa servir vocabularis controlats ni metadades

²⁴ Permaenllaços (permalinks): un permalink és una URL permanent que fa referència a un article específic d'un blog o foro.

estàndards ni en la cerca, ni en la navegació ni en la indexació, i el segon sistema en el moment de l'anàlisi encara no havia desplegat tots els vocabularis controlats en la cerca i en la indexació creuades.

A continuació, es presenten els resultats finals per a cadascun dels sistemes; es poden veure les puntuacions finals o nivell d'IQ que s'han valorat per cadascuna de les dimensions i paràmetres que caracteritzen un sistema d'informació de qualitat. Tots els qüestionaris, d'on s'han extret les dades, es poden consultar a l'Annex 2.9.

Atributs	1.Dimensió Sistema (Interrogació)	2.Dimensió Serveis (Tipologia)	3.Dimensió Col·lecció
Accessibilitat / Disponibilitat (1,2,3)	3	5	5
Actualització (2,3)	5	5	5
Autoria (2,3)	3	3	5
Completesa (2,3) [Cobertura, Metadades,etc .]	5	5	5
Comprensibilitat (1)	5	--	--
Estructura (1)	5	--	--
Integritat (1)	5	--	--
Precisió (1,2,3)	--	5	5
Rellevància (2,3)	--	5	5
Seguretat (1)	5	--	--
Total IQ	5	5	5

Taula 5. Intute: avaluació d'IQ

Atributs	1.Dimensió Sistema (Interrogació)	2.Dimensió Serveis (Tipologia)	3.Dimensió Col·lecció
Accessibilitat / Disponibilitat (1,2,3)	5	5	5
Actualització (2,3)	3	5	5
Autoria (2,3)	1	5	5
Completesa	3	--	--

(2,3)			
Comprensibilitat (1)	3	--	--
Estructura (1)	5	--	--
Exhaustivitat (2,3) [Cobertura, Metadades, etc.]	3	5	5
Integritat (1)	5	--	--
Precisió (2,3)	--	5	5
Rellevància (2,3)	--	3	3
Seguretat (1)	5	--	--
Total IQ	5	5	5

Taula 6. Vascoda: avaluació d'IQ

Atributs	1.Dimensió Sistema (Interrogació)	2.Dimensió Serveis (Tipologia)	3.Dimensió Col·lecció
Accessibilitat / Disponibilitat (1,2,3)	5	5	5
Actualització (2,3)	--	5	5
Autoria (1,2,3)	3	5	5
Completesa (2,3) [Cobertura, Metadades, etc.]	--	3	3
Comprensibilitat (1)	3	--	--
Estructura (1)	3	--	--
Integritat (1)	5	--	--
Precisió (2,3)	--	5	5
Rellevància (2,3)	--	3	3
Seguretat (1)	3	--	--
Total IQ	5	5	5

Taula 7. AustLII: avaluació d'IQ

Atributs	1.Dimensió Sistema (Interrogació)	2.Dimensió Serveis (Tipologia)	3.Dimensió Col·lecció
-----------------	--	---------------------------------------	------------------------------

Accessibilitat / Disponibilitat (1,2,3)	5	5	5
Actualització (2,3)	--	5	5
Autoria (1,2,3)	3	5	5
Completesa (2,3) [Cobertura, Metadades,etc .]	--	5	5
Comprensibilitat (1)	5	--	--
Estructura (1)	5	--	--
Integritat (1)	5	--	--
Precisió (1,2,3)	--	5	3
Rellevància (2,3)	--	3	5
Seguretat (1)	5	--	--
Total IQ	5	5	5

Taula 8. LIBRIS: Avaluació d'IQ

Atributs	1.Dimensió Sistema (Interrogació)	2.Dimensió Serveis (Tipologia)	3.Dimensió Col·lecció
Accessibilitat / Disponibilitat (1,2,3)	3	5	5
Actualització (2,3)	--	5	5
Autoria (1,2,3)	3	5	3
Completesa (2,3) [Cobertura, Metadades,etc .]	--	5	5
Comprensibilitat (1)	5	--	--
Estructura (1)	5	--	--
Integritat (1)	3	--	--
Precisió (1,2,3)	--	5	5
Rellevància (2,3)	--	3	5
Seguretat (1)	5	--	--
Total IQ	5	5	5

Taula 9. Europea: avaluació d'IQ

2.5. Conclusions

Tots els sistemes distribuïts d'informació de qualitat presentats segueixen un mateix model. La idea central és que són plataformes úniques per intercanviar d'informació en un àmbit cooperatiu, tant a nivell interinstitucional (cooperació entre diferents tipus d'institucions d'un mateix país o regió, amb àmbits de recerca multidisciplinaris, amb l'interès de posar en comú el coneixement científic o difondre'l: Vascoda, Intute, LIBRIS), com a nivell disciplinari (diferents institucions de diferents països que tenen un interès temàtic en comú: Europeana, AustLII).

En aquest model cooperatiu, són sistemes que integren informació de gran qualitat exhaustivament en la seva especialització, que intenten arribar al major nombre d'informació disponible a la xarxa i donar la màxima visibilitat als continguts per als membres de les comunitats. Contenen diferents tipus d'informació i documentació, integren tant referències bibliogràfiques com documents a text complet en els diferents tipus de fons bibliotecari/institucional d'un mateix sistema, mitjançant una interfície comuna, amb continguts propis i aliens. Es tracta, doncs, de models de sistemes d'informació que s'allunyen dels models tradicionals on hi havia una entitat local gestora i productora. En aquests models de sistemes d'informació, s'elimina el sentit d'identitat, no només per la col·lectivitat, sinó perquè són sistemes que es retroalimenten també de fons externs i aliens dels quals no en són propietaris (*Linked Open Data*).

Els os tipus de sistemes d'informació distribuïts analitzats, tipus passarel·la temàtica i sistemes semàntics, comparteixen una arquitectura de la informació similar, és a dir, utilitzen els vocabularis controlats en la d'indexació, cerca i navegació creuades com a eines d'interoperabilitat entre diferents col·leccions i sistemes d'informació. La diferència entre els dos tipus de models de sistemes presentats, els semàntics i els que no ho són, rau bàsicament en què els sistemes semàntics se centren principalment en la contextualització del contingut a través de les dades entrelaçades, és a dir, la resta de sistemes proporcionen poca relació del contingut que mostren (text, web, dades, imatges, vídeo, so, etc.) amb el context del propi contingut. Addicionalment, els sistemes semàntics, gràcies a l'estandardització i la interoperabilitat, faciliten la gestió i desenvolupament de *software*, la col·laboració entre diferents institucions i la interrelació entre diferents tipus d'informació i documentació (*Linked Data*). Per tant, són eficients amb el maneig de dades distribuïdes i en la seva visualització. Per contra, la gestió de la informació en la resta de sistemes és molt més feixuga i costosa, i la prova és que molts dels sistemes tipus passarel·la temàtica han anat desapareixent en el temps i les innovacions que es produeixen van cap a models de sistemes semàntics.

Una altra dada rellevant és que els sistemes semàntics s'aproximen cada cop més als usuaris. Dit d'una altra manera, els sistemes semàntics fan que els usuaris puguin ser partícips de la retroalimentació dels sistemes. No obstant això, la major part de les innovacions centrades en els usuaris encara estan en fase de desenvolupament (vegeu Capítol 1). Per contra, les passarel·les temàtiques amb una major trajectòria històrica, tot i que no incorporaven serveis semàntics, tenen una major experiència amb els serveis orientats als usuaris (vegeu el cas d'Intute).

CAPÍTOL 3. INDEXACIÓ I NAVEGACIÓ CREUADES

3.1. Introducció

El propòsit d'aquest capítol és estudiar les passarel·les temàtiques com a sistemes ecològics de recuperació de la informació a Internet amb els principals processos que comporta l'arquitectura de la informació d'aquests sistemes: la navegació creuada (*cross-browsing*) i la indexació creuada (*cross-indexing*).

L'interès d'aquest estudi sorgeix per la importància de les passarel·les temàtiques que tracten amb qualitat el contingut que gestionen. A vegades també s'anomenen biblioteques digitals perquè segueixen procediments molt similars en la catalogació, indexació i classificació dels continguts, així com en el control de qualitat a l'hora de seleccionar els recursos que formaran part de la col·lecció.

Un altre objectiu d'aquest treball és fer un estudi exploratori sobre l'estat de la la indexació i la navegació creuades, com són emprades per les diferents passarel·les, com funcionen, i avantatges i desavantatges. En aquest treball bibliogràfic es pretén estudiar on se centren les investigacions en l'àmbit de la indexació i navegació creuades i es vol mostrar les possibilitats que hi ha per desenvolupar recerques futures sobre el tema.

En aquest capítol no s'han limitat ni l'abast cronològic ni geogràfic, però la bibliografia existent sobre la tecnologia d'interès es fixa en l'àmbit geogràfic europeu. Els casos d'estudi analitzats han estat les passarel·les del Regne Unit (RDN: SOSIG, PSIGate, HUMBUL, Altis, Artifact, EEVL, Gesource i BIOME) entre altres projectes de passarel·les europees (Renardus i *DESIRE*) i les passarel·les nòrdiques (NISBIG). En aquest treball es parla també del projecte alemany CARMEN (model de base de moltes d'aquestes passarel·les mencionades), però se n'ha descartat l'estudi per problemes lingüístics, ja que la major part de la informació que ens proporcionava aquesta passarel·la temàtica estava en alemany.

3.2. Les passarel·les temàtiques

Les passarel·les temàtiques o *subject gateways* es coneixen com a «webs, basats en un mecanisme per accedir a una col·lecció d'alta qualitat, on s'avaluen tots els recursos identificats per donar suport a la recerca en una determinada matèria o disciplina» (*australian subject gateways forum*). Però en aquest estudi exploratori s'analitzen les passarel·les temàtiques com a sistemes avançats de recuperació de la informació que compleixen uns valors ecològics, com la implementació de la indexació creuada i la navegació creuada, característica principal d'aquest tipus de sistemes.

3.2.1. Context

Les passarel·les temàtiques apareixen juntament amb el desenvolupament d'Internet a mitjan dels noranta. En el seu desenvolupament s'identifiquen quatre fases (Dempsey, 2000):

1. L'esotèrica. En aquesta etapa les passarel·les es consideren xarxes de treball o instruments esotèrics per a les comunitats científiques.
2. Comunitat. L'ús se centra en les comunicacions, la recerca i en els hàbits col·laboratius entre diversos camps, com l'acadèmic i la recerca. Es comença a crear un espai social comú i, al mateix temps, comencen a ser utilitzades com a serveis d'informació. Un exemple d'aquesta fase és l'aparició de la *Virtual Library (VLib)* creada per Bernners-Lee (1994), altres exemples d'aquesta fase són *Bubl* i *Aliweb* (Regne Unit, 1994).
3. Infraestructures d'informació en l'àmbit acadèmic i de recerca. Aquest tipus de xarxes integren informació acadèmica i recursos estratègics per a la recerca i l'aprenentatge.
4. Infraestructures d'informació pública. Internet en aquesta etapa passa de ser un espai de comercialització i privatització a tenir infraestructures d'informació públiques. Cal destacar que les passarel·les temàtiques en el seu recorregut històric han estat considerades infraestructures d'informació acadèmica o de recerca i, posteriorment, espais de col·laboració entre diferents comunitats científiques. La seva aparició coincideix també amb l'aparició d'infraestructures públiques.

Una de les primeres iniciatives importants es va desenvolupar al Regne Unit amb un programa de biblioteques electròniques (eLib) dut a terme al 1994 pel Joint Information Systems Committee (JISC). Les passarel·les temàtiques formaven part de l'àrea ANR (*Acces to Network Resources*) on es desenvolupaven eines de navegació en xarxa i de creació de matèries d'àmbit local que esdevenien eines i servidors d'informació. Un altre projecte desenvolupat per eLib va ser ROADS (*Resource Organisation and Discovery for subject-based services*, 1994) un *software* gratuït i un conjunt de normes per crear passarel·les i portals temàtics que també va ser implementat a Finlàndia. «ROADS està compost per una sèrie de mòduls PERL que es descarreguen com un paquet i s'instal·len a un servidor, també inclou la utilització del protocol de cerca i recuperació de la informació WHOISS++» (Garcia Delgado, P, 2001).

Les passarel·les comencen a evolucionar al 1995 amb l'objectiu de crear xarxes amb serveis efectius per a la recerca i l'aprenentatge, que responguessin a la necessitat de proveir i gestionar col·leccions de recursos mitjançant la descripció del recurs i, que donessin accés a través de la matèria (navegació o *browsing*) i la cerca (*searching*). SOSIG (Ciències Socials) va ser

la passarel·la pionera en el projecte *eLib*, que el 1994 ja descrivia uns 300 recursos; després van aparèixer d'altres passarel·les com ADAM (Art, Disseny, Arquitectura i mitjans de comunicació), EEVL (Enginyeria), OMNI (Medicina), History (Història) i Biz/Ed. També hi havia altres països que desenvolupaven iniciatives similars, com Holanda, Dinamarca, Finlàndia i Austràlia.

3.2.2. Característiques de les passarel·les temàtiques

Les passarel·les temàtiques es caracteritzen pels controls de qualitat, ja que preval la rellevància o qualitat de la informació per sobre de la quantitat. Els elements que formen part d'aquest control de qualitat són:

- Selecció i desenvolupament de la col·lecció. Esforç humà i intel·lectual en el procés de selecció dels recursos que comprenen la col·lecció (generalment els criteris de selecció no són publicats). En aquest punt, també s'hi inclou una política de selecció pròpia de la institució, com la cobertura temàtica, geogràfica i el tipus de documents.
- Gestió de la col·lecció. És un procés regular per tal de mantenir el nivell de qualitat de la col·lecció, com per exemple comprovar que els enllaços funcionin, etc.
- Creació de metadades. És un procés manual i intel·lectual fet per experts o professionals de la informació i la documentació.
- Descripció de recursos i metadades. Els documents seleccionats són descrits sota unes normes i amb metadades (normalment estàndards internacionals) que identificaran i localitzaran el recurs.
- Accés a les matèries. Tots els recursos són indexats a través d'un esquema de classificació de matèries o vocabularis controlats que constitueixen la "navegació per matèries" del sistema. Aquest és el punt més característic del control de qualitat.
- Accés a la cerca i a la navegació. Els sistemes han de permetre accedir a la cerca i a la navegació de la informació a través de processos més o menys avançats, però que permetin als usuaris cercar per camps, operadors booleans, cerques combinades a través de diferents elements de metadades, navegar a través d'arbres temàtics, permetre opcions de filtratge (limitar per llengua o tipus de documents), etc. Dublin Core, suport del protocol Z39.50 per realitzar cerques, ús d'estàndards internacionals com l'HTML, XML, RDF, etc. Aquesta característica permet cooperació entre diferents sistemes.

Altres elements característics de les passarel·les temàtiques de tipus general:

- Cooperació. S'estableix cooperació i coordinació entre diferents països amb l'objectiu de donar cobertura temàtica i, en la majoria dels casos, també geogràfica a una comunitat d'usuaris, i facilitar la millora dels serveis oferts per les passarel·les.
- Accés integrat a recursos impresos i electrònics. La infraestructura de les passarel·les, gràcies a les tecnologies d'aplicació del protocol Z39.50, permet integrar recursos impresos (provinents d'un OPAC, per exemple) i recursos web de la xarxa. També s'ha de tenir en compte la interacció de formats o esquemes de descripció per recuperar aquests recursos impresos i electrònics (indexació creuada).
- Recuperació sostenible. Donar accés lliure a la informació i que aquesta respongui a uns criteris de selecció comuns: que la recuperació i el manteniment dels sistemes sigui sostenible.
- Eines de cerca creuada. En una sola interfície l'usuari pot accedir mitjançant la cerca creuada a diferents passarel·les temàtiques de diversos països.
- Navegació creuada. Generalment permet a l'usuari cercar informació per matèries des d'una mateixa interfície però en diverses passarel·les.
- Duplicació d'esforços. S'estableixen mecanismes de cooperació i coordinació per establir uns criteris comuns que evitin la duplicació d'esforços entre les diferents passarel·les integrants d'una xarxa:
 - Desenvolupar criteris comuns per a la selecció de recursos.
 - Normalització de sistemes. Llenguatges documentals, codis específics (país, dates, etc.), esquemes de metadades, etc. Per facilitar la interoperabilitat entre els sistemes.
 - Crear uns principis d'accés a la informació compatibles amb el disseny d'estructura dels camps utilitzats.
 - Determinar condicions sobre drets de propietat intel·lectual.
 - Avaluar i aconsellar tant en la provisió d'aspectes tècnics com en l'arquitectura dels sistemes.

3.3. Navegació creuada i cerca creuada

3.3.1. Navegació creuada o *cross-browsing*

Els aspectes principals a resoldre en els primers projectes de passarel·les temàtiques eren els processos de navegació i recerca.

Quant a la navegació s'optava per estructures jeràrquiques que podien contenir organitzacions de tipus geogràfic o distribuïdes per matèries (CDU).

La navegació creuada és un servei clau en totes les passarel·les temàtiques perquè facilita l'accés al contingut per mitjà de la representació temàtica, independentment de si abasta o no diferents passarel·les temàtiques, tenint en compte que cadascuna utilitza les seves classificacions temàtiques. Aquests esquemes es basen en estàndards i classificacions pròpies, a més, l'estructura de navegació (*browsing*) de cada passarel·la és diferent, es poden trobar d'1 a 5 nivells, a més de subdivisions temàtiques per cadascun dels conceptes representats. En el cas de Renardus, per exemple, s'intenta mapejar²⁵ Dewey (*Decimal Classification*). És a dir, hi ha un procés on s'identifiquen termes equivalents dels tesaurus o classificacions locals a una classificació universal com per exemple la DDC. En aquest cas, la navegació creuada es fa amb el sistema de navegació local i no directament en els registres, després es mapeja tota l'estructura del sistema de navegació incloent les relacions entre termes i la sintaxi d'aquestes relacions. Es poden consultar exemples de les relacions de mapeig que fa Renardus als annexes 3.1. i 3.3.

Cadascun dels participant de les passarel·les tenen un sistema propi de lectura dels esquemes de classificació. A més, s'utilitza una eina que proporciona OCLC (WebDewey) per facilitar la producció del mapeig d'informació en una sintaxi correcta. Veure un exemple del Sistema de mapeig Renardus en l'Annex 3.1. (exemple 2). En el cas de Renardus la navegació creuada fa el seu procés de mapeig per mitjà de la DDC, i, a més, en l'apartat de "cerca avançada" podem fer una recerca per la mateixa classificació (Veure Annex 3.2., exemple 1).

Finalment la navegació creuada en una passarel·la es pot dur a terme de diverses formes, seguint l'exemple de Renardus s'observen els següents processos:

- Emprar l'estructura de la DDC per navegar per les pàgines web de Renardus, utilitzant la "cerca avançada" on també es selecciona incloure o no els diferents esquemes de classificacions de les passarel·les integrades. (Veure Annex 3.2., exemple 1)
- Navegar per les estructures temàtiques de la passarel·la. Aquestes normalment apareixen en la pàgina principal i solen ser de tipus jeràrquic. (Veure Annex 3.2., exemple 2)
- Utilitzar les estructures de navegació de les passarel·les locals. Aquest procés de navegació, anomenat "navegar i saltar" (Annex 3.2, exemple 2.3), és el resultat del procés de "navegació per l'estructura temàtica" on

²⁵ Consultar la definició del terme mapeig al Glossari final d'aquesta tesi.

a través d'una matèria seleccionada apareixen els diferents nivells d'equivalència existents sobre el terme entre diferents passarel·les temàtiques locals. Altrament, cada nivell d'equivalència et permet "saltar" directament a la passarel·la local. Els nivells d'equivalència identificats per Renardus són:

- Plena Equivalència (*Full equivalence*)
 - Equivalència d'Acceptats (*Broader equivalence*)
 - Equivalència de "No Acceptats" (*Narrower equivalence*)
 - *Major overlap*. El contingut de la pàgina local s'adequa en gran part al terme seleccionat en la pàgina de navegació.
 - *Minor overlap*. Indica que l'equivalència del terme seleccionat representa una part menor en la col·lecció local representada. Navegació amb l'estructura DDC mitjançant les pàgines web de Renardus.
- Navegació gràfica. Amb la navegació per l'estructura temàtica de Renardus apareix l'opció de navegar per mitjà d'un gràfic facilitant la cerca a l'usuari inexpert (amb manca de recursos o que no sàpiga com buscar la informació desitjada) (Veure l'Annex 3.1)

3.3.2. Mapeig de DDC per a la navegació creuada

El fet que cadascun dels participants que formen part de les passarel·les temàtiques tingui el seu propi esquema de classificació per a la navegació creuada fa que hi hagi la necessitat de "mapejar" aquests esquemes per crear un sistema de classificació comú.

L'organització del coneixement basat en els esquemes de classificació temàtics té els següents avantatges:

- Navegació. És el principal avantatge perquè facilita la recerca d'informació als usuaris inexperts que no estan familiaritzats amb una matèria, la seva estructura (termes relacionats jeràrquics o associatius), o bé, la terminologia associada a la matèria en qüestió.
- Interoperabilitat. Permet i facilita el treball en xarxa gràcies a l'ús d'esquemes de classificació. Per exemple: quan un usuari cerca una interfície el sistema interroga diferents passarel·les que permeten al client emprar esquemes de classificació, i, al mateix temps el sistema utilitza aquests esquemes de classificació per a la recuperació de la informació.

- Accés multilingüe a una col·lecció. Molts esquemes de classificació universal com és la DDC o la CDU es publiquen en diferents llengües. Aquests esquemes ofereixen accés multilingüe als recursos sense la necessitat de traduir el contingut del document.

La DDC és el sistema de classificació utilitzat per la major part de passarel·les temàtiques com és el cas de la xarxa de passarel·les del Regne Unit RDN (SOSIG, PSigate, HUMBUL, Altis, Artifact, EEVL, Gesource i BIOME), Renardus (passarel·les europees) o NISBIG (passarel·les nòrdiques).

La DDC aporta molts avantatges que altres esquemes de classificació no tenen (Neuroth i Koch, 2001):

- Aplicabilitat d'eines online. La DDC és una eina que s'adapta còmodament a l'ús de sistemes digitals a Internet, i és fàcil d'aplicar-se en navegadors i plataformes (web Dewey).
- Ús global. La DDC és el sistema més utilitzat a nivell global, i s'escolleix molt sovint per a la classificació de recursos d'Internet. Algunes passarel·les temàtiques com Renardus utilitzen la pròpia DDC com a sistema de mapeig.
- Adequació dels sistemes de classificació i els seu funcionament. En molts casos s'opta per la DDC per la seva senzillesa, ja que la CDU, tot i ser (molt) més nova i permetre estructures facetades, és tres vegades més extensa que la DDC (conté 61.000 classes més). Alhora de representar el coneixement al web, l'usuari vol estructures senzilles que li permetin navegar pel sistema de forma clara i precisa, per això s'estandarditza l'ús de la DDC i no el de la CDU entre altres sistemes de classificació.

Dos bons exemples d'esquemes de classificació utilitzats a Internet són BUBL LINK i Yahoo. BUBL LINK és un servei que proporciona recursos web de cobertura acadèmica en totes les àrees. En la seva pàgina principal s'observa l'ús de la DDC com a sistema de navegació i organització. Yahoo és un cercador comercial que dona cobertura a matèries populars però, també utilitza aquest esquema de classificació universal amb les 14 categories principals.

Altres casos de sistemes d'informació com són les passarel·les SOSIG i EELS utilitzen esquemes de matèries més específics. SOSIG en la seva estructura de navegació usa tant un thesaurus especialitzat com la CDU, i EELS emprava esquemes de classificació especialitzats amb enginyeria.

3.3.3. Navegació creuada i cerca creuada

Tradicionalment la cerca apareixia com a complement de la navegació per l'extensa informació que facilitava l'exploració i l'accés als usuaris. Però, els

cercadors no eren precisos i sovint no aportaven cap mena de descripció, per tant, tot i refinar la seva recerca, l'usuari es trobava amb una sobreabundància d'informació.

Amb l'aparició de les metadades, com a eina per descriure el contingut dels recursos d'informació, la cerca assoleix un nivell superior de recuperació de la informació (més rellevància i precisió). Actualment, el format estàndard de *metadades* DublinCore.

Cal destacar que una passarel·la ha de complir uns criteris de qualitat i definir unes metadades comunes utilitzades per promoure l'accés a la cerca i la navegació. En aquest procés cada participant integrat en una passarel·la serà responsable de mapejar els seus formats de metadades per construir-ne un de comú, per això és necessari conèixer:

- Elements de metadades que són suportats per cadascun dels participants.
- Definició semàntica de cada element (refinaments i esquemes de codificació).
- Definició semàntica de cada element (regles, codis, estàndards, etc.)
- Regles de catalogació utilitzades.
- Repetibilitat de cada element.
- Forma d'obligatorietat de cada element (obligatori, recomanació, opcional, etc.).
- Qualificador de llengua (per al títol, descripció i matèria) per informar sobre la llengua de les metadades.

L'ús de metadades en les passarel·les temàtiques no només ha servit com a mitjà de recuperació de la informació, sinó que també s'ha emprat per indexar la informació d'Internet. Aquest és el cas del *Projecte DESIRE* que utilitza diversos softwares per indexar la informació del web (Rybarzyk, s.a).

3.4. Indexació creuada

La indexació com a eina de recuperació en els entorns digitals és més complexa que en els entorns tradicionals. Les passarel·les temàtiques, en els seus inicis, estudiaven la possibilitat d'utilitzar llistes d'encapçalaments com ara la *Library of Congress Subject Headings* (LCSH) com a sistemes d'indexació, però, es van adonar que aquestes passarel·les tenien una estructura molt complexa. Llavors, tot i simplificar el seu vocabulari, els sistemes de classificació es van imposar per ser més esquemàtics, sobretot en el cas de la DDC.

Les classificacions permetien a l'indexador representar les col·leccions locals d'una forma més específica, i els sistemes de classificació aportaven una classificació més acurada que la llarga pre-coordinació de la LCSH i la simple post-coordinació dels descriptors d'un tesaurus.

Actualment en la indexació d'Internet, fora de les passarel·les temàtiques i iniciatives similars, no hi ha controls de qualitat establerts, ja que s'utilitzen simples algorismes on s'extreuen les paraules més significatives d'un document (paraules del títol o de la descripció). També es té en compte altres elements com el posicionament dels termes i la seva possible reiteració (mesures de freqüència, ponderació, etc.), a més, molts cercadors presenten la informació per la seva popularitat (nombre de cites que rep un web). L'ús d'algorismes produeix una sobreabundància d'informació que arriba a saturar l'usuari, llavors, per evitar-ho (tot i això) es poden seguir altres criteris com la cerca mitjançant les metadades o per un sistema d'indexació de continguts web que cada cop és més utilitzat. Cal recalcar que en les passarel·les temàtiques, com s'ha explicat anteriorment, la indexació és creuada i es centra en el mapeig de continguts utilitzats també en la navegació creuada.

3.4.1. Indexació creuada i mapeig

El mapeig és un procés d'identificació de termes, conceptes i relacions jeràrquiques que són aproximadament equivalents entre els diferents llenguatges controlats en un mateix sistema. Les relacions de mapeig s'expressen entre un parell de classes i no entre una classe i un recurs individual. En l'apartat de navegació creuada, el mapeig se centrava en els sistemes de classificació com un procés unidireccional d'una classificació DDC cap a una o varies classificacions locals (sistemes de navegació). En aquest procés quan es tracta una classe s'ha de veure el nivell d'especificitat del contingut de la matèria i trobar una plena equivalència entre aquesta classe i les seves subclasses, i finalment, mirar les coincidències. (Veure exemple 1, Annex 3.1.).

La DDC és per moltes passarel·les, juntament amb les metadades, el sistema de classificació més utilitzat per indexar els seus continguts. Però en la indexació creuada també es pot parlar de mapeig de tesaurus (Doerr, M. 2001) com a un vocabulari controlat que permet la millora la qualitat en la recuperació de la informació i la interrogació a les bases de dades. El mapeig de tesaurus es pot concebre com un procés per fusionar tesaurus, metatesaurus i construcció de concordances creuades.

El tesaurus es contempla també com una ajuda a la cerca, on l'usuari s'aproxima al llenguatge expert que s'ha utilitzat tant per indexar recursos d'informació com en els conceptes associats als termes, un bon exemple és SOSIG.

Les dificultats amb el mapeig dels tesaurus són les diferències semàntiques i les relacions jeràrquiques que poden haver entre ells. Aquests problemes semàntics i de relacions, que també es donaven amb el mapeig de classificacions, estan associats als mètodes estadístics i a les xarxes neuronals emprades en el mapeig. Una altra dificultat dels tesaurus, citada per molts autors, és l'alta complexitat dels vocabularis controlats que comporta una migració cap a nous tipus de terminologia. Per això, actualment es busquen sistemes que siguin adaptables i capaços de fer processos de transició, sobretot es plantegen dos temes:

- Correlació de tesaurus (es busca la substitució de conceptes). Un exemple el tenim amb el projecte alemany CARMEN on es duen a terme correlacions de diferents *tesaurus* alemanys en ciències socials i literatura, aquests processos es basen amb mètodes intel·lectuals i estadístics simultàniament.
- Federació de tesaurus (es busca la relació de conceptes). Un exemple és el projecte MACS (Consorti de biblioteques Nacionals Europees) que actualment investiguen la viabilitat de la correlació de LCSH amb totes les llengües europees.

Un altre gran obstacle en el mapeig de tesaurus és l'heterogeneïtat:

- Utilització de diferents paraules, diferents llengües, diferents nivells en les llengües, etc.
- Diferents cobertures (segons les necessitats dels usuaris. Els tesaurus poden tenir diferents cobertures geogràfiques, per països, per regions, àrees administratives, per comunitats, etc.
- Diferències semàntiques. Diferències entre conceptes incloent *tesaurus* de diverses llengües
- Diferències entre les relacions semàntiques. Diferències en monojeràrquies o en aspectes de classificació.
- Multilingüisme. No s'ha de confondre el multilingüisme (concordança entre termes de diferents vocabularis tenint en compte les llengües naturals amb la traducció de conceptes de tesaurus.

L'heterogeneïtat de les estructures jeràrquiques dels diferents sistemes que hi ha en una passarel·la esdevé un altre punt a tractar. La correlació entre dos tesaurus suposa la interrelació entre jerarquies i relacions (termes acceptats, no acceptats, etc.). Aquest procés de mapeig a vegades comporta ambigüitats, no obstant això, s'ha demostrat que la inconsistència entre els termes es dona per errors humans. Doerr (2001) proposa que les relacions jeràrquiques i semàntiques haurien de fer-se clarament abans que els tesaurus siguin

correlacionats, a més, aquestes correlacions haurien de concordar amb les metadades.

Un altra problemàtica és la confusió de l'usuari a causa de la polisèmia i l'homonímia, per això s'haurien d'evitar termes ambigus, o bé, posar indicadors per diferenciar el seu significat. Tot i que els tesaurus es basen en jerarquies de paraules i conceptes no hauria d'haver confusió entre termes en entorns digitals on l'usuari navega per jerarquies virtuals.

Finalment, podem dir que el procés de mapeig amb tesaurus estableix termes i/o conceptes d'indexació entre conceptes provinents de diferents tesaurus que formen part d'una passarel·la temàtica. El procés de mapeig serveix per substituir els termes originals provinents de les col·leccions locals quan l'usuari fa una interrogació amb una passarel·la als diversos sistemes o passarel·les integrants en el sistema. (Veure exemple 3 de l'Annex 3.1.)

3.5. Conclusions

Aquesta premissa plantejada per Sutter es té en compte al llarg d'aquest estudi i també en les passarel·les temàtiques concebudes sempre com a eines que permeten l'accés a la informació de qualitat. L'eix d'aquest capítol està focalitzat principalment en la indexació i navegació creuades com a eines ecològiques (Capurro, 1990), ja que intenten donar accés a la informació sense cap tipus de barrera social ni lingüística. És cert, que els continguts de cadascuna de les passarel·les temàtiques es dirigeixen cap a un públic en concret i especialitzat, però l'accés a la informació està dissenyat per arribar a tothom. La navegació creuada és un exemple de disseny senzill per a què els usuaris més inexperts puguin accedir a la col·lecció. Després, hi ha la cerca creuada per aquells usuaris familiaritzats en l'entorn, a més, l'opció simple i avançada del buscador permet a l'usuari introduir un text lliure per fer les seves cerques. Una altra possibilitat que ofereixen les passarel·les temàtiques és l'exploració amb l'ús de vocabularis controlats. En el cas de SOSIG s'ofereix la recerca amb tesaurus, Renardus, i amb l'opció avançada es pot buscar mitjançant la classificació decimal de Dewey (DDC) i fer una cerca combinada per diversos camps (títol, autor, etc.). En canvi, la passarel·la EELS²⁶ d'enginyeria permet fer cerques creuades amb els vocabularis combinats amb llenguatge natural o limitant.

Perquè la indexació i la navegació creuades es produeixin hi ha d'haver un procés de mapeig de la informació entre diferents passarel·les temàtiques. El

²⁶ Cerca creuada d'EELS: <http://eelsdb.lub.lu.se/cgi-bin/egwcgi/egwirtcl/eelsscreen.tcl/name=find&lang=eng&service=aeels2>

mapeig es pot considerar com una gran eina ecològica que possibilita la consistència i la interoperabilitat entre les passarel·les temàtiques.

El mapeig amb els sistemes de classificació permet l'accés a la informació i l'organització de la informació. Els avantatges i usos ecològics que proporciona el mapeig es poden resumir en quatre punts:

- La conversió entre diferents sistemes per incorporar registres d'una estructura local a una passarel·la temàtica. En aquest procés també hi ha una conversió entre els esquemes de classificació (indexació creuada). Aquests sistemes poden ser OPACS, bases de dades o serveis d'Internet.
- Suport per a la traducció de categories i termes entre diferents llengües. El mapeig representa la cobertura de termes en diferents llengües i busca la seva equivalència (dimensió lingüística, Capurro, 1990).
- Provisió de la navegació creuada entre diferents serveis els quals tenen, amb els seus propis sistemes de classificació, la provisió de la cerca creuada entre diferents sistemes, poden tenir) amb els seus propis sistemes de classificació, així com també la provisió de la cerca creuada entre diferents sistemes.
- La interoperabilitat entre diferents sistemes, esquemes de classificació, regles de catalogació, metadades, etc.

En l'estudi de casos de les passarel·les temàtiques s'han detectat diferents punts de conflicte com són la discussió que hi ha entre la utilització d'esquemes de classificació i la indexació de matèries. No obstant això, es conclou que les eines més utilitzades per a la indexació i navegació creuades són els estàndards DDC i les metadades Dublin Core considerades actualment unes de les eines més eficaces en el mapeig de la informació del sistema d'informació per la seva fàcil aplicació. Cal recordar que la DDC és l'estàndard de classificació més conegut arreu del món amb una estructura clara i precisa per a l'usuari, per això sol ser escollida com a eina d'indexació i navegació creuada, i, al mateix temps, pot interoperar amb altres vocabularis controlats. Aquesta hipòtesi és una línia de recerca que caldria estudiar en profunditat per tal de veure com afecten aquests processos en la recuperació de la informació científica (especialitzada), ja que els sistemes de classificacions estan pensats per a informació de tipus generalista (multidisciplinària). També és necessari analitzar les estructures de cerca i navegació de diversos vocabularis controlats per comprovar si els sistemes de classificació són, o no, els vocabularis controlats més eficaços. Entre els casos estudiats hi ha exemples de sistemes d'informació de tipus passarel·la que no utilitzen esquemes de classificació com a eines d'indexació i navegació creuada, com SOSIG que utilitza tesaurus en la indexació i navegació, o el projecte CARMEN que utilitza llistes

d'encapçalaments de matèries. De fet, en els capítols 7 i 8 d'aquesta tesi es planteja la hipòtesi de l'ús de mapeig de tesaurus com a vocabulari controlat representatiu en l'àmbit científic i com a mecanisme facilitador de la interoperabilitat de diferents sistemes d'informació.

Finalment, només cal dir que s'ha de seguir treballant cap al perfeccionament d'aquests sistemes ecològics perquè actualment són l'aproximació més clara a una ecologia de la informació com a models de qualitat i de no pol·lució informativa.

CAPÍTOL 4. ESTÀNDARDS I FORMATS

4.1. Introducció

Els estàndards bàsics que s'han de tenir en compte en qualsevol sistema d'informació en un entorn digital fan referència, d'una banda, als **formats de descripció bibliogràfica** i **d'objectes digitals** dels recursos d'informació continguts en el sistema, i, d'altra banda, als **estàndards i protocols d'intercanvi d'informació** que es poden donar en un entorn federat o distribuït²⁷.

Els estàndards en sistemes d'informació digitals (distribuïts o no) serveixen principalment per proporcionar interoperabilitat entre la informació continguda en el sistema d'informació i els serveis d'informació integrats en la plataforma digital. Per tant, es considera que és necessari integrar tant la interoperabilitat semàntica com la sintàctica.

La **interoperabilitat sintàctica** fa referència als **sistemes de codificació i representació sintàctica** que permet que tant les dades com la informació no es puguin barrejar ni se'n puguin canviar els seus **formats, codificacions, propietats, valors i tipus de dades** (DELOS Project núm. 507618, pàg. 7 i 9). En aquest sentit es fa referència a la utilització de formats estandarditzats de codificació de documents (com els formats XML, MARC o RDF, és a dir, **formats d'estructures de dades**). La **interoperabilitat semàntica** fa referència a la utilització d'estàndards basats en la **representació semàntica del contingut documental**. La interoperabilitat semàntica és un dels majors objectius que cal aconseguir en la recerca sobre biblioteques digitals. Doerr (2004), defineix la interoperabilitat semàntica com una tasca complexa que afecta a diversos nivells i funcions dels sistemes d'informació; està formada per tres nivells bàsics: **estructures de dades** (metadades, dades de contingut, dades de gestió de col·leccions, dades de serveis de descripció), **dades categòriques** (dades universals, com classificacions, tipologies o matèries en general, o dades categoritzades, és a dir, vocabularis controlats), i **dades factuais** (dades particulars o específiques que es refereixen a gent, ítems o llocs, és a dir, dades que fan referència a controls d'autoritats, geogràfics, etc.).

Pel que fa a aquest capítol, la recerca s'ha centrat bàsicament en els formats i estàndards d'estructura de dades (metadades que afecten a la sintàxi i la semàntica) per descriure i recuperar la informació; en d'altres capítols es tracten les dades categòriques o factuais (sistemes d'indexació i classificació, vegeu capítols 3, 5 i 7). Com apunta Greenberg (2003), les metadades són ja

²⁷ Totes les taules i gràfics d'aquesta secció són d'elaboració pròpia. Ambdós estan confeccionats a partir de les guies citades i corresponents a cadascun dels estàndards i formats presentats. Ans al contrari, es cita la font corresponent. Afegir que totes les sigles que apareixen en aquest capítol es troben desenvolupades en l'apartat del Glossari.

dades estructurades sobre un objecte basades en unes normes. Cal tenir en compte que l'objectiu de les metadades va més enllà que la descripció formal i de contingut tradicional de l'àmbit de la biblioteconomia i la documentació; segons la NISO 2004, les metadades són definides com a "informació estructurada que descriu, explica, localitza o facilita la recuperació, ús o gestió d'un recurs d'informació". Pel que fa a estàndards, generalment es poden definir quatre grans grups:

- Grup 1: Estàndards d'identificació i localització: DOI, URI, etc.
- Grup 2: Estàndards de llenguatge de marcat: HTML, XML, RDF, MARC, etc.
- Grup 3: Estàndards protocols d'intercanvi d'informació: Z39.50, SRU, OAI-PMH, etc.
- Grup 4: Estàndards de metadades: Dublin Core, MODS, METS, etc.

Tot i que és necessari tenir en compte tots els estàndards en el disseny d'un sistema d'informació, aquest capítol se centra només en els estàndards rellevants per a la interoperabilitat que afecten el contingut dels sistemes d'informació, és a dir, estàndards del grup 2 i del grup 4, que representen metadades que afecten la descripció i la recuperació de la informació. La resta d'estàndards (grup 1 i 3) es refereixen a aspectes tècnics i tecnològics que fan possible l'accés i intercanvi de la informació en un o més sistemes. Els estàndards d'identificació i localització necessaris per localitzar i donar accés al document/objecte/informació estan integrats i descrits en els estàndards de metadades (grup 4). I els estàndards tecnològics (grup 3), fan possible la interoperabilitat entre els sistemes, per exemple, en el cas del Z39.50 o SRU + SRW tradicionalment han servit per permetre la interoperabilitat de metadades utilitzades en biblioteques (Gradmann, S., 2008). L'estàndard OAI-PMH ha servit com a mètode per permetre còpies de pàgines i arxius web i fer més simple l'intercanvi de la informació basada en OAI-PHM, entre d'altres aproximacions a serveis web com SPAP/UDDI i la API, basada en JAVA definida en JCR (JSR 170/283), així com plataformes basades en GRID com iRods.

Evidentment, per assegurar la interoperabilitat de sistemes d'informació digital, és primordial utilitzar metadades elaborades tenint en compte **estàndards** d'ús comú en el seu àmbit d'aplicació per fer possible que recursos descrits amb esquemes diferents puguin ser recuperats independentment del seu origen. Convè destacar que les metadades són un conjunt d'informació estructurada que permet descriure, explicar i localitzar la informació; així, doncs, ajuden a recuperar, usar i gestionar la informació (NISO, 2004). Aquestes metadades estan formades per esquemes de metadades que tenen l'objectiu de descriure un recurs d'informació específic; segons el grup de treball sobre metadades de

l'American Library Association Committee on Cataloging: Description and Access (CC:DA, 2000) «un esquema de metadades proporciona una estructura formal dissenyada per identificar l'estructura de coneixement d'una disciplina determinada i vincular-la a la informació de la disciplina a través d'un sistema que ajudi a identificar el descobriment i ús de la informació dins d'aquesta disciplina». Els esquemes de metadades inclouen tant aspectes de contingut, semàntics, com de codificació (especificació del llenguatge de marcatge utilitzat, per exemple, XML). En canvi, els estàndards de metadades corresponen a un conjunt de regles definides per una comunitat d'usuaris que fan referència a la representació del contingut (identificació d'un autor, títol, etc.) i contempnen també els valors d'aquest contingut (per exemple, si un valor o metadata s'ha d'agafar d'un vocabulari controlat).

El context de les metadades és molt heterogeni, ja que poden haver-n'hi tants esquemes com àmbits d'aplicació, i poden correspondre o no a estàndards. De manera que hi ha diferents tipus d'esquemes que cal tenir en compte:

- Basats en els estàndards. Esquemes normalitzats, aprovats per organismes oficials de normalització (ISO, ANSI, AENOR, etc.). Per exemple, l'estàndard Dublin Core (ISO 15836:2003), el MODS o el EAD, entre d'altres exemples consultables al web de la *Library of Congress*.
- Basats en les especificacions. Esquemes desenvolupats per consorcis o entitats industrials. Aquestes iniciatives serveixen per fer arribar la necessitat de normalitzar-los als òrgans de normalització corresponents i per tant, poden servir de base per a l'elaboració de normes. Per exemple, IMS (*Learning Resource Metadata Information Model*), Richardson i Powell (2003).
- Basats en les implementacions. Anomenats també perfils d'aplicació. Són esquemes que responen a necessitats específiques de comunitats d'usuaris. Solen recollir elements d'un o diversos estàndards o especificacions, i s'implementen o bé apliquen restriccions d'ús (nombre d'elements o repeticions) o bé afegeixen extensions (nous elements o vocabularis de noms, etc.).

L'aproximació als estàndards de metadades de biblioteques i centres de documentació en ciències aquàtiques i marines, que es pot agrupar en un context comú de biblioteques especialitzades i/o científiques (incloent-hi també les universitàries) no divergeix del d'altres disciplines i/o biblioteques especialitzades. Com es va estudiar en el capítol d'interoperabilitat semàntica, a través de les entrevistes a professionals de la xarxa europea i mediterrània de biblioteques i centres de documentació en ciències aquàtiques i marines (EURASLIC i MedLibs), els estàndards i formats d'estructures de dades més

utilitzats són el format MARC²⁸ i MARC 21²⁹, tant per a dades bibliogràfiques com per a autoritats (dades factuais). A més, són formats que tenen el suport de la *Library of Congress* a través de la xarxa *Network Development and MARC Standards Office*.

El segon format més utilitzat és l'estàndard de metadades Dublin Core, les funcions del qual se centren en el descobriment de recursos digitals i contribueixen també a la cerca, a la recuperació i a l'accés a la informació. Dublin Core és una iniciativa internacional que apareix el 1995 (Dublin Core Metadata Ekenebt Set, v.1.1) amb l'explosió d'Internet, però no es consolida fins al 2005 com a recomanació internacional i es formalitza i s'usa com a marc per a la construcció formal en el marc de recursos del W3C. Dublin Core utilitza *Resource Description Framework (RDF)* com a model de dades, implementat amb XML, que permet la representació de la informació sobre recursos del World Wide Web. L'Objectiu de Dublin Core és facilitar el vincle entre les dades i servir com a base per desenvolupar especificacions de sintaxi independents a l'aplicació i a la restricció d'idiomes. Per tant, Dublin Core és integrable a qualsevol sistema d'informació digital i compatible amb altres formats de descripció com MARC o MARC21.

Així doncs, en aquest capítol s'han agrupat i estudiat els formats i estàndards que afecten principalment l'àmbit de les biblioteques i centres de documentació especialitzats en ciències aquàtiques i marines. Tanmateix és evident que són formats i estàndards d'ús internacional i que l'àmbit d'aplicació és extensible a qualsevol tipus d'unitat d'informació. S'avaluen també aquells formats i estàndards relacionats amb els principals formats de referència i també els formats i estàndards que es preveuen com a referents en les biblioteques digitals i/o sistemes distribuïts, independentment del context d'aplicació temàtic o geogràfic.

Pel que fa als formats referents citats anteriorment (MARC i Dublin Core), cal especificar que, tot i que es caracteritzen per la interoperabilitat, encara no se n'ha estès l'ús en un mateix sistema d'informació. Així mateix, moltes biblioteques usen els estàndards en diferents col·leccions i bases de dades. Per exemple, en l'àmbit català, l'ús de Dublin Core està estès en repositoris digitals (com és el cas de RECERCAT) i l'ús del format MARC21 en catàlegs de biblioteques individuals o col·lectius (CCUC); els dos sistemes estan coordinats pel Consorci de Biblioteques Universitàries de Catalunya (CBUC). El motiu és simplement que cada format va ser creat amb un objectiu diferent: com s'ha esmentat anteriorment, Dublin Core apareix per la necessitat de descriure recursos web, ja que Internet començava a ser un mitjà d'informació

²⁸ Tots els formats MARC (MARC Standards) estan regulats per l'estàndard internacional ISO 2709:1996 Information and documentation -- Format for Information Exchange; Actualitzada per: ISO 2709:2008

²⁹ MARC 21, està promogut per la IFLA (International Federation of Libraries Association), i s'utilitza internacionalment i està traduït a 19 llengües. Veure servei de traduccions de la MARC Office: <http://www.loc.gov/marc/translations.html>. MARC21,

creixent i aquest nou format permetria fàcilment descobrir i descriure aquests tipus de recursos digitals (Dutta, B., 2003). L'estàndard Dublin Core té el seu suport actual de la iniciativa internacional *The Open Archives Initiative Protocol for Metadata Harvesting*, que indica l'ús de Dublin Core simple (DC) com a requisit mínim de qualsevol repositori obert o base de dades *open access* (OA).

Actualment, s'està produint ja el canvi esperat cap al desenvolupament de sistemes d'informació més interoperables i flexibles que permeten l'ús de diverses col·leccions i, al mateix temps, permeten diferents formats de descripció i accés a la informació. Tot i així, la majoria d'iniciatives es duen a terme en l'àmbit patrimonial i amb sistemes d'informació distribuïts. Un exemple català és el Catàleg Col·lectiu de les Biblioteques dels Museus de Barcelona, publicat el 2011, on s'utilitzen formats estàndards i d'interoperabilitat com RIS, Dublin Core, MARC21-XML. La *Library of Congress*, en els darrers anys (2010-2013), també ha impulsat la utilització de RDA (*Library of Congress*, 2012), un nou format estàndard de contingut (com les AAR2) emprat per saber quina informació es registra i com. RDA prové del format conceptual FRBR (1998), que és un format complementari al format MARC. Té l'objectiu d'afegir més atributs (elements de marcatge) i relacions en les dades bibliogràfiques i d'autoritats, per tal que els usuaris puguin aprofitar aquest coneixement i perquè les dades puguin ser reutilitzables de diverses formes, i no només per a la descripció/cita bibliogràfica habitual. Tot i que l'aplicació del RDA és limitada (principalment s'usa a Nord-Amèrica), el 2013 s'hi van afegir biblioteques i arxius del Canadà, la *National Library of Australia*, la *German Speaking Library Community*; a nivell nacional hi ha l'exemple del CCUC, que publica la guia per la integració de registres RDA l'abril del 2013 (CCUC, 2013), també s'ha de tenir en compte la *British Library*, que forma part del comitè d'administració de les RDA (Miller, 2013). Aquest format es considera el format de catalogació futur, ja que actualment es publica a l'*Open Metadata Registry* com a *Linked Open Data (LOD)*, per tant, ofereix un vocabulari més ric per descriure els recursos i per expressar relacions entre ells.

El 2002 apareix el format MODS (*Metadata Object Description Schema*), un altre format i esquema de descripció d'objectes digitals que, a diferència del Dublin Core, es va crear pensant en l'àmbit bibliotecari (biblioteques digitals). Aquest format s'usa sobretot en biblioteques nacionals i té el suport de la *Library of Congress* (vegeu *MODS: uses and features*, 2012) i la seva oficina MARC, *The Network Development and MARC Standards Office*. MODS va ser creat per complementar el format MARC21, també descriu com a format capaç de suportar dades seleccionades de registres ja existents amb format MARC21, i també té la capacitat de crear registres de descripció de recursos.

La diferència més important entre Dublin Core i MODS és l'àmbit d'aplicació. DC s'ha estès entre repositoris digitals, però alhora és molt popular en altres

contextos com l'àmbit de les biblioteques (*DCMI Libraries Community*), en l'àmbit geoespacial (Projecte INSPIRE, 2008)³⁰, científic (*DCMI Science And Metadata Community*) i en l'àmbit educatiu (*Dublin Core Education Community*), entre d'altres. En canvi, l'àmbit d'aplicació de l'estàndard MODS es limita a les biblioteques nacionals.

En l'àmbit científic, que és el cas d'estudi, l'escenari és encara molt més ampli, ja que les metadades per descriure objectes varien segons la matèria o àrea d'especialització. Molts cops no és suficient amb l'ús d'un esquema de metadades per descriure recursos/objectes. Com a exemple, l'ús de Dublin Core només proporciona metadades en la font d'informació. Altrament, l'ús d'unes metadades específiques, per exemple, el cas del model de metadades geoespacial (Projecte INSPIRE), permet incloure tant les metadades de la font d'informació com les metadades en el seu registre. És a dir, si es vol adaptar el model de Dublin Core per proporcionar metadades en el seu registre és necessari definir cada registre com un recurs en si mateix, i provocar així una càrrega innecessària del sistema d'informació; és per això que en molts casos s'opta per l'ús de models de metadades específics o fets a mida segons el tipus d'informació a descriure.

Es poden trobar altres exemples similars també en el camp de les ciències aquàtiques i marines, on s'identifiquen uns 62 esquemes de metadades o estàndards de continguts (vegeu *International Metadata Marine Interoperability Project*), fet que el converteix en un àmbit temàtic molt heterogeni. Un altre exemple és l'àmbit de les ciències agràries; el 2002 es van presentar unes guies de bones pràctiques per descriure informació i objectes digitals que es van publicar el 2005, *AGRIS Application Profile*, és esquema de metadades específic que agrupa tres esquemes de metadades referents en el camp: Dublin Core (DC), [Australian Government Locator Service Metadata \(AGLS\)](#) i [Agricultural Metadata Element Set \(AgMES\) namespaces](#); els tres esquemes poden ser representats amb els formats HTML, XML o RDF. Actualment, a través del portal *Agricultural Information Management Standards (AIMS)* de la FAO, s'identifiquen fins a 30 tipus de perfils d'aplicació de metadades relacionats en l'àmbit agrari (vegeu *Metadata Set VEST Registry*).

Aquesta diversitat de formats i estàndards en àmbits científics es deu principalment a la heterogeneïtat de les disciplines i a la diversitat de la tipologia d'informació, gestionada per diferents tipus de bases de dades (bases de dades bibliogràfiques, bases de dades d'espècies, de gens, taxonòmiques, de dades atmosfèriques, oceanogràfiques, etc.), ja que cadascuna correspon a necessitats d'informació diferent (recullen diferents recursos i nivells de descripció). Normalment, els perfils i models de metadades prenen com a

³⁰ Projecte INSPIRE, té en compte Dublin Core (DC) per a la difusió de metadades pròpies a altres comunitats geoespacial i en l'administració electrònica, i també s'usa DC per a la cerca bàsica de recursos geoespacial.

referència la base de dades per als quals són creats, per tant, és normal que hi hagi tants models diferents com bases de dades. No obstant això, actualment diferents comunitats amb un interès comú (INSPIRE, MMI o AGRIS) centren els seus esforços a desenvolupar projectes comuns i fomentar l'ús d'estàndards i esquemes de metadades comuns.

De manera paral·lela, cal tenir en compte també l'escenari global que envolta la informació digital i els sistemes que gestionen i permeten l'accés i consum de la informació. El fet que l'accés, la creació i la publicació de la informació a través d'Internet estigui a l'abast de tothom fa que la informació digital creixi de forma desmesurada, en conseqüència, augmenta la informació desestructurada i heterogènia, i en general hi ha dificultats per accedir i recuperar la informació de qualitat. Per exemple, l'aparició del web 2.0 ha fet aparèixer nous formats d'intercanvi d'informació (RSS, Atom, etc.) i serveis web o xarxes socials (blogs, wikis, etc.), en els quals l'usuari i/o consumidor d'informació és també creador i gestor d'informació (Alonso-Arevalo, et al., 2010). És a dir, el nivell de continguts a la xarxa incrementa fàcilment, així com la veracitat i el rigor de la informació disponible, i es viu un període de infoxicació digital (Cornella A., 1996). Així doncs, els sistemes d'informació digital i el desenvolupament del World Wide Web tendeixen a la construcció d'un web més estructurat amb nous formats i estàndards que permetin aquesta recuperació de la informació de qualitat; alhora, els sistemes d'informació digitals tenen la capacitat de comprendre i assimilar la informació i el coneixement que contenen o que poden descobrir al web. En aquest nou escenari, les biblioteques i els sistemes d'informació digital tenen un paper molt important, perquè són tradicionalment fonts d'informació estructurades amb continguts de qualitat.

Pel que fa a les biblioteques digitals o als sistemes d'informació digitals que les integren, cal dir que també hi ha hagut un canvi de mentalitat. Les biblioteques han patit canvis en les col·leccions, en els formats, en les comunitats d'usuaris i en el comportament informacional, en conseqüència han canviat els serveis cap als usuaris. Les biblioteques digitals, en general, són més heterogènies que les biblioteques tradicionals. En el cas de les col·leccions, en el model de biblioteca tradicional, s'associa la col·lecció amb el fons de la biblioteca, l'usuari hi pot accedir a través de la consulta amb el catàleg de la biblioteca, que és l'índex de les col·leccions de la biblioteca i que conté les metadades dels elements d'aquestes col·leccions. En canvi, en una biblioteca digital, el concepte de col·lecció és "un grup d'objectes", i aquests objectes no són necessàriament físics i/o propietat de la biblioteca. Una col·lecció pot ser un conjunt de metadades que apunta als recursos distribuïts de diverses col·leccions seleccionades en la xarxa per la rellevància que ha definit la pròpia biblioteca digital (similitud temàtica, cooperació bibliotecària, etc.).

En definitiva, l'escenari actual pensa en el web 3.0 o web semàntica (W3C, *copyright* 1994-2005). L'objectiu principal d'aquesta web és donar més sentit als recursos d'Internet per tal de poder recuperar la informació digital d'una manera més fàcil i més ràpida que abans, i donar més coherència als recursos obtinguts d'una consulta. El web 3.0 busca solucions que facilitin la cerca d'informació a través d'una estructura comuna amb la qual poder compartir, reutilitzar, processar i transferir informació amb facilitat. Aquest web 3.0 té més significat i es basa en llenguatges universals que poden resoldre problemes del web relacionats amb la manca de semàntica. Fins ara l'accés al web es feia a través de buscadors (per exemple, Yahoo o Google) que donen accés a la informació a través d'unes paraules clau que proporciona un individu o usuari d'Internet. Els resultats obtinguts depenen del vocabulari usat en la cerca i no sempre es recupera la informació desitjada, ja que els documents significatius solen contenir una terminologia diferent a la usada en la cerca. El procés de recuperació de la informació en els cercadors presenta problemes greus d'exhaustivitat (*recall*) i una precisió baixa en els resultats obtinguts, tot i que cerques semànticament iguals haurien de donar resultats similars. Així doncs, el web semàntic intenta que les màquines puguin interpretar oracions i extreure'n la informació útil per al usuari. S'intenta representar el contingut web d'una forma més fàcilment processable per a una màquina. Tot i que el terme *web semàntica* va ser creat el 1998 per Tim Berners Lee, no es troben projectes i iniciatives on es comencen a veure les aplicacions reals del web semàntic fins a l'actualitat. El terme web semàntica es refereix a la visió del W3C sobre un web amb dades enllaçades. Les dades enllaçades poden ser expressades a través dels formats i estàndards RDF, SPARQL, OWL, i SKOS.

L'especificació RDF (Resource Description Framework) es compon d'un conjunt de recomanacions del W3C, publicat el 2002, i la primera versió RDF Schema va ser el 1999 (Bikakis N., et al. 2012). RDF és un model de dades per especificar objectes digitals o recursos i relacions entre ells, que ofereix una semàntica simple i compatible amb la resta de formats estàndards per al web semàntic. És un format que pot ser representat amb una sintaxi XML, a més a més, les tecnologies OWL i SKOS es basen en RDF, ambdós formats permeten una major integració i interoperabilitat entre les dades que RDF. RDF Schema és un llenguatge de descripció de vocabularis que permet especificar les propietats i les classes de recursos RDF, amb una semàntica per establir jeràrquies de generalització d'aquestes propietats i classes

SPARQL és un llenguatge de consulta per a dades estructurades en RDF que permet extreure informació i al mateix temps construir-ne de nova (extreu dades en forma de URI i literals, i també pot extreure subgrafs RDF i construir nous grafs) a través de les consultes a repositoris RDF. Les consultes es poden realitzar a través de repositoris via API o bé a través d'un punt de consulta web amb entrada SPARQL. És a dir, es pot arribar a consultar una col·lecció

específica d'informació, de documents o d'objectes o bé realitzar una consulta genèrica a qualsevol arxiu RDF accessible al web.

OWL (Ontology Web Language) està dissenyat per ser utilitzat en aplicacions que necessiten processar el contingut de la informació en lloc de representar només la informació publicada per qualsevol persona (web 2.0). Es publica per primer cop el juliol de 2002. OWL és un llenguatge de descripció que proporciona un vocabulari enriquit, amb una semàntica formal, característica que el converteix en un format més complet i complex que la resta de formats (XML, RDF o SKOS) utilitzats per a les descripcions de propietats i classes, per exemple en les relacions entre classes (per exemple, disjunció), la cardinalitat, la igualtat, el triplat enriquit de les propietats, les característiques de les propietats (per exemple, simetria) i l'enumeració de les classes. OWL té tres subllenguatges: OWL Lite, OWL DL i OWL Full. OWL és un format que facilita un major mecanisme de interoperabilitat amb el contingut web i que és compatible amb la resta de formats, com l'XML, RDF o RDF esquema (RDF-S) que s'havien utilitzat fins ara.

SKOS (Simple Knowledge Organization System) proporciona un model de metadades per representar l'estructura bàsica i el contingut dels esquemes conceptuals de vocabularis controlats o de sistemes d'organització del coneixement (per exemple, encapçalaments de matèria, taxonomies, esquemes de classificació, tesaurus, etc.), que funcionen com a llenguatges de recuperació de la informació en sistemes d'informació o biblioteques digitals.

SKOS és una ontologia OWL Full, però com que té característiques simples no contempla una semàntica formal i completa. Els orígens del en es remunten a la redacció d'un diccionari de sinònims de les activitats dins el projecte SWAD-Europe2. La primera versió de SKOS Core es va presentar a l'any 2003.

Recollint tot el que s'ha exposat, el web semàntic també afecta els sistemes d'informació digital com les biblioteques digitals. Destacant el que exposa el W3C, els formats de llenguatge ontològic permeten una major interoperabilitat amb el contingut web. Per tant, diferents sistemes d'informació digital podrien ser fàcilment integrables si utilitzessin aquests formats i tecnologies. De manera que, dintre d'un mateix entorn digital (biblioteca o sistema distribuït), no caldria separar catàlegs, repositoris de documents o qualsevol altre sistema d'informació digital. Així doncs, en aquest capítol s'exposen, d'una banda, els tipus de formats que faciliten la creació d'un sistema digital d'informació distribuïda, s'avalua la compatibilitat entre diferents sistemes d'informació existents, s'estudien els formats i estàndards més estesos segons el cas d'estudi (biblioteques especialitzades d'àmbit científic) i, d'altra banda, es mostren les tendències actuals amb exemples pràctics que demostren la funcionalitat del web 3.0 i la compatibilitat i interoperabilitat amb altres formats en sistemes d'informació més tradicionals.

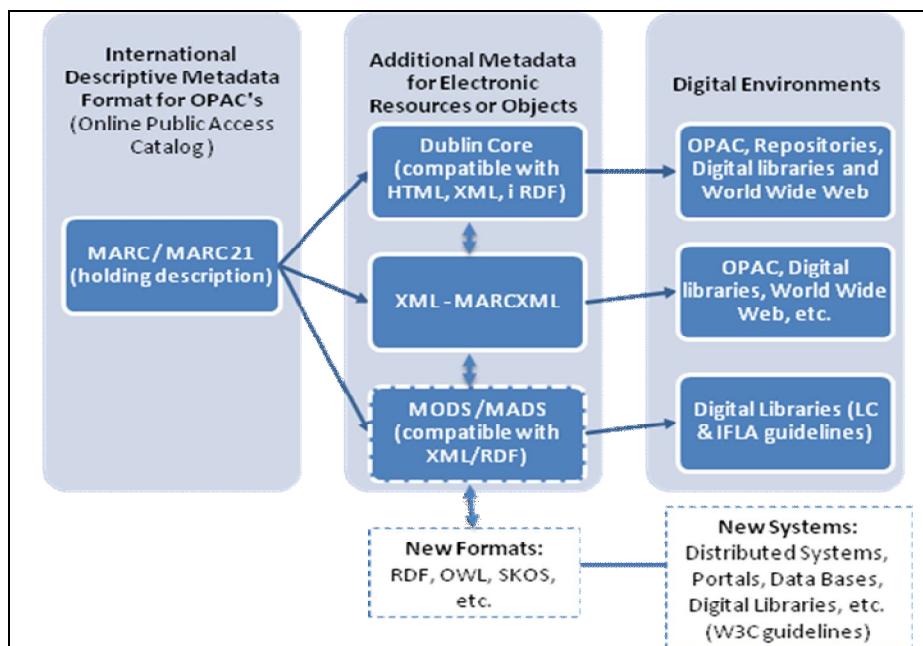


Figura 15. Formats tradicionals i nous formats: compatibilitat en l'àmbit d'aplicabilitat (Font: Sebastià-Salat; Reverté, C., 2011)

Tots els formats tenen un nivell de interoperabilitat alt perquè són estàndards i estan definits com a mínim amb XML Schema. Alhora, a partir de l'exemple d'un entorn tradicional com un catàleg de biblioteca (sistema d'informació mínim que es pot trobar en una biblioteca especialitzada), que conté com a mínim un estàndard com el MARC, es mostra com pot ser mapejat o com pot ser convertible en diversos formats existents en sistemes d'informació i al mateix temps pot ser interoperable amb nous formats i estàndards en nous sistemes d'informació digital. El principi bàsic és que XML és un format estàndard de descripció de recursos web, per tant, l'ús de qualsevol tipus de metadades per descriure recursos i objectes digitals en entorns tradicionals que utilitzin XML o MARCXML fa possible que hi hagi una major flexibilitat en els sistemes d'informació. Així doncs, la utilització d'esquemes de metadades XML permet també que sistemes bibliotecaris puguin rebre i compartir registres externs de qualsevol tipus de sistema d'informació o del web. XML permet un ús col·laboratiu d'intercanvi de metadades a través de servidors com l'OAI o serveis web Z39.50 com SRU.

4.2. Estàndards i formats

En aquesta secció es descriuen els estàndards i formats des del punt de vista de la interoperabilitat entre sistemes d'informació d'àmbit especialitzat, concretament en l'àmbit de les ciències aquàtiques i marines, que és l'àrea objecte d'estudi. Es descarta l'anàlisi de nous formats d'aplicació recent com el cas de RDA, perquè durant tot el període d'estudi no hi havia aplicacions ni indicis d'una aplicació del format en sistemes d'informació de biblioteques i centres de documentació de l'àmbit de les ciències aquàtiques i marines.

L'objectiu d'aquest capítol és definir un model de recuperació d'informació de qualitat dins d'un entorn digital o sistema d'informació amb l'ús dels estàndards i formats més adequats per desenvolupar sistemes d'informació interoperables. Al mateix temps s'avalua el nivell d'interoperabilitat entre els diferents formats estàndards més estesos, per tal d'estudiar la viabilitat d'un sistema distribuït en un entorn de biblioteques i centres de documentació especialitzats en àmbits científics, concretament en el camp de les ciències aquàtiques i marines.

El capítol es divideix en dos grups de normes i formats que contempnen metadades que afecten el contingut d'un sistema d'informació, que pot ser distribuït o no, però que vol facilitar l'intercanvi d'informació:

1. **Estàndards i formats per a la descripció dels recursos i/o objectes digitals.** També són anomenats *formats de dades estructurades*, perquè representen un document o un objecte per l'estructura bibliogràfica.
2. **Estàndards i formats per a l'intercanvi d'informació** i per permetre la **interoperabilitat** entre diversos sistemes, també anomenats *formats de dades de valor*, ja que representen un valor en relació amb el document i la descripció (per exemple: llocs, persones o subjectes).

4.2.1. Estàndards i formats per a la descripció de recursos i/o objectes digitals

Aquest primer grup d'estàndards, coneguts també com a formats d'estructura de dades, es defineixen com a formats d'intercanvi d'informació. Bàsicament s'utilitzen en sistemes bibliotecaris tradicionals (catàlegs) i actuals (biblioteques digitals, sistemes distribuïts, etc.). Els formats principals que hi ha en aquesta categoria són els formats bibliogràfics, d'autoritats, de classificació i de descripció d'objectes. Els més destacats en els entorns bibliotecaris són el format MARC i el Dublin Core.

4.2.1.1. MARC (*Machine-Readable Cataloguing*).

És un format estàndard per a l'emmagatzematge i intercanvi bibliogràfic de registres d'informació (materials textuais impresos i manuscrits, arxius d'ordinador, mapes, música, recursos continus, materials visuals i mixtes, etc.) llegibles per màquina. També es coneix com a format de descripció bibliogràfica que defineix sintaxi, significat i regles de descripció de recursos recollits en una biblioteca o entitat similar. El MARC va néixer amb l'objectiu de crear un sistema comú per descriure recursos de biblioteques, i va ser originat per la Library of Congress cap als anys seixanta. Tots els estàndards MARC responen a la Norma ISO 2709:1996 (*Information and documentation -- Format for Information Exchange*), són mantinguts per la Library of Congress i tenen el seu suport.

Els formats MARC són normes per a la representació i a la comunicació d'informació bibliogràfica i relacionada, representades amb un format llegible per màquina (definició extreta de la Library of Congress). El seu ús, estès a nivell mundial per diferents tipus de biblioteques, evita la duplicació d'esforços, ja que una comunitat de biblioteques pot compartir l'adquisició i catalogació del mateix material, i suposa un estalvi econòmic per a tot els implicats. És a dir, l'ús del format MARC en biblioteques ha comportat un treball col·laboratiu i cooperatiu entre biblioteques, que es tradueix molt cops amb catàlegs col·lectius d'àmbit nacional o especialitzat.

Un dels formats MARC més estesos i recents és el MARC 21, format per USMARC i CANMARC (MARC21, 1998). El MARC 21 s'utilitza també per a la representació i comunicació de registres bibliogràfics i també és llegible per màquina. El MARC 21 és mantingut per la per la Library of Congress, la Library and Archives Canada i la British Library en consulta amb les comunitats d'usuaris. La British Library va adoptar MARC 21 com a format de catalogació el juny de 2004 com a part de la implementació d'un sistema integrat de biblioteques. Existeixen cinc MARC 21 Concise Formats, usats per a la descripció de dades rellevants: bibliogràfic, autoritats, registres, classificació i comunitat. Un registre MARC està format per una capçalera, el registre amb informació general sobre el contingut del recurs, i un nombre de camps que contenen les dades del recurs.

100 1# \$a Darwin, Charles \$d 1809-1882					
100	1#	\$a	\$d	Darwin, Charles	1809-1882
Etiqueta	Indicadors	Codis de subCamps		Valor dels subcamps	

Taula 10. Exemple de camp de dades MARC

Cada camp MARC és definit per una *etiqueta*, formada per un nombre del 001 al 999 (no tots els camps s'utilitzen). Els camps amb etiquetes de la 001 al 008 s'utilitzen com a camps de control, i la resta d'etiquetes són camps de dades del registre. Cada *etiqueta* pot ser definida amb més detall amb l'ús de dos *indicadors*, que són nombres del 0 al 9. Finalment, cada camp està definit per subcamps, que, a la vegada, són definits per codis alfabètics delimitats pel símbol de dòlar (\$), com s'observa en l'exemple 1. Cal afegir que les etiquetes representades per 0XX i 8XX es classifiquen com a etiquetes de "control d'autoritat". Finalment, cal explicar que les etiquetes es divideixen en centenars i que el contingut pot ser descrit de forma resumida amb 9 camps mínims (MARC 21, 2011).

- | |
|---|
| <ul style="list-style-type: none"> - 0XX: Camp d'informació de control - 1XX: Entrada principal (per exemple: autoria) - 2XX: Títols, edició, publicació, impressió, etc. - 3XX: Descripció física del document/objecte a descriure |
|---|

- 4XX: Informació sobre la col·lecció
- 5XX: Camps d'accés de matèria
- 7XX: Camps d'entrades secundàries (entrades a altres formes d'autoritat 70X-75X; o referent a enllaços 76X i 78X)
- 8XX: Camps d'entrada secundària de col·lecció (80X-83X) i camps de fons, localització, grafies alternatives, etc.

Taula 11. Exemple de camps mínims MARC21

4.2.1.2. Dublin Core (DC)

És un estàndard de metadades que va sorgir com una iniciativa internacional el 1995 amb l'explosió d'Internet. Però no es va establir ni reconèixer com a un format internacional fins al 2005. Avui dia, és un dels estàndards i esquemes de metadades més estesos en les biblioteques i centres de documentació, entre d'altres àmbits d'aplicació, i és també reconegut com a marc per a la construcció de recursos digitals en el context del W3C. En el camp de les ciències aquàtiques és també el més utilitzat juntament amb el format MARC i les seves versions subseqüents.

El propòsit de Dublin Core és fer d'enllaç entre les dades i servir de base per desenvolupar aplicacions. A més a més, és un estàndard que es pot utilitzar de forma independent a les especificacions o restriccions de les aplicacions, tant pel que fa al llenguatge com a la sintaxi. Per tant, Dublin Core és defineix com a format interoperable amb qualsevol sistema d'informació digital i, com ja s'ha introduït anteriorment, és compatible amb els formats de descripció, com MARC o MARC21, utilitzats en OPAC. No obstant això, hi ha molts centres d'informació i biblioteques digitals que encara fan diferència en l'ús d'aquests tipus de formats segons el tipus de sistema digital emprat. L'exemple més estès és l'ús de Dublin Core per a repositoris digitals (RECERCAT) i l'ús de MARC21 en catàlegs de biblioteques (CCUC). La raó és lògica: simplement cada format va ser creat en un ordre diferent i amb un objectiu diferent. És a dir, Dublin Core neix amb la necessitat de descriure recursos web, ja que Internet era una font creixent d'informació i aquest nou format permetria fàcilment el descobriment i la descripció d'aquests recursos web. En canvi, els formats MARC es creen per representar i/o descriure un fons bibliogràfic, localitzable físicament en una institució, per tal de facilitar l'accés a la informació a través d'un entorn digital.

En els darrers anys, el moviment *Open Access* (OA) i el creixement conseqüent de repositoris digitals han facilitat també l'ús de l'estàndard Dublin Core, reconegut en les directrius per a l'aplicació del Protocol de la Iniciativa d'Arxius Oberts per a la Recollida i Preservació de Metadades (OAI-PMH); a més, ha estat ratificat per les normatives ISO 15836:2009 de febrer de 2009, ANSI /

NISO Z39.85-2007 estàndard de maig de 2007, i IETF RFC 5013 de l'agost de 2007.

Per comparar els dos formats (MARC i DC), d'entrada, cal valorar l'estructura de cadascun. D'una banda, els formats MARC es componen d'una capçalera (prové de dades per al processament del registre), un directori (entrada de dades corresponents a l'etiqueta, longitud i el punt d'inici de cada camp d'informació d'un registre) i uns camps variables, que poden ser de dades o de valors (és on s'organitzen les dades del registre; són identificats amb una etiqueta numèrica de tres caràcters que s'emmagatzema a l'entrada corresponent del directori).

A continuació es mostra un quadre comparatiu dels dos estàndards, basat només en l'anàlisi de contingut d'un registre bibliogràfic:

MARC 21	Dublin Core Simple amb XML
CAPÇALERA: 00000nam a2200361 a 4500	[no es contempla]
DADES DE CONTROL: 001 .b24552471 007 cr 008 000224s1999 ne 001 eng c 020 0792356314 (versió paper) 020 0306475855 035 0792356314 9ES-BaCBU001 035 1792-47660 9ES-BaCBU035 080 632	[no es contempla]
245 00 \$a (Title statement/ Title proper) e.g. 245 00 \$a Integrated pest and disease management in greenhouse crops /\$c R. Albajes ... [et al.]	<dc:title> Title: El nom del recurs </dc:title>
650 #0 \$a (Subject added entry- Topical term) 653 ## \$a (Index term- uncontrolled) e.g. 650 04 Plantes \$xMalalties i plagues	<dc:subject> Subject: La matèria que correspond al contingut del recurs. </dc:subject>
520 ## \$a (summary, etc., note) e.g. 520 ## \$a This book is mainly directed towards postgraduate students and professionals in the field of research and implementation of integrated pest and disease management programmes in greenhouse crops. The book originated from an international course taught at the International Centre for Advanced Mediterranean Agronomic Studies in Zaragoza, Spain. The authors are specialists from universities, research institutions and companies in Europe, America, Asia, Africa and Oceania. 505 0# \$a (Formatted contents note)	<dc:description> Description: An account of the content of the resource. </dc:description>
786 0# \$n (Data source entry/ note)	<dc:source> Source: Un recurs relacionat del qual es deriva el

	recurs descrit. </dc:source>
546 ## \$a (Language note) e.g. 546 ##\$a English	<dc:language> Language: llengua/idioma del contingut intel·lectual del recurs. </dc:language>
787 1# \$n (Nonspecific relationship entry/note)	<dc:relation> Relation: Referència a un recurs relacionat </dc:relation>
500 ## \$a (General note) 522 ## \$a (Geographic coverage note) e.g. 500 ## \$a A la portada: CIHEAM; e.g. 504 ## \$a Índex.	<dc:coverage> Coverage: L'extensió o abast del contingut del recurs </dc:coverage>
100 1# \$a (Main entry- personal name) 700 1# \$a (Added entry- personal name) e.g. 700 1# \$a Albajes Garcia, Ramon	<dc:contributor> Contributor: Una entitat que ha contribuït amb l'elaboració del contingut del recurs.</dc:contributor>
710 2# \$a (Added entry- corporate name) 711 2# \$a (Added entry- conference name) 720 ## \$a (Added entry-uncontrolled name/name) e.g. 710 2# Centre international de hautes études agronomiques méditerranéennes	[no es contempla]
260 ## \$b (Publication, description, etc. (Inprint)/ Name of publisher, distribution, etc.) e.g. 260 ## \$a Dordrecht [etc.]:\$b Kluwer Academic Publishers,\$c 1999	<dc:publisher> Publisher: Entitat responsable de fer possible l'accés del recurs. </dc:publisher>
100 1# \$a (Main entry- personal name) 700 1# \$a (Added entry- personal name) 710 2# \$a (Added entry- corporate name) 711 2# \$a (Added entry- conference name) 720 ## \$a (Added entry-uncontrolled name/name) e.g. 700 1# \$a Albajes Garcia, Ramon; e.g. 700 2# \$a Centre international de hautes études agronomiques méditerranéennes	<dc:creator > Creator: Entitat responsable de crear el contingut del recurs.</dc:creator >
540 ## \$a (Terms governing use and reproduction note) 856 42 \$3 (Rights)	<dc:rights> Rights: informació sobre els drets del recurs </dc:rights>
260 ## \$c (Date of publication, distribution etc.) e.g. 260 ## \$c 1999	<dc:date> Date: la data associada amb un esdeveniment del cicle de vida del recurs (data de creació, de publicació, etc.). La pràctica recomanada és utilitzar un esquema de codificació, com ara el perfil de W3CDTF ISO 8601 [W3CDTF].</dc:date>
655 #7 \$a (Index term- genre/form)	<dc:type> Type: The nature or genre of the content of the resource. </dc:type>
856 ## \$q (Electronic format type)	<dc:format xsi:type="dcterms:IMT"> Format:

	The physical or digital manifestation of the resource (per exemple: text/html)</dc:format>
856 4# \$u (Electronic location and access/ URL) e.g. 856 4# \$zAccés restringit als usuaris de la UB, UAB, UPC, UPF, UdG, UdL, URV, UOC, BC, UJI, UVIC http://ebooks.springerlink.com/UrlApi.aspx?action=summary&v=1&bookid=78492	<dc:identifier xsi:type="dcterms:URI"> Identifier: Una referència ambigua del recurs que l'identifica com a un element únic i que està associat amb un context donat. Exemple: http://hdl.handle.net/----/ </dc:identifier>
Altres elements no contemplats en DC simple:	Equivalència però amb Dublin Core Qualificat
300 545 p.	<dcterms:extent>545 p.</dcterms:extent> </dcterms:extent>
830 0 Developments in plant pathology ; v14	[no es contempla]

Taula 12. Comparativa entre Dublin Core (XML) i Marc21

Dublin Core s'expressa amb l'esquema XML del W3C, de manera que la forma de presentació de l'estructura de les dades és molt similar a la representada en la descripció de documents per MARC21. No obstant això, les etiquetes que representen els 15 elements bàsics de Dublin Core mostren que ofereix un esquema molt més simple si es compara amb l'estructura de dades de MARC21. Amb aquest exemple també s'observa la dificultat que hi ha amb l'equivalència de camps, ja que MARC21 contempla molts més elements de descripció, per tant, l'especificitat i el nivell de recuperació de la informació és molt major. De fet el format Dublin Core és usat per recursos web que no tenen la mateixa complexitat d'un registre bibliogràfic, i qualsevol persona pot publicar informació al WWW.



Dipòsit de la Recerca de Catalunya

[Pàgina inicial de RECERCAT >](#)
[Centres i instituts de recerca >](#)
[IRTA, Recerca i Tecnologia Agroalimentàries >](#)
[22 - Nitroso Animal >](#)
[Congressos i simposiums >](#) Document

Empreu sempre aquest identificador per citar o enllaçar aquest document <http://hdl.handle.net/2072/4544>

Registre complet

Camp Dublin Core	Valor
dc.contributor	IRTA, Recerca i Tecnologia Agroalimentàries
dc.contributor.author	Esteve Garcia, Enric
dc.contributor.author	García Regueiro, José Antonio
dc.date.accessioned	2007-10-26T12:16:37Z
dc.date.available	2007-10-26T12:16:37Z
dc.date.created	2005
dc.date.issued	2005-05-23
dc.identifier.citation	Esteve-Garcia, E.; García Regueiro, J.A. "Egg quality: chemical residues in respect to food safety". En: Xith European Symposium on the Quality of Eggs and Egg Products. Doorwerth, The Netherlands, 23-26 May, 2005, p. 154-163
dc.identifier.uri	http://hdl.handle.net/2072/4544
dc.format.extent	26 p.
dc.format.extent	322048 bytes
dc.format.mimetype	application/vnd.ms-powerpoint
dc.language.iso	eng
dc.rights.uri	Aquest document està subjecte a una llicència d'ús de Creative Commons, amb la qual es permet copiar, distribuir i comunicar públicament l'obra sempre que se'n citin l'autor original i l'IRTA i no se'n faci cap ús comercial ni obra derivada, tal com queda estipulat en la llicència d'ús (http://creativecommons.org/licenses/by-nc-nd/2.5/es/)
dc.subject.other	Ous -- Indústria i comerç -- Control de qualitat
dc.subject.other	Ous -- Qualitat -- Congressos
dc.subject.other	Seguretat alimentària
dc.title	Egg quality: chemical residues in respect to food safety
dc.type	jsp.search.advanced.type.pap???

Apareix a la col·lecció: [Congressos i simposiums](#)

Figura 16. Exemple de Dublin Core (Font: RECERCAT)

Figura 17. Exemple MARC 21 amb sortida de dades Millenium (Font: CCUC)

Interoperabilitat en Dublin Core

Dublin Core és un estàndard que inclou dos nivells de descripció, un nivell simple (exemple 2) i un altre nivell de descripció més avançat el Dublin Core qualificat (*qDC* o *DCMI Metadata Terms*). El nivell simple és la versió més utilitzada i eficaç per descriure una gamma de recursos àmplia en xarxa d'una forma fàcil. La versió simple, com s'ha comentat, inclou 15 elements i el qualificat inclou tres elements addicionals (públic, procedència i titular de drets), així com un grup de millores d'elements (o d'elements de classificació) que refinen la semàntica dels elements en formes que poden ser útils en el descobriment de nous recursos en un sistema d'informació. Dublin Core també és defineix com a "petit llenguatge usat per realitzar una classe particular de declaracions sobre els recursos". En aquest llenguatge trobem dues classes de termes, elements (noms) i qualificadors (adjectius), que poden ser ordenats en un patró simple dels estats. Els mateixos recursos són els subjectes implicats en el propi llenguatge. La semàntica de Dublin Core s'ha establert per un grup internacional i multidisciplinari de professionals de biblioteconomia, ciències de la computació, de codificació de textos, museòlegs i altres camps relacionats.

Però el que cal considerar en aquest capítol és la interoperabilitat que aquest tenir amb altres esquemes de metadades, que és en definitiva el que ens permetrà l'intercanvi d'informació en un sistema federat. Els nivells d'interoperabilitat de DC (DCMI versió 1.1., juny 2012) són prou elevats o

adequats tant pel que fa a la descripció (compartir vocabularis formals i limitacions dels registres) com a la sintaxi (compartir vocabularis formals en registres intercanviables), a la semàntica (compartir vocabularis basats en semàntica formal) i a la terminologia (compartir vocabularis definits en llenguatge natural). En aquest sentit no és comparable amb el format MARC.

DCMI Metadata Terms (*Dublin Core Metadata Initiative -- elements, element refinements, encoding schemes, and vocabulary terms*) han estat descrits com a metadades compatibles amb els principis de Linked Data (web semàntica). El primer nivell d'interoperabilitat (terminològic) es dona entre les metadades usades en les aplicacions basades en llenguatge natural. En el segon nivell d'interoperabilitat (semàntic) es basa entre les metadades usades en les aplicacions i a compartir un model formal que prové del RDF el qual és suportat per la Linked Data. De fet, SKOS és un tipus de llenguatge RDF i, per tant, facilita l'enllaç entre dades, informació i coneixement, ja que a través de l'ús de URI (adreces web) i RDF és possible una web semàntica o *Linked Data*. També el nivell 3 (sintàctic) i el 4 (descripció) són compatibles amb el model *Linked Data*. El tercer nivell fa que les aplicacions siguin compatibles amb el model de dades enllaçades a la "Descripció del conjunt". El quart nivell fa referència als registres d'intercanvi de metadades que usen les aplicacions; a més a més d'un conjunt de limitacions, utilitzen el mateix vocabulari i reflecteixen un model compartit.

Termes DCMI	Definició
Term Name: abstract	A summary of the content of the resource.
Term Name: accessRights	Information about who can access the resource or an indication of its security status.
Term Name: accrualMethod	The method by which items are added to a collection.
Term Name: accrualPeriodicity	The frequency with which items are added to a collection.
Term Name: accrualPolicy	The policy governing the addition of items to a collection.
Term Name: alternative	An alternative name for the resource
Term Name: audience	A class of entity for whom the resource is intended or useful.
Term Name: available	Date (often a range) that the resource became or will become available.
Term Name: bibliographicCitation	A bibliographic reference for the resource.
Term Name: conformsTo	An established standard to which the described resource conforms.
Term Name: contributor	An entity responsible for making contributions to the resource
Term Name: coverage	The spatial or temporal topic of the resource, the spatial applicability of the

	resource, or the jurisdiction under which the resource is relevant.
Term Name: created	Date of creation of the resource
Term Name: creator	An entity primarily responsible for making the resource.
Term Name: date	A point or period of time associated with an event in the lifecycle of the resource.
Term Name: dateAccepted	Date of acceptance of the resource
Term Name: dateCopyrighted	Data del Copyright
Term Name: dateSubmitted	Data de presentació del recurs
Term Name: description	An account of the resource. Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.
Term Name: educationLevel	A class of entity, defined in terms of progression through an educational or training context, for which the described resource is intended.
Term Name: extent	The size or duration of the resource
Term Name: format	The file format, physical medium, or dimensions of the resource. Recommended best practice is to use a controlled vocabulary, such as the AGROVOC thesaurus
Term Name: hasFormat	A related resource that is substantially the same as the pre-existing described resource, but in another format.
Term Name: hasPart	A related resource that is included either physically or logically in the described resource.
Term Name: hasVersion	A related resource that is a version, edition, or adaptation of the described resource.
Term Name: identifier	An unambiguous reference to the resource within a given context.
Term Name: instructionalMethod	A process, used to engender knowledge, attitudes and skills, that the described resource is designed to support.
Term Name: isFormatOf	A related resource that is substantially the same as the described resource, but in another format.
Term Name: isPartOf	A related resource in which the described resource is physically or logically included.
Term Name: isReferencedBy	A related resource that references, cites, or otherwise points to the described resource.
Term Name: isReplacedBy	A related resource that supplants, displaces, or supersedes the described resource.
Term Name: isRequiredBy	A related resource that requires the

	described resource to support its function, delivery, or coherence.
Term Name: issued	Date of formal issuance (e.g., publication) of the resource.
Term Name: isVersionOf	A related resource of which the described resource is a version, edition, or adaptation.
Term Name: language	Recommended best practice is to use a controlled vocabulary such as RFC 4646 [RFC4646].
Term Name: license	A legal document giving official permission to do something with the resource.
Term Name: mediator	An entity that mediates access to the resource and for whom the resource is intended or useful.
Term Name: medium	The material or physical carrier of the resource.
Term Name: modified	Date on which the resource was changed.
Term Name: provenance	A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation.
Term Name: publisher	An entity responsible for making the resource available.
Term Name: references	A related resource that is referenced, cited, or otherwise pointed to by the described resource.
Term Name: relation	A related resource.
Term Name: replaces	A related resource that is supplanted, displaced, or superseded by the described resource.
Term Name: requires	A related resource that is required by the described resource to support its function, delivery, or coherence
Term Name: rights	Information about rights held in and over the resource.
Term Name: rightsHolder	A person or organization owning or managing rights over the resource.
Term Name: source	A related resource from which the described resource is derived
Term Name: spatial	Spatial characteristics of the resource.
Term Name: subject	The topic of the resource.
Term Name: tableOfContents	A list of subunits of the resource
Term Name: temporal	Temporal characteristics of the resource
Term Name: title	A name given to the resource.
Term Name: type	The nature or genre of the resource.
Term Name: valid	Date (often a range) of validity of a resource.
Term Name: identifier	An unambiguous reference to the resource within a given context.

Term Name: language	A language of the resource.
Term Name: publisher	An entity responsible for making the resource available.
Term Name: relation	A related resource.
Term Name: rights	Information about rights held in and over the resource.

Taula 13. Elements Qualified Dublin Core (qDC). (Font: DCMI)

4.2.1.3. Altres estàndards i formats d'estructures de dades en entorns digitals

Segons els resultats obtinguts amb les enquestes fetes a les xarxes de biblioteques especialitzades en ciències aquàtiques i marines, l'únic estàndard d'estructura de dades anomenat i utilitzat és MARC o Dublin Core, però en aquest apartat es mostraran altres estàndards utilitzats i estesos en el camp de les biblioteques digitals, com el format MODS (*Metadata Object Description Schema*), METS (*Metadata Encoding & Transmission Standard*), que s'orienta a l'empaquetament d'objectes digitals (utilitzat en les metadades descriptives i administratives creades a partir d'altres esquemes) o el MADS (*Metadata Authority Description Standard*), aquest darrer format s'exclou de l'anàlisi perquè no és un format de descripció bibliogràfica o d'objectes digitals. MADS és un llenguatge de marcatge XML addicional al MODS per a la descripció de registres d'autoritats en MARC21.

4.2.1.3.1. MODS (vegeu *MODS user guide*)

És un esquema de metadades per descriure objectes bibliogràfics, que es poden utilitzar amb diferents propòsits: com a format especificat de SRU (servei web basat en el protocol Z.39.50), com a extensió de l'esquema METS, com a recollidor de metadades, etc., però en aquest cas s'analitza com a format de descripció d'objectes digitals en aplicacions bibliotecàries. Aquest estàndard es va desenvolupar al 2002 i és mantingut per la Xarxa de Desenvolupament i Oficina de normes MARC de la Library of Congress juntament amb els usuaris del format. MODS es descriu com a format de descripció d'esquemes de metadades d'objectes i com a llenguatge de marcatge XML per descriure registres MARC21. Per tant, és un esquema de metadades XML dissenyat per ser capaç de portar les dades seleccionades a partir de registres MARC21 existents i per permetre la creació de registres originals per descriure recursos. Inclou un subconjunt de camps MARC i utilitza un llenguatge basat en el llenguatge de marcat en lloc d'un valor numèric com en MARC, en alguns casos, reagrupa elements del format bibliogràfic MARC21.

El desenvolupament i l'aplicació del format MODS és posterior al Dublin Core, de manera que no està tan estès com el DC. A més a més, l'àmbit d'aplicació se centra principalment en biblioteques d'àmbit públic (Missingham, 2004), principalment biblioteques nacionals, entre d'altres unitats d'informació que

segueixen les directrius marcades per la Library of Congress, on l'ús del format MARC és també més estès, ja que MODS representa una evolució i/o millora per a la descripció d'objectes digitals. Certament, el desenvolupament de MODS és una iniciativa que sorgeix per cobrir les mancances de Dublin Core. Tot i que la versió de Dublin Core qualificat està molt més desenvolupada que la versió simplificada, Dublin Core no representa completament un registre bibliogràfic com pot ser representat en format MARC. Així doncs, la Library of Congress ha desenvolupat amb MODS un subtipus de MARC que s'adapta als sistemes d'informació, s'usa per descriure objectes digitals i, al mateix temps, és compatible amb les descripcions ja existents al catàleg. Convé destacar que és un sistema més flexible, però no per això més complicat, que el format MARC. Com ja apuntava Guenther, R. el 2003, una de les desenvolupadores de MODS és en aquest segle on ens podem plantejar si realment el format MARC21 ha quedat obsolet o no, i que l'evolució cap al XML és un ordre lògic; de fet, MODS es pot entendre com una alternativa entre un format de metadades simple (Dublin Core) i un format molt detallat amb molts tipus de dades amb diferents complexitats estructurals, com el MARC21. MODS té un grau alt de compatibilitat amb registres MARC perquè hereta els elements de metadades de la semàntica del MARC. Algunes dades han estat reagrupades i, en alguns casos, diferents elements de dades en MARC poden ser agrupats en un mateix element a través de MODS. A més, MODS no assumeix l'ús de cap codi específic de catalogació. Diversos elements tenen un atribut ID opcional per facilitar la coordinació de l'element.

Avantatges: MODS intenta ser complementari a altres formats de metadades; té els següents avantatges:

- Ofereix elements més rics que Dublin Core.
- Els tipus d'elements són més compatibles amb les dades contingudes en sistemes bibliotecaris.
- L'esquema està més orientat a l'usuari final que l'esquema complet MARCXML.
- El tipus d'element és més simple que el format complet MARC i per això més senzill d'aplicar.

Desavantatges: MODS inclou un subtipus de dades provinents del MARC 21 *Format for Bibliographic Data* (MARC 21, núm. 12, 2012). A més, com a conjunt d'elements que permet la representació de les dades en sistemes basats ja en MARC, permet la conversió dels camps bàsics; per contra, algunes dades específiques poden no ser incloses.

Un altre inconvenient és que l'esquema MODS no funciona per a MARC21. Concretament, un registre original en MARC21 convertit a MODS, després no

es pot convertir de nou a MARC21 en la seva totalitat sense una pèrdua d'especificitat en l'etiquetatge o pèrdua de dades. En alguns casos, encara que es pugui reconvertir en MARC21, les dades no poden ser col·locades exactament en el mateix camp que a MARC perquè es pot haver assignat a un altre camp més general. No obstant això, aquests inconvenients no fan que es perdin dades, només hi pot haver pèrdues en la identificació detallada del tipus d'element que representa. En altres casos l'element MARC no pot tenir un element equivalent a MODS i, en conseqüència, les dades específiques es podrien perdre en la conversió. Les directrius de MODS no indiquen que la conversió amb MARC no sigui possible, sinó que no és directa com passa amb el format MARC i que segurament implica conversions addicionals específiques que no són del tot pràctiques.

4.2.1.3.2. METS (*Metadata, Encoding and Transmission Standard*)

És un estàndard mantingut també per la Library of Congress i desenvolupat com a una iniciativa de la *Digital Library Federation* (DLF), expressat mitjançant el llenguatge d'esquema XML i creat per codificar les metadades descriptives, administratives, estructurals i tècniques, entre d'altres, que fan referència a la descripció d'objectes d'una biblioteca digital (Wendler, R., 1999). METS és definit com a especificació per a l'intercanvi i emmagatzematge; tanmateix, conté les metadades bàsiques per preservar objectes digitals (metadades administratives, descriptives, estructurals i tècniques que fan referència als arxius físics). En canvi, l'esquema de metadades definit com a estàndard de metadades per a la preservació és el PREMIS (PREMIS, 2009), el qual pretén especificar les unitats semàntiques necessàries per donar suport a les funcions bàsiques de conservació (*Digital Preservation Metadata Standards*, 2010). Cal afegir, que aquest tipus de metadades no són objecte d'estudi en aquesta recerca i és per això que no s'han comparat amb la resta de formats representats. No obstant això, la interoperabilitat amb els formats estàndards MODS i METS és possible (Habing, 2007).

Pel que fa a un document METS, cal tenir en compte que s'estructura en set parts bàsiques:

1. Capçalera: conté metadades que descriuen el document, com per exemple dades sobre l'autor, editor, etc.
2. Metadades descriptives: es refereix a metadades descriptives externes al document METS; però si el propi document conté metadades descriptives també s'inclourien en aquesta secció
3. Metadades administratives: poden ser internes (incrustades al document) o externes, i proporcionen informació sobre com el document que s'ha creat o emmagatzemat, els drets de propietat intel·lectual,

informació sobre la font original de l'objecte descrit, informació sobre procedència dels arxius compresos a la biblioteca digital, etc.

4. Secció d'arxius: es fan llistes dels arxius que tenen contingut sobre versions electròniques de l'objecte digital. Inclou el mapa estructural (apartat nucli del document METS) que descriu l'estructura jeràrquica de l'objecte de la biblioteca digital i enllaça els elements d'aquesta estructura amb els arxius que contenen les metadades pertanyents a cada element.
5. Enllaços estructurals: apartat que permet als creadors del document METS arxivar hipervincles dirigits als nusos del mapa estructural.
6. Comportament: aquesta secció s'utilitza per associar comportaments executables amb el contingut de l'objecte METS. Cada comportament té un element que identifica un mòdul de codi executable que implementa i executa els comportaments definits d'una manera abstracta per la interfície de definició.

4.2.1.4. Anàlisis i comparativa entre els formats

De manera global, els formats exposats en aquest capítol es diferencien bàsicament pel seu ús i/o objectiu pel qual van ser creats. Els formats MARC, MODS o MADS són definits per la *Library of Congress* com a formats de descripció de recursos digitals (MODS i MADS específicament per a la descripció d'objectes digitals, basat en la comunitat bibliotecària tradicional que utilitza MARC). En canvi, el format METS és descrit com a estàndard de codificació de metadades i especificacions per a la transmissió de metadades que permet d'una forma robusta i flexible la definició d'objectes digitals. Altrament, el format PREMIS proporciona maneres de descriure objectes i processos que són essencials per a la preservació digital, la qual cosa fa que es descarti aquest format en aquest capítol, ja que la preservació d'objectes digitals no és objecte d'estudi. No obstant això, l'interès principal d'aquests estàndards està amb les seves capacitats d'interoperabilitat, i en aquest sentit, cal remarcar que totes aquestes especificacions de metadades estan construïdes sobre XML, per tant, existeix interoperabilitat quant a la seva estructura de dades. A més, les seves comunitats d'usuaris i enfocaments subjacents també tenen molt en comú i cal destacar també que el contingut d'aquests formats és de tipus independent, la qual cosa fa possible que puguin ésser compatibles i coexistir en un mateix sistema, com per exemple és el cas del projecte *eJournal Archival Information Package* o AIP de la *British Library* (Dappert i Enders, 2008).

A continuació es presenta una taula que mostra les possibles equivalències entre els tres formats destacats com a formats de descripció o d'estructura de

dades. A posteriori s'avaluen les equivalències dels diferents elements de metadades que conformen els formats, partint del model de metadades simple de Dublin Core, amb la finalitat de veure el nivell d'interoperabilitat existent entre els estàndards de descripció bibliogràfica més utilitzats davant d'un possible mapeig de metadades (*crosswalk*) entre diferents sistemes d'informació.

DUBLIN CORE SIMPLE	MODS	METS ³¹
<p><title></p> <p>Títol: nom pel que formalment es coneix el recurs</p>	<p><titleInfo><title>.</p> <p>MODS permet subelements <title info> els quals poden ser parsejats (extreure els elements de forma individual) amb els següents elements: <nonSort>, <title>, <subtitle>, <partNumber> i subelements <partName></p>	<p><metsHdr></p> <p>Encapçalament: anomenat <i>METS header</i> conté informació relativa a la creació del document METS (nom del arxiu, data de creació i modificació del mateix, nom del responsable de la seva creació, etc.)</p>
<p><creator></p> <p>Creador: persona o entitat responsable de la creació del recurs o la versió del mateix, etc.</p>	<p><name><namePart></p> <p>També permet incloure el tipus de contribució del creador entre d'altres elements complementaris com la descripció, afiliació, etc. Per exemple:</p> <pre><name type="personal"> <namePart type="termsOfAddress">Ms.</na mePart> <namePart>Reverté, C. </namePart> <affiliation>IRTA</affiliation> <role> <roleTerm type="text" authroity="marcrelator">creator</ roleTerm></role></name></pre>	<p><MetsHdr></p> <p>L'encapçalament és on s'inclou el nom del responsable de la creació de l'objecte digital, etc.</p>
<p><subject></p> <p>Matèria: tema del que tracta el recurs.</p>	<p><subject> i <classification></p> <p>L'element <subject> inclou els subelements:</p> <p><topic> <name> <ocupation></p>	<p><mdRef> [metadades descriptives externes] <mdWrap> [metadades descriptives internes] Metadades descriptives: a) Apunta a metadades descriptives externes al document METS (per exemple, un registre MARC en</p>

³¹ No s'està fent un mapeig exacte, sinó mostrant una equivalència aproximada dels elements. Tal i com s'observa, hi ha alguns tipus de metadades METS (capçalera/creator) que poden encaixar en diferents elements de DC i MODS.

		un OPAC o un document EAD disponible en un servidor web); b) conté internament les metadades descriptives, o c) combinar ambdues aproximacions.
<Description> Descripció: text lliure on s'explica el contingut del recurs.	<abstract> <note> <tableOfContents>	<mdRef> [metadades descriptives externes] <mdWrap> [metadades descriptives internes]
<Publisher> Editor: entitat responsable de la publicació del recurs.	<originInfo><publisher>	<metsHdr> Nom del responsable de la seva creació, etc.
<contributor> Col·laborador: persona o entitat amb responsabilitat parcial en la creació del recurs.	<name><namePart> Aquerst element és com l'element "creator", però no té una codificació diferent, simplement canvia el 'rol' de l'autoritat/tipus de responsabilitat: <role> <roleTerm type="text" authroity="marcrelator"> contributor </roleTerm></role>	<metsHdr> Nom del responsable de la seva creació, etc.
<date> Data: data de creació o publicació del recurs o altres dades associades al seu cicle de vida.	<originInfo><dateIssued> <originInfo><dateCreated> <originInfo><dateCaptured> <originInfo><dateOther>	<metsHdr> L'element d'encapçalament incorpora també la data de creació i modificació del mateix, nom del responsable de la seva creació, etc.
<type> Tipus de recurs: naturalesa del recurs, en funció del contingut.	<typeOfResource> <genre>	<behaviorSec> o bé, <mdRef> [metadades descriptives externes] <mdWrap> [metadades descriptives internes]
<format> Format: naturalesa del recurs, en funció de les seves característiques tècniques.	<physicalDescription> Inclou els subelements: <form> <reformattingQuality> <internetMediaType> <extent> <digitalOrigin> <note>	<behaviorSec> Comportament: associa comportaments d'execució al contingut del document METS, de tal manera que, per exemple, sigui possible especificar el tipus d'aplicació necessària per utilitzar el recurs o per indicar que el recurs requereix una determinada fulla d'estil per a ser mostrat correctament.

<p><identifier> Identificador: referència per a la identificació inequívoca del recurs (URI, URL, DOI, etc.)</p>	<p><identifier> <location><uri></p>	<p><amdSec> Metadades administratives que inclouen quatre subcomponents: 1) metadades tècniques, 2) drets d'accés i ús (<i>rights metadata</i>), 3) origen (<i>source metadata</i>), i 4) metadades de conservació (<i>preservation metadata</i>). Al igual que el component de metadades descriptives, poden fer referència a un registre extern o bé conformar un paquet de metadades intern.</p>
<p><source> Font: fa referència a l'<i>identifier</i> del recurs descrit. Veure <relation element></p>	<p><relatedItem type="original"> + <titleInfo><title> or <location><url></p>	<p><amdSec> Metadades administratives de tipus 3 (<i>source metadata</i>)</p>
<p><language> Idioma: idioma o idiomes utilitzats en el recurs.</p>	<p><language> (els elements són equivalents per al mapeig)</p>	<p><mdRef> [metadades descriptives externes] <mdWrap> [metadades descriptives internes]</p>
<p><relation> Relació: referència al identificador del recurs o recursos amb els que està relacionat el recurs descrit.</p>	<p><relatedItem> més <titleInfo><title> o <location><url></p>	<p><fileSec> Directori d'arxius o <i>file inventory</i> permet comptar amb una llista d'arxius associats a un document i les seves relacions. Els arxius que el formen poden ser enllaçats o poden estar continguts en el registre com un codi binari.</p>
<p><coverage> Cobertura: cobertura espacial, temporal o jurisdiccional associada al contingut del recurs.</p>	<p><subject><temporal> <subject><geographic> <subject><hierarchicalGeographic> <subject><cartographics></p>	<p><mdRef> [metadades descriptives externes] <mdWrap> [metadades descriptives internes]</p>
<p><Rights> Drets: dades relatives al règim de protecció de drets d'autor que afecten al ús del recurs descrit.</p>	<p><accessCondition> (és un element equivalent a <i>Rights</i>)</p>	<p><amdSec> Metadades administratives: al igual que el component de metadades descriptives, pot fer referència a un registre extern o bé conformar un paquet de metadades.</p>
<p>No existeix l'element</p>	<p>RecordInfo (Dades administratives del recurs: data de creació, modificació, etc.)</p>	<p><admSec> [Metadades administratives] <fileSec> Directorio d'archivos o "<i>file inventory</i>": permet incloure una llista d'arxius associats a un document i les seves</p>

		relacions. Els arxius que comprenen poden ser enllaçats o poden estar continguts en el registre com un codi binari.
No existeix l'element	No existeix l'element	<structMap> El mapa estructural descriu l'estructura d'un recurs, presentant una organització jeràrquica dels registres de metadades i els continguts als que es refereixen. Aquest mapa fomenta el desenvolupament d'instruments que faciliten la navegació dels usuaris en col·leccions extenses de documents.
No existeix l'element, encara que s'aproxima a l'element <Relation>	No existeix l'element, encara que s'aproxima a l'element <RelatedItem>	<structLink> Enllaços estructurals: especialment útil per descriure llocs web, pretén recollir hiperenllaços referents a divisions indicades en el mapa estructural.

Taula 14. Interoperabilitat entre els formats d'estructura de dades DC, MODS i METS (Font: DCMI i *Library of Congress*)

Notes en relació al mapeig del format MODS a Dublin Core i viceversa:

Element Títol: l'element no és variable i el contingut es idèntic.

Element Creador: MODS a diferència de DC descriu el tipus de rol al que fa referència l'element "creador", a més, els elements són repetibles. També permet crear subelements de <name>: <namePart>, <displayForm>, <affiliation>, <role>, <description>.

```
<name type="personal">
  <namePart
type="given">Ash</namePart>
  <namePart
type="family">Amin</namePart>
  <role><roleTerm
type="text">author</roleTerm></role>
</name>
```

Taula 15. Exemple de l'element Creador (Font: Elaboració pròpia)

Amb Dublin Core no es poden especificar els elements com en MODS, per tant, tenim informació que és perd i el mateix succeeix amb l'element matèria (*Subject*).

Element Matèria: MODS permet incloure elements més específics o subelements que poden fer referència a diferents vocabularis controlats, el mateix passa amb Dublin Core i l'element "Subject". Cal afegir que MODS inclou un element específic per "Classificació" separat de l'element "Subject", i, en canvi, en DC el mateix element es troba dintre de "Subject". Els subelements MODS específics per l'element "Subject" fan referència al tema del document/objecte digital, àmbit geogràfic, època, coordenades, etc. (<topic>, o <geographic>, <temporal>, <coordinates>, <name>, <titleInfo>, <hierarchicalGeographic>, etc.).

En principi no hauria d'haver pèrdua d'informació si el mapeig es produís amb els mateixos tipus de vocabularis, altrament, el mapeig no seria tan directe ni tan fàcil. Avui dia trobem molts exemples de projectes internacionals que estan fent mapejos entre diferents vocabularis (NAL-AGROVOC-LCSH, etc.), Però aquests són convertits prèviament a llenguatges ontològics (SKOS o OWL).

Element Descripció: en Dublin Core hi ha informació que ve del "text lliure", en canvi amb MODS s'inclou la descripció en altres tipus d'elements més definits como poden ser el camp resum, notes o contingut ([<abstract>](#), [<note>](#) i [<tableOfContents>](#)). Així doncs, aquesta informació es perdria en la conversió de DC a MODS, tot i que cal recordar que Dublin Core qualificat (qDC) sí té un element de refinament que es refereix al resum ("E: description"; "ER: abstract") i un altre element per descriure la taula de continguts "ER:TableOfContents".

Element Editor: DC inclou dades de publicació del registre on apareix la "date" i l'"editor", per contra, MODS en només s'obté informació de l'editor. Per tant, pot haver pèrdua d'informació quan es mapeja aquest camp de DC a MODS.

Element Data: en MODS hi ha diferents elements que es refereixen a l'àrea d'informació i que en DC només es defineix en un element, "date", que es refereix a la data de creació o publicació del recurs o altres dates associades al seu cicle de vida, però no hi ha els subelements que existeixen en MODS.

Element Tipus de recurs: En DC només hi ha l'element "Type" que fa referència al tipus de recurs descrit, que amb MODS equival <typeOf Resource>, però a més a més hi ha una altre element de major especificitat que proporciona MODS, <genre>, que designa una categoria que caracteritza un determinat estil, forma o contingut (artística, composició musical, literari, etc.) a un recurs.

Element Format: un altre cop més MODS proporciona més subelements per descriure l'element "Format", **<physicalDescription>** i amb DC només hi ha l'element <format>.

<physicalDescription>	Definició
<u><form></u>	Designació de la presentació física del recurs.
<u><reformattingQuality></u>	Indica avaluació de la qualitat física d'un recurs electrònic en relació al seu ús (Per exemple: accés, preservació, reemplaçament, etc.)
<u><internetMediaType></u>	Identificació del tipus de format electrònic, o de representació de les dades del recurs. Per exemple: text/html, ASCII, Postscript file, executable application or JPEG image
<u><extent></u>	Extensió física del recurs
<u><digitalOrigin></u>	Designa els antecedents del recurs. La procedència del fitxer digital d'on prové el recurs en relació a la seva creació, ús i gestió.
<u><note></u>	Informació textual sobre la descripció física del recurs.

Taula 16. Subelements de l'element de descripció del format (Font: MODS; Library of Congress)

Element Identificador: MODS diferencia entre "identifier", que és l'identificador del recurs i l'element idèntic a DC, i l'element "location" com a element que indica l'ubicació del recurs. A més, MODS utilitza atributs específics dintre els elements per identificar els diferents tipus de recursos. Per exemple: <identifier type="uri">.

Element <relatedItem> amb l'atribut "original" (<relatedItem type="original">) és l'element equivalent per a MODS de l'element **<source>** de DC, és a dir, es refereix a un identificador de recurs del qual deriva el recurs descrit. MODS utilitza el mateix element **<related>** com a equivalent a l'element de DC **<relation>**, el qual fa referència a un *identifier* del recurs o recursos amb els que està relacionat el recurs descrit. En aquest sentit, MODS utilitza el mateix element però ofereix diferents atributs que serveixen per definir el que són elements diferents en DC. Tanmateix, l'element <relatedItem> en MODS pot utilitzar diferents [subelements](#) que afavoreixen la interoperabilitat entre recursos i que pel contrari no ofereix DC.

<relatedItem>
<u><titleInfo></u> ; <u><name></u> ; <u><typeOfResource></u> ; <u><genre></u> ; <u><originInfo></u> ; <u><language></u> ; <u><physicalDescription></u> ; <u><abstract></u> ; <u><tableOfContents></u> ; <u><targetAudience></u> ; <u><note></u> ; <u><subject></u> ; <u><classification></u> ; <u><relatedItem></u> ; <u><identifier></u> ; <u><location></u> ; <u><accessCondition></u> ; <u><part></u> ; <u><extension></u> ; <u><recordInfo></u>

Taula 17. Subelements de l'element "RelatedItem"

Element Coverage: tal i com s'observa en la taula 12, MODS utilitza diferents subelements de l'element **<subject>** per definir el tipus de cobertura (encara que no té subelement per definir la cobertura espacial), és a dir, en MODS no existeix un element específic equivalent a **<Coverage>**. Així doncs, es preveu que no hi haurà un correcte mapeig entre els elements de DC i MODS.

4.2.1.5 Resultats

Recollint tot el que s'ha dit, es determina que el format MODS té una complexitat major que Dublin Core. Així com METS és un format encara més complex que els anteriors descrits. Fonamentalment, aquesta complexitat es deu a que els objectius de creació dels mateixos formats són diferents i estan dissenyats per ser utilitzats en diferents contextos.

D'una banda, DC i MODS estan orientats cap a la descripció de recursos bibliogràfics. Tot i que DC originalment s'utilitzava per descriure recursos electrònics continguts en el web, aviat es va estendre i és avui dia utilitzat en la comunitat bibliotecària en general, i per això es utilitza també en la descripció de registres bibliogràfics. Pel que fa a l'aparició del format MODS, enfocat a la descripció de registres bibliogràfics, l'objectiu és oferir un model de metadades més complet que els habitualment utilitzats en el camp de les biblioteques, principalment formats MARC. D'altra banda, pel que fa al format METS, com és definit en la seva pròpia descripció, és un esquema de metadades utilitzat per objectes digitals que tenen l'objectiu de gestionar col·leccions més complexes (col·leccions impreses/digitals de biblioteques, museus, etc.). Conseqüentment, es necessita la seva aplicació perquè per a la gestió de diferents tipus de col·leccions es necessiten no només metadades descriptives sinó que també s'han d'utilitzar d'altres metadades de tipus administratiu, tècnic i estructural entre d'altres.

Es poden trobar moltes diferències entre METS i MODS/DC, però cal fer una especial referència a la importància de les metadades referents a l'accés, preservació i conservació dels documents i/o arxius:

- Metadades administratives: informació relativa al emmagatzematge, a la procedència dels arxius i les seves còpies, com també a la seva propietat intel·lectual.
- Arxiu: metadades que fan referència a la traçabilitat dels arxius i les seves versions.
- Metadades estructurals: que permeten registrar els hipervincles i arxivar-los, etc.
- Metadades de comportament: execució dels arxius, etc.

Entenent que l'evolució i aparició de diferents formats i estàndards poden causar dubtes o discussions sobre quins són els formats de metadades més adequats, no s'ha d'oblidar que en el disseny d'un sistema d'informació distribuït (col·lectiu) s'imposaran d'altres qüestions relacionades amb la popularitat (coneixença i ús de formats, normalment estàndards), la sostenibilitat (aprofitar recursos, eines i formats existents, i minimització de costos) i la interoperabilitat (compartir i intercanviar informació i coneixement sense construir un nou sistema, ús de processos de mapeig entre d'altres). En síntesi, i seguint una de les premisses d'aquesta tesi, cal perseguir un sistema d'informació basat en l'ecologia de la informació (Davenport i Prusak, 1997), que sigui col·laboratiu i cooperatiu (Reverté, C. i Sebastià-Salat, M., 2009).

4.2.2. Estàndards i formats per a l'intercanvi d'informació i interoperabilitat entre sistemes

4.2.2.1. Semàntica

El segon grup d'estàndards es relacionen amb el web semàntic. Com s'ha comentat anteriorment, el web semàntic es basa en la idea d'afegir metadades de tipus semàntic i/o ontologies en el World Wide Web. Aquesta informació addicional permet la descripció i donar significat als continguts, i, contribueix a la relació de dades d'una forma més òptima i eficient del que fins ara proporciona la web sintàctica (representació de la informació basada amb els documents). Per exemple, els actuals cercadors web tot i ser cada cop més sofisticats, se centren en cercar informació per la seva popularitat i recuperen documents web que poden ser candidats a contenir el contingut desitjat per un usuari. El resultat és la provisió d'un llistat d'hiperenllaços que rarament donen resposta a preguntes específiques, i, a més, es recuperen documents web que són relacionats junts amb un hiperlink unidireccional (no hi ha una relació semàntica entre ells). Per contra, el web semàntic, pretén donar respostes complertes i rellevants que responguin a preguntes específiques formulades a un sistema d'informació. Atès que la semàntica aporta la descripció del significat d'un concepte, i aquesta especificitat permet que la "resposta" d'un sistema sigui rellevant a la "pregunta" formulada per un usuari. Actualment ja hi ha disponibles molts cercadors³² semàntics en funcionament. Un exemple, és el cercador intern de la web kukushumushu.com (dissenyat per ISOCO) amb tecnologia del web semàntic on es visualitza aquesta comprensió semàntica entre la interrogació d'un usuari i la capacitat de resposta d'un sistema semàntic. Per a veure un exemple pràctic del seu funcionament, s'interroga al sistema amb 2 preguntes que no tenen a veure amb els productes en venda de la tenda online.

³²Llista de cercadors semàntics publicada en la W3C. Url: http://www.w3.org/wiki/Search_engines

1.1 Pregunta: ¿Será el fin del mundo en el 2013?



1.2. Pregunta 2: ¿Terminará la crisis al 2013?

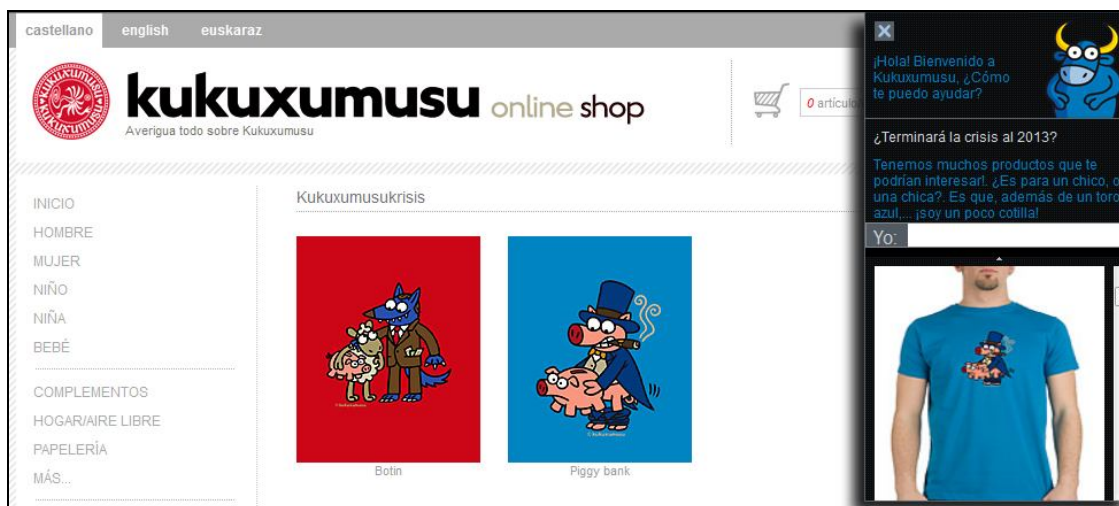


Figura 18. Exemple de interrogació del sistema (Font: kukuxumusu.com)

Les dues interrogacions realitzades anteriorment al cercador de kukuxumusu.com, el sistema dóna una resposta concreta i no un llistat de resultats amb soroll documental que és el que donaria un cercador convencional. En la primera pregunta no apareixen resultats però el cercador dóna una resposta: “por el momento la habilidad de adivinar no tengo...pero dales tiempo a mis colegas de Kukuxumusu”. El cercador ha analitzat la interrogació de forma semàntica i ha detectat que l'acció del verb era en futur, per això ha donat aquesta resposta, a més de saber que no té resultats per donar a l'usuari. En canvi, en la segona pregunta sí apareixen resultats, identifica el concepte crisis i apareixen dos productes del seu catàleg de camisetes de la línia kukuxumusukrisis que sí estan relacionats amb el concepte. Però a més, el cercador continua fent interrogacions a l'usuari per refinar la seva consulta: “tenemos muchos productos que te pueden interesar, es para un chico o una chica”.

L'objectiu bàsic del web semàntic és proporcionar un mecanisme per estructurar la informació distribuïda (en el web o en bases de dades, etc.) per finalment permetre el processament del contingut de forma més fàcil i eficient. La idea és proporcionar a la web dades ben definides, estructurades i enllaçades, facilitant que les aplicacions descobreixin aquestes dades, les

integren, les raonin i utilitzen o reutilitzen aquesta informació trobada al web. És per això que a través de llenguatges de marcatge estandarditzats per a la representació de la informació, com és el cas dels formats d'ontologies (OWL i SKOS) a través dels quals diferents sistemes poden comprendre's mitjançant un intercanvi d'informació semàntica (descripció del contingut a través d'ontologies). Aquests formats milloren doncs la interoperabilitat entre diferents sistemes digitals que comparteixen un llenguatge semàntic. Les ontologies, que s'originen en la disciplina de la intel·ligència artificial (AI), són considerats "agents intel·ligents" capaços de descriure la semàntica de forma explícita, però el veritable valor del web semàntic es dona quan un sistema o programari pot recollir i processar contingut web a partir de fonts diverses i intercanviar informació amb altres sistemes.

4.2.2.2 Estàndards per a la descripció de sistemes organització del coneixement (KOS)

Els sistemes d'organització del coneixement més habituals són: tesaurus, sistemes de classificació i llistes d'encapçalament de matèries.

Aquest apartat està orientat a l'estudi dels tesaurus, perquè han estat els sistemes d'organització del coneixement (KOS) més utilitzats en el camp de les biblioteques en ciències aquàtiques i marines, i també perquè com ja s'ha vist en anteriors capítols són vocabularis controlats heterogenis i multilingües que poden ocasionar més problemàtiques en un sistema d'informació distribuït d'aquest àmbit multidisciplinari i per això són cas d'estudi. A més, en l'actual web semàntica o web 3.0, els tesaurus són usats també com a Sistemes Simples d'Organització del Coneixement (SKOS). La web semàntica ha creat l'expectativa d'oferir uns sistemes avançats de recuperació de la informació on l'usuari es dirigeix a un agent/programari amb el seu propi idioma i l'agent respon de manera intel·ligent. La interpretació de la consulta es dona amb les transaccions d'informació que es realitzen entre altres agents que hi ha a la web, on a través de les "ontologies" (referents a temes específics o sinònims relacionats amb la consulta de l'usuari, relacions entre aquests termes i un conjunt de regles d'inferència) es facilita l'expansió de la consulta realitzada per l'usuari, però també ajuden a que un sistema sigui més precís i augmenti així el nivell de rellevància dels resultats obtinguts d'una consulta. Per aquest motiu, és també imprescindible estudiar i conèixer els formats del web semàntic i veure les possibilitats de convertir els tesaurus en llenguatges ontològics o ontologies terminològiques (Lassila, O. i McGuinness, D.L., 2001a). En la següent secció, es presenten els formats i estàndards essencials en el web semàntic (XML, RDF, OWL i SKOS), però en cap cas s'aprofundirà amb els aspectes tècnics. L'objectiu plantejat en l'estudi dels formats semàntics és donar una visió general d'aquests formats centrada amb la seva aplicació en

sistemes distribuïts i principalment en els sistemes d'organització del coneixement (KOS). Per tal d'ampliar informació cal consultar la web del W3C.

4.2.2.2.1. XML, RDF i RDF Schema

XML, RDF i RDF Schema (RDFS) són els estàndards clau per a el web semàntic. Tots ells han anat evolucionant per tal de millorar la interoperabilitat entre sistemes, oferint un intercanvi i representació de dades més flexible i eficient que els formats antecessors (SGML i HTML).

XML (*eXtensible Markup Language*) és un metallenguatge que prové del SGML (Llenguatge de Marcatge Generalitzat Estàndard), estàndard ISO 8879 que serveix per a la definició de vocabularis orientats per intercanviar dades estructurades. El seu desenvolupament ve per les carències de l'HTML, ja que aquest darrer no conté dades estructurades, només informació sobre fragments d'un document. La representació del format XML és per etiquetes, igual que HTML (<titulo>...</titulo>) però amb la gran diferència de que XML és obligatori l'ús d'una etiqueta d'obertura i una de finalització per tal de representar el contingut d'un document. Un document XML està format per un pròleg (consisteix en una declaració del XML i una referència opcional a documents externs d'estructuració), un nombre d'elements i opcionalment pot tenir un epíleg. Veure exemple de descripció XML amb DC de la Pàgina Principal de la EPA (*European Environment Agency*):

Pròleg

```
<?xml version="1.0"?>
<metadata
  xmlns="http://www.ukoln.ac.uk/metadata/dcdot/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

  xsi:schemaLocation="http://www.ukoln.ac.uk/metadata/dcdot
/ http://www.ukoln.ac.uk/metadata/dcdot/dcdot.xsd"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
```

Elements

```
<dc:title xml:lang="es" >
Agencia Europea de Medio Ambiente – Agencia Europea del
Medio Ambiente (AEMA)
</dc:title>
<dc:subject xml:lang="es"> Agricultura; Contaminación
atmosférica; Biodiversidad; Productos químicos; Cambio
climatic; Costas y mares; Energía; Medio ambiente y
salud; Tecnología medioambiental; Pesca; Economía verde;
Consumo de los hogares; Industria; Usos del suelo;
Recursos naturales; Ruido; Policy instruments; Suelo;
Regiones específicas;Turismo; Transporte; Medio ambiente
urbano; Residuos y recursos materiales; Agua
</dc:subject>
<dc:date>2012-11-19</dc:date>
<dc:type>Text</dc:type>
<dc:format>text/html; charset=utf-8</dc:format>
```

```
<dc:format>100765 bytes</dc:format>
<dc:identifier>http://www.eea.europa.eu/es
</dc:identifier>
</metadata>
```

Figura 19. Exemple de descripció XML amb DC

No obstant això, XML, tot i representar els continguts de forma estructurada, manca de mitjans per parlar de semàntica (significat) de les dades. Per exemple, nidificació d'etiquetes que proveïen d'estructura a un document, no té cap significat predefinit, de manera que cada aplicació té que interpretar la nidificació. Observem els següents exemples que descriuen amb XML la següent frase:

“European Environment Agency (EEA) és una agència de la Unió Europea”

Exemple 1:

```
<OrganismeInternacional="Unió Europea">
  <Agència Nom>European Environment Agency</Agència Nom>
</OrganismeInternacional>
```

Figura 20. Exemple 1 descripció de contingut amb XML

Exemple 2:

```
<Agència Nom="European Environment Agency">
  <OrganismeInternacional>Unió
  Europea</OrganismeInternacional>
</Agència Nom >
```

Figura 21. Exemple 2 descripció de contingut amb XML

Exemple 3:

```
<Tipus d'Organisme>
<OrganismeInternacional="Unió          Europea"></Organisme
Internacional>
  <Agència Nom>European Environment Agency</Agència Nom>
</Tipus d'Organisme >
```

Figura 22. Exemple 3 descripció de contingut amb XML

En els dos primers exemples, s'observa que es representa la mateixa informació però amb una nidificació dels elements de forma oposada. Per tant, no hi ha una forma estàndard per assignar el significat a la nidificació d'etiquetes.

Amb **RDF** (Infraestructura per a la Descripció de Recursos³³), que més que un llenguatge és un model de dades, s'intenta millorar aquesta manca de significat o semàntica que té XML. RDF es compon d'un bloc bàsic format per una tripleta d'elements, objecte-atribut-valor (subjecte–predicat–objecte), anomenada sentència (*statement*). El tercer exemple, exposat anteriorment,

³³ Traducció extreta de la W3C: *Resource Description Framework*(RDF): Especificación del Modelo y la Sintaxis. Eva Méndez (traductora). Traducción 3 enero de 2001. Url: <http://www.sidar.org/recur/desdi/traduc/es/rdf/rdfesp.htm#intro>

seria una bona representació, ja que hi ha una sentència on s'està definit el tipus d'organisme al que s'està descrivint que en format de graf podria tenir la representació següent:

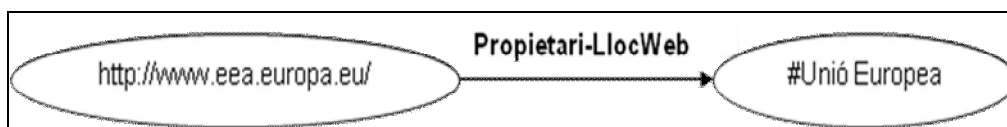


Figura 23. Exemple de graf simple

Aquest model de dades abstracte definit per RDF pot ser tan ampli com es desitgi i pot ser representat per XML, però també amb d'altres esquemes de representació de dades com és XML-Schema (xsls) i Dublin Core (DC):

```
<?xml version="1.0"?>
<!DOCTYPE                                rdf:RDF                                SYSTEM
"http://dublincore.org/documents/2002/07/31/dcmes-xml/dcmes-
xml-dtd.dtd">

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://www.eea.europa.eu/es">
    <dc:title>
      Agencia Europea de Medio Ambiente – Agencia Europea del
      Medio Ambiente (AEMA)
    </dc:title>
    <dc:subject>
      Agricultura; Contaminación atmosférica; Biodiversidad;
      Productos químicos; Cambio climatic; Costas y mares; Energía;
      Medio ambiente y salud; Tecnología medioambiental; Pesca;
      Economía verde; Consumo de los hogares; Industria; Usos del
      suelo; Recursos naturales; Ruido; Policy instruments; Suelo;
      Regiones específicas; Turismo; Transporte; Medio ambiente
      urbano; Residuos y recursos materiales; Agua
    </dc:subject>
    <dc:date>2012-11-19</dc:date>
    <dc:type>Text</dc:type>
    <dc:format>text/html;charset=utf-8</dc:format>
    <dc:format>100765 bytes</dc:format>
  </rdf:Description>
</rdf:RDF>
```

Figura 24. Exemple de descripció Pàgina web de la EEA

Com s'observa en l'exemple 1 (figura 20), ja en la primer línia s'especifica l'ús d'xml (<?xml version="1.0"?>). RDF es representat per un element xml amb l'etiqueta *rdf:RDF*, i, el contingut d'aquest element és un nombre de descripcions que utilitzen les etiquetes *rdf:Description*. Cada descripció ofereix una sentència sobre un recurs que s'identifica amb una de les tres possibilitats següents:

- Atribut *about*, referit a un recurs ja existent (vegeu figura 24),
- Atribut *ID*, que permet crear un nou recurs, i

- Sense nom, per crear un recurs anònim.

Les descripcions s'han de donar en un ordre determinat, és a dir, XML utilitza una serialització de la informació. L'ordre de les descripcions o recursos, no és significativa si es segueix el model abstracte d'RDF. És a dir, el que determina el model real de dades és el model de grafs representat per RDF i XML és només una possible representació serialitzada del graf.

Així doncs, RDF va molt més enllà del model d'estructura jeràrquica d'XML, ja que es basa en un model més complex i reutilitzable. Els grafs permeten definir la informació aportant semàntica al contingut, és a dir, amb RDF, les descripcions dels recursos es basen en tripletes com "subjecte-predicat-objecte", on el subjecte és el recurs a descriure, el predicat és una propietat o relació del recurs i l'objecte és el valor assignat a aquesta propietat o recurs amb el que s'estableix una relació.

Un recurs (Subjecte) és tot allò que pot ser descrit per una expressió RDF, de manera que pot ser una pàgina web, un lloc complet, una part d'una plana web com una entitat abstracta, una persona o un objecte que no té perquè estar accessible via web (com per exemple, un llibre, una persona, un lloc o un esdeveniment). Tots els recursos són identificats mitjançant una URI (*Uniform Resource Identifier*). En RDF, els recursos s'identifiquen mitjançant URI, mentre que la seva localització (i accés) a la xarxa es realitzen a través de la URL (Uniform Resource Locator). És important recalcar que qualsevol cosa que es vulgui definir pot tenir associada una URI i la seva extensió permet la definició d'identificadors per a qualsevol objecte imaginable i no ha d'estar necessàriament dirigit a un recurs web.

Les propietats (Predicat) dels recursos són les característiques, atributs o relacions utilitzades per descriure al recurs, i cada propietat té un significat específic, ja que defineix els seus valors possibles, els tipus de recursos sobre els que es pot aplicar per a la seva descripció, així com la seva relació amb altres propietats.

Les sentències (Objecte) és un recurs o objecte al qual s'associa una propietat que a la vegada s'assigna a un determinat valor que descriu un aspecte del recurs o sentència. En el cas que el valor de la propietat es refereixi a un altre recurs s'ha d'identificar mitjançant una URI associada. Una sentència és per exemple l'autor de la web "<http://www.eea.europa.eu/>" que és la Unió Europea. El valor pot ser una cadena de caràcters literals com "Unió Europea" o bé, un recurs com pot ser la pàgina principal o "home page" de la web http://www.eea.europa.eu/publications#%C3%A7=en&%C9=all&%C11=5&%B_start=0 (pàgina principal: <http://www.eea.europa.eu/>).

Al mateix temps, RDF permet escriure sentències sobre altres sentències per tal d'aportar coneixement sobre altres declaracions, i per a que això sigui

possible cada sentència ha de tenir un identificador. És a dir, RDF permet convertir una sentència en un recurs (mecanisme de cosificació), en aquest procés s'introdueix un objecte auxiliar que ha d'estar relacionat amb la sentència original a través de les propietats de subjecte, predicat o objecte. Aquest objecte és el contenidor de la sentència original (cosificada), i aquest es pot utilitzar per formar noves sentències relacionades amb la sentència original.

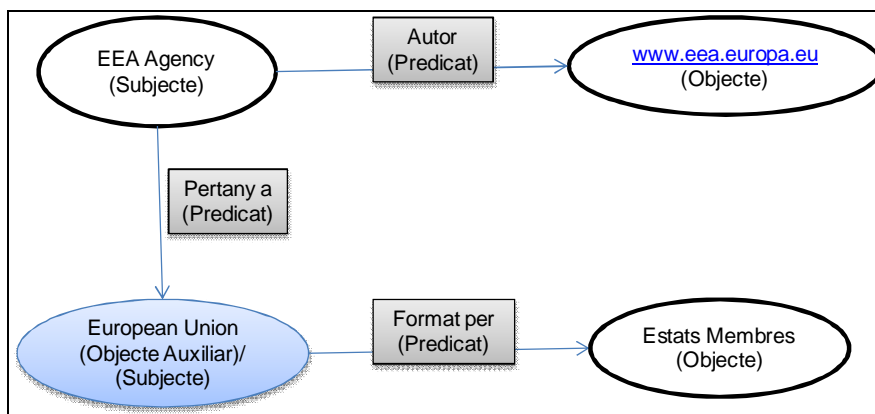


Figura 25. Grafo simple de sentència sobre sentència (reidificació)

RDF també permet relacionar els recursos amb una col·lecció (recursos) determinada. Això és possible perquè hi ha una la definició de tipus de contenidors que permeten gestionar llistes de recursos o literals. Els tipus de contenidors són: *Bag* (una llista de recursos o literals sense ordre que permet duplicats), *Sequence* (una seqüència o llista ordenada de recursos o literals, la seqüència permet declarar un conjunt en el qual l'ordre d'aparició dels membres és important i també es permeten valors duplicats), *Alternative* (és una seqüència de recursos o literals per un únic valor o propietat, és a dir, d'aquesta llista de recursos o literals només es pot seleccionar un), *Collections* (conjunt de recursos o literals agrupats amb una llista que no pot tenir duplicats).

Es conclou doncs que RDF és un model de dades abstracte que proporciona declaracions sobre els recursos, descrivint i representant de forma explícita les característiques de tot tipus de recursos. RDF utilitza l'XML (RDF/XML) per serialitzar aquestes declaracions en una sintaxi concreta. RDF facilita la comprensió de les màquines d'un conjunt de dades que són definides amb RDF i que posteriorment poden ser utilitzades per extreure informació o executar diferents processos (reutilització de la informació). No obstant això, RDF és un model d'ontologia bàsic (identificació d'objectes, relacions, reidificació, contenidors i col·leccions), per aquest motiu, es va desenvolupar una extensió del model amb funcionalitats addicionals, l'RDFS (RDF Schema). L'RDFS afegeix característiques als recursos i proporciona un vocabulari estandarditzat per als conceptes, és a dir, es descriuen meta-conceptes, l'objectiu és proporcionar l'enriquiment dels recursos amb semàntica

(ontologies), cal recordar que el web fins ara ha estat concebut com un web sintàctic basat amb HTML.

Des de que el Word Wide Web va proposar RDF com a estàndard, aquest ha estat considerat un dels majors avenços en aplicacions web basades en tesaurs (Ali Asghar Shiri i Crawford Revie, 2000) i en la millora de la interoperabilitat entre sistemes. RDF, descrit amb una sintaxis XML (Koch, T., 1999) fa possible l'intercanvi de dades amb vocabularis controlats de diferents aplicacions i serveis.

RDFS pot definir propietats a través de la descripció d'esquemes senzills de classes i subclasses, el domini d'aplicació i el rang de valors possibles. Però com es tracta d'un llenguatge de descripció molt bàsic és necessari que es complementi amb altres llenguatges de modelatge com OWL o SKOS.

Els principals elements del modelatge per a la descripció de recursos són:

- **Classes:** permeten definir una jerarquia explícita de classes. Una classe es definida com un recurs i té un identificador (ID) únic. I existeixen tres tipus de classe que permeten definir noves classes.
 - *Rdfs:Resource*: qualsevol objecte que pugui ser objecte d'una descripció a través d'una sentència RDF.
 - *Rdfs:Property*: és la classe que permet definir noves propietats per a la descripció dels recursos.
 - *Rdfs:Class*: s'utilitza per definir altres classes o instàncies.
- **Propietats:** descriuen les característiques o relacions entre els recursos RDFS. Les propietats bàsiques són:
 - *Rdfs:type*: descriu modela interrelacions del tipus instància entre recursos i classes. Un recurs pot ser una instància de més d'una classe.
 - *Rdfs:subClassOf*: defineix la jerarquia de classes. Una classe pot ser subclasse d'altres subclasses. Per exemple, si una classe A2 és subclasse d'A1 i un recurs R pertany a la classe A2, llavors també pertany a la Classe A1.
 - *Rdfs:PropertyOf*: permet crear jerarquies de propietats. Cal tenir en compte que si una propietat P2 és subpropietat de la propietat P1, i un recurs R té una propietat P2 amb valor V, això implica que el recurs R té la propietat P1 amb valor V.

- Restriccions de propietats: són elements necessaris per indicar sobre quines classes de recursos es poden aplicar determinades propietats, així com possibles valors que aquestes puguin agafar. Restriccions:
 - *Rdfs:domain*: permet indicar les classes sobre les que aplicar una propietat
 - *Rdfs:range*: indica els possibles valors que pot agafar una propietat

Aquestes restriccions són instàncies de la classe *rdfs:ConstraintProperty*, per tant, és a la vegada una subclasse de *rdf:Property* com de *rdfs:ConstraintResource* i aquesta a la vegada ho és de *Rdfs:Resource*. A continuació es representa un graf de RDFS de la sentència “Charles Darwin és autor del llibre “The Origin of species””:

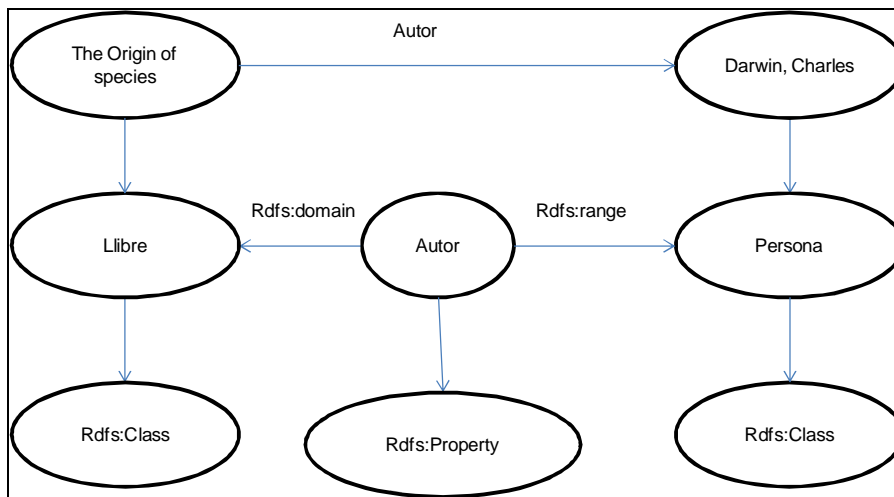


Figura 26. Representació de graf RDFS

El principal objectiu d’RDF és proporcionar informació descriptiva simple sobre els recursos que es troben a la web. Dos dels exemples més coneguts d’aplicació d’RDF són RSS i FOAF.

RDF Site Summary (RSS) 1.0 (<http://web.resource.org/rss/1.0/>) és un vocabulari RDF que permet la catalogació d’informació com notícies i esdeveniments de manera que sigui possible localitzar aquesta informació adaptada a les preferències dels usuaris. Els arxius RSS contenen metadades sobre fonts d’informació especificades pels usuaris la funció principal és avisar els usuaris que els recursos que ells han seleccionat han estat actualitzats sense necessitat de comprovar directament la pàgina.

FOAF (<http://www.foaf-project.org/>) és un projecte que permet crear descripcions sobre persones, els vincles que existeixen entre ells i les coses que fan i/o creen. Es tracta novament d’un vocabulari RDF que permet tenir

disponible informació personal de manera senzilla perquè pugui ser processada, compartida i reutilitzada.

4.2.2.2.2. OWL

El Llenguatge d'Ontologies web (OWL), segons W3C, està dissenyat per a ser usat en aplicacions que necessiten processar el contingut de la informació en lloc d'únicament representar la informació per als humans. OWL facilita un millor mecanisme d'interpretabilitat de contingut web que la resta de formats presentats com XML, RDF, i RDFS (RDF-Schema). OWL és un mecanisme per desenvolupar sistemes d'organització del coneixement o ontologies (vocabularis controlats) específics amb els quals s'associen recursos i poden ser utilitzades a través de diferents sistemes d'informació. RDF es limita a establir predicats bàsics i RDF Schema es limita a representar jerarquies (subclasses i propietats) i oferir definicions d'un domini i el rang d'aquestes propietats. En canvi, amb OWL s'aconsegueix una major expressivitat i representació del coneixement.

Així doncs, OWL és una recomanació del W3C d'un llenguatge específic per crear ontologies. L'objectiu dels llenguatges d'ontologies és permetre escriure la conceptualització explícita i formal dels models del domini que es vulguin representar, és a dir, descriu conceptes i relacions entre els conceptes d'un camp del coneixement o comunitat concret. Els components característics de les ontologies són (Gruber, 1993):

- **Conceptes:** són idees bàsiques que s'intenten formalitzar. Aquests conceptes poden ser objectes, mètodes, plans, estratègies, processos de raonament, etc.
- **Relacions:** representen la interacció i enllaç entre els conceptes del domini. Les relacions formen part de la taxonomia del domini. Per exemple: classe, subclasse-de, part-de, part-exhaustiva-de, connectat-a, etc.
- **Funcions:** són un tipus concret de relació on s'identifica un element mitjançant el càlcul d'una funció que considera diferents elements de l'ontologia. Per exemple, poden aparèixer funcions com: categoritzar-classe, assignar-data, etc.
- **Instàncies:** s'utilitzen per representar objectes determinats d'un concepte
- **Axiomes:** són teoremes que es declaren sobre les relacions que han de complir els elements de l'ontologia. És a dir, requisits com "si A i B són de la classe C, llavors A no és subclasse de B", etc.

O sigui, les ontologies inclouen definicions de conceptes bàsics sobre una disciplina concreta i la relació entre aquests conceptes. A nivell d'estructura

organitzativa podem imaginar un model d'Entitat-Relació (model conceptual d'una base de dades relacionals) on cada recurs associat ha de ser descrit per les seves propietats i classes (com passava amb RDF), i, ahora es permet establir jerarquies de generalització entre les propietats i les classes, a més a més, de poder descriure el tipus de relacions entre les classes com per exemple, relacions de disjunció, cardinalitat (si és repetible o no), igualtat, tipologies de propietats, caracterització de les propietats o classes enumerades.

A tot això, per tal d'aconseguir una ontologia consistent cal considerar els següents requisits principals (Antoniou, G, i VanHarmelen, F., 2010):

- Una sintaxis ben definida (es refereix a un llenguatge de programació que permet que la màquina processï la informació) i una semàntica formal: es refereix a la lògica matemàtica i és un requisit previ al raonament. La semàntica formal descriu de forma precisa el significat del coneixement i no està subjecta a interpretacions. Per exemple, es poden definir els següents raonaments:
 - Establir l'equivalència entre classes: $A=B$ i $B=C \rightarrow A=C$
 - Pertinença de classes: Si X és una instància de C i C és una subclasse de D , llavors X és una instància de D .
 - Consistència: si es defineix que X és una instància de la classe A , i A és una subclasse de $B \cap C$, on A és una subclasse de D , i B i D són disjunctes, es produeix una situació d'inconsistència, perquè A hauria d'estar buida però conté la instància X , això indica un error en l'ontologia.
 - Classificació: quan es declara que certs parells de propietat-valor són suficients per pertànyer a una classe. Per exemple, agafant una classe A , si un individu X compleix les condicions de la classe A , es pot concloure que X hauria de ser una instància de A .
- Oferir suport al raonament, és a dir, es permeten establir relacions de semàntica formal de forma automàtica. A més, permet comprovar la consistència de l'ontologia, si existeixen relacions que no estan previstes entre classes i si es realitza la classificació automàtica d'instàncies en les classes.
- Tenir suficient poder expressiu i conveniència d'expressió: no és només suficient amb l'organització del coneixement en jerarquies de tripletes (RDF/RDFS) amb les seves restriccions, relacions, etc. Sinó que també és necessari definir el rang d'una propietat, indicar diferents tipus de classes no jeràrquiques com classes disjunctes, combinacions booleanes (establir noves classes a partir de la combinació d'altres classes),

restriccions de cardinalitat o característiques especials de les propietats (transitives, inverses, etc.).

OWL té tres tipus de llenguatges, OWL Lite, OWL DL i OWL Full. Totes les varietats d'OWL utilitzen RDF per a la seva sintaxis, les instàncies es declaren igual que RDF, utilitzant descripcions RDF i informació en triples (Subjecte-Predicat-Objecte). També hi ha construccions d'OWL que són especialitzacions dels seus homòlegs en RDF, com per exemple `owl:Class` i `rdf:Class`, o bé, `owl:DatatypeProperty` o `owl:ObjectProperty` i `rdf:Property`.

OWL Full és un llenguatge complex que utilitza totes les primitives del llenguatge OWL, però també permet la combinació d'aquestes de forma arbitrària amb RDF i RDFs. Aquesta característica fa que sigui possible que canviï el significat de les primitives predefinides de RDF o OWL, aplicant les primitives d'un o altre llenguatge. Per exemple, es pot posar una restricció de cardinalitat en la classe de totes les classes, limitant el nombre de classes que es poden descriure en una ontologia. Tot i que aquesta flexibilitat del llenguatge és un avantatge, és també un inconvenient, ja que aquesta complexitat acaba dificultant el raonament i fent que sigui un llenguatge indecidible.

OWL DL (*OWL Description Logic*) és un subllenguatge de OWL Full que es crea amb l'objectiu d'obtenir una eficiència computacional. OWL DL restringeix la forma amb la que es poden utilitzar els constructors de OWL i RDF assegurant que el llenguatge tingui una lògica descriptiva, és a dir, no permet que s'apliquin a la vegada els constructors d'OWL i RDF.

OWL Lite és una restricció addicional que limita a OWL DL a un subconjunt dels constructors del llenguatge. Per exemple, exclou la possibilitat de declarar classes enumerades, enunciats disjunts i una cardinalitat arbitrària. L'avantatge del llenguatge és que facilita la seva comprensió per als usuaris i és fàcil d'implementar per als desenvolupadors. I l'inconvenient principal és la pròpia restricció del llenguatge que no té tanta expressivitat com la resta.

La principal característica diferenciadora entre OWL Full i l'altre parell de llenguatges d'OWL és que OWL Full té una major expressivitat que la resta de llenguatges, essent capaç de representar construccions complexes del món real, i també ofereix una major flexibilitat (combinació de constructors OWL i RDF). Però al mateix temps, la complexitat del propi llenguatge dificulta el raonament de les ontologies que es desenvolupin. Per tant, l'ús d'OWL requereix una major capacitat i recursos de computació per poder validar les ontologies i inferències creades.

L'ús d'OWL DL per crear una ontologia sembla ser molt més avantatjós (López 2012, Antoniou 2010, Lacy 2005). D'una banda hi ha un ampli mercat d'aplicacions que l'utilitzen, i d'altra banda, perquè OWL DL es va crear per donar suport i proporcionar propietats de càlcul desitjables per als sistemes de

raonament. És a dir, la interrogació a un sistema d'informació que utilitzi OWL DL, en la majoria dels casos, pot donar una resposta lògica en un temps determinat ja que té unes bones propietats computacionals. No obstant això, es perd la compatibilitat amb RDF que existeix amb OWL Full.

Un document OWL és conegut com ontologies OWL i són documents amb RDF. L'estructura de la qual és la següent: capçalera (header), cos (body) i el peu de pàgina (footer).

Document OWL (estructura, definició i abast)	Exemples de Representació OWL (representació tipus document XML) ³⁴
Capçalera	Capçalera
Declaració XML i RDF Start Tag <ul style="list-style-type: none"> - Declaració espais de noms 	<pre><rdf:RDF xmlns:owl ="http://www.w3.org/2002/07/owl#" xmlns:rdf ="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs ="http://www.w3.org/2000/01/rdf-schema#" xmlns:xsd ="http://www.w3.org/2001/XMLSchema#"</pre>
Element d'ontologia <ul style="list-style-type: none"> - Informació de la versió - Elements importats 	<pre><owl:Ontology rdf:about=""> <rdfs:comment>Exemple d'ontologia OWL</rdfs:comment> <owl:priorVersion rdf:resource="www.exempleOwl.cat"/> <owl:imports rdf:resource="www.exempleOwl.cat"/place "> <rdfs:label>Universitat de Barcelona </rdfs:label> </owl:Ontology></pre>
Cos	Cos
<p>Classes: on es defineixen els diferents tipus de classes i individus, així com les seves jerarquies i enumeracions. Hi ha tres tipus de classes: <i>owl:Class</i>, <i>owl:Thing</i>, <i>owl:Nothing</i>. <i>owl:Class</i> defineix l'element a descriure, la classe <i>owl:Thing</i>, és la classe més general i la que ho conté tot (<i>totes les classes són subclasses de owl: Thing</i>), i <i>owl:Nothing</i> és una classe buida (<i>totes les subclasses tenen una superclasse owl:Nothing</i>). I les subclasses i instàncies es defineixen per <i>rdf:subClassOf</i> i <i>rdfs:type</i></p>	<pre><owl:Class rdf:ID="ésser viu"> <rdfs:subClassOf rdf:source="#animal"/> rdf:type="#peix"/> </owl:Class></pre>
<p>Propietats: Propietats de tipus objecte que relacionen els objectes amb altres</p>	<pre><owl : ObjectProperty rdf: ID="esImpartidaPer"> <rdfs: domain rdf:resource="#conferencia"/></pre>

³⁴Nota: La representació dels formats ontològics pot ser representada com en l'exemple, en forma de document XML, o bé en format de graf (presentat en apartats anteriors) o en format turtle, un altre tipus de sintaxis RDF consultable a: <http://www.w3.org/TeamSubmission/turtle/>.

<p>objectes (entre instàncies de classes): <i>owl:ObjectProperty</i> Propietats de tipus de dades, que relacionen els objectes amb valors de dades (relació entre una classe i un literal, i permet ús de dades XML Schema): <i>owl:DatatypeProperty</i>. Per derivar unes propietats a partir d'unes altres es pot utilitzar "rdfs:subPropertyOf". També es poden utilitzar propietats RDFS per definir el domini (<i>rdfs:domain</i>) o el rang d'una propietat (<i>rdfs:range</i>)</p>	<pre><rdfs:range rdf:resource="#PersonalInvestigador"/> <rdfs:subPropertyOf rdf:resource="#implica"/> <owl : ObjectProperty> <owl : DatatypeProperty rdf:ID="edad"> <rdfs:range rdf=resource=http://www.w3.org/2001/.../> <owl : DatatypeProperty></pre>
<p>Característiques especials de les propietats (individus o instàncies de classes):</p> <p>1. Propietats de les relacions:</p> <ul style="list-style-type: none"> - Transitiva: defineix relacions del tipus "és més alta que" o "té més nota que", o "pertany a" o "és del tipus". Per exemple, sabem que peix (A) i animal (C) són éssers vius (B), però volem definir que 'peix' és un tipus d'animal: R(A, B) y R(B, C) \Rightarrow R(A, C) - Simètrica³⁵: s'utilitza per expressar les instàncies com "té la mateixa nota que" o "viuen a la mateixa regió" o "és germà de". O bé es representa per: R(A, B) \Rightarrow R(B, A) - Funcional: defineix una propietat que agafa com a màxim un valor per a cada objecte, com per exemple, la edat, l'altura, el sexe, estat civil, etc. És a dir, R(A, B) y R(A, C) \Rightarrow B=C; - Inversa: és el tipus de relació habitual, per exemple, "si un investigador imparteix una conferència, la conferència és impartida per l'investigador". És a dir, R(A, B) \Rightarrow S(B, A) - Inversa funcional³⁶: defineix una propietat per a la qual dos objectes diferents no poden agafar el mateix valor, per exemple "Número de la 	<pre>[owl:TransitiveProperty]: <owl:ObjectProperty rdf:ID="Animaltipus"> <rdf:type rdf:resource="&owl;TransitiveProperty"/> <rdfs:domain rdf:resource="#ÉsserViu "/> <rdfs:range rdf:resource="#ÉsserViu "/> </owl:ObjectProperty> [owl:SymetryProperty]: <owl:ObjectProperty rdf:ID="viuMateixaRegió"> Rdf:type rdf:resource="&owl ; SymetricProperty" /> <rdfs:domain rdf:domain rdf:resource="#Persona" /> <rdfs:range rdf:resource="#Persona" /> </owl:ObjectProperty> [owl:FunctionalProperty]: <owl:ObjectProperty rdf:ID="EstatCivil"> <rdf:type rdf:resource="&owl;FunctionalProperty" /> <rdfs:domain rdf:resource="#Casat" /> <rdfs:range rdf:resource="#Solter" /> </owl:ObjectProperty> [owl:InverseOf]: <owl : ObjectProperty rdf: ID="imparteix"> <rdfs:range rdf:resource="#conferencia"/> <rdfs: domain rdf:resource="#PersonalInvestigador"/> <owl:inverseOf rdf=resource="#esImpartidaPer"/> <owl : ObjectProperty> [owl:InverseFunctionalProperty] <owl:InverseFunctionalProperty rdf:ID="biologicalMotherOf"> <rdfs:domain rdf:resource="#Woman"/> <rdfs:range rdf:resource="#Human"/></pre>

³⁵ Nota: Sintàcticament, una propietat es defineix com simètrica pel fet de que és una instància incorporada a la Classe OWL.

³⁶ Nota2: OWL DL i les propietats de "tipus de dades" són disjunctes, de manera que una propietat inversa-funcional no pot ser definida per les propietats de tipus de dades.

<p>Seguretat Social” o el “DNI”: R(B, A) y R(C, A) \square B=C</p>	<pre></owl:InverseFunctionalProperty></pre>
<p>2. Restriccions de les propietats³⁷: es defineixen com un recurs declarat com una instància de la classe <i>owl:Restriction</i>; hi ha variacions segons el tipus d'OWL utilitzat, per ampliar informació cal visitar la pàgina de cadascun dels estàndards³⁸ del web semàntic del W3C. En aquesta secció es presenten les restriccions de propietats generals³⁹:</p> <ul style="list-style-type: none"> - <i>Rdfs:subClassOf</i>, permet especificar que una classe C1 és una subclasse de la classe C, per tant, totes les instàncies de C1 seran també de C. - <i>owl:allValuesFrom</i>, utilitzada per delimitar la classe dels possibles valors que la propietat que s'especifica amb <i>owl:onProperty</i> pot prendre (tots els valors de la propietat han de sorgir d'aquesta classe). En l'exemple, estem dient, que només els Investigadors Principals (IP's) són permesos com a valors de la propietat "ImpartidaPer". - <i>owl:allhasValue</i>, assenjala un valor concret que ha de tenir la propietat especificada per l'element <i>owl:onProperty</i>; al mateix temps fa referència a una "quantificació universal", el valor indica un "tot". Seguint, l'exemple, aquest tot indicaria, que "totes" les conferències són impartides per IP's. - <i>owl:somehasValue</i>, fa referència a una restricció anomenada de "quantificació existencial", el valor de la qual indica "alguns", en l'exemple, es fa referència a que algunes conferències són impartides per IP's. 	<pre> Example: rdfs:subClassOf / owl:allValuesFrom <owl:Class rdf:about="#ConferenciesEU"> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty rdf:resource="#esImpartidaPer" /> <owl: rdf:resource="#InvestigadorPrincipal"/ > <owl:Restriction> </rdfs:subClassOf> </owl:Class> Example: owl:hasValue <owl:Class rdf:about="ConferenciesEU" /> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty rdf:resource="#esImpartidaPer"/> <owl:hasValue rdf:resource=#123455"/> <owl:Restriction> <rdfs:subClassOf> </owl:Class> Exemple: owl:hasValue <owl:Class rdf:about="ConferenciesEU"A <rdfs:subClassOf> <owl:Restriction> <owl: onProperty rdf:resource="#esImpartidaPer"/> <owl:hasValue rdf:resource=#123455"/> <owl:Restriction> <rdfs:subClassOf> </owl:Class> Exemple de: owl:somehasValue <owl:Class rdf:about="PersonalIP"/> <rdfs:subClassOf> <owl:Restriction> <owl: onProperty rdf:resource="#Imparteix"/> <owl:someValuesFrom rdf:resource=#ConferenciesEU"/> <owl:Restriction> <rdfs:subClassOf> </owl:Class> </pre>
<p>Peu de Pàgina</p>	<p>Peu de Pàgina</p>

³⁷ W3C està treballant per desenvolupar noves propietats i restriccions en la versió d'OWL2, consulta a: <http://www.w3c.org/TR/owl2-overview/>

³⁸ Semantic web Standards. W3c Wiki. Url: http://semanticweb.org/wiki/Semantic_web_standards [darrera consulta: 10 desembre, 2012]

³⁹ OWL. Capítol IV. Manual del web semàntic. Grigoris Antoniou; Frank Van Harmelen (autors). Granada: Comares, 2010. Pàgina. 138-141. ISBN:978-84-9836-780-5

- RDF End Tag	</rdf:RDF>
---------------	------------

Taula 18. Exemple de tipus de document OWL

Una altra característica de OWL és que es poden fer operacions sobre les classes, es poden definir classes com a resultat de la intersecció (AND), unió (OR) o complement (NOT) entre classes i/o propietats (combinacions booleanes). Aquestes combinacions booleanes es poden definir amb *owl:Class*, o mitjançant expressions de les classes. Un altre tipus de combinació de classes és la relació disjunta (*owl:disjointWith*), és a dir, dues classes que no poden tenir relació, per exemple: si un individu és declara com a una classe de tipus 'revista' no pot ser de tipus 'llibre' i viceversa.

```
<owl:Class rdf:ID="personesUniversitat">
  <owl:unionOf rdf:parseType="Collection">
    <owl:Class rdf:about="#membrePersonal" />
    <owl:Class rdf:about="#estudiant" />
  </owl:unionOf>
</owl:Class>
```

Figura 27. Exemple de propietat d'unió de classes (owl:unionOf)

```
<owl:Class rdf:ID="DocentenBiologia">
  <owl:intersectionOf rdf:parseType="Collection">
    <owl:Class rdf:about="#cosdocent" />
    <owl:Restriction>
      <owl:onProperty rdf:resource="#perteneixA" />
      <owl:hasValue rdf:resource="#DepartamentBiologia"/>
    </owl:Restriction>
  </owl:intersectionOf>
</owl:Class>
```

Figura 28. Exemple de Propietat d'intersecció de classes (owl:intersectionOf)

```
<owl:Class rdf:ID="assignatura">
  <rdfs:subClassOf>
    <owl:Class>
      <owl:complementOf rdf:resource="membrePersonal">
    </owl:Class>
  </rdfs:subClassOf>
</owl:Class>
```

Figura 29. Exemple de propietat de complement (owl:complementOf)

La propietat *owl:complementOf* es pot substituir per *owl:disjointWith*

Una altra característica interessant d'OWL és que permet elaborar sentències d'equivalència (o mapeig) entre dues o més classes (*owl:equivalentClass*), entre propietats (*owl:equivalentProperty*), o bé entre individus (*owl:sameAs*).

```
<owl:ObjectProperty rdf:ID="DonaConferenciesDe">
  <owl:equivalentProperty rdf:resource="#imparteix" />
</owl:ObjectProperty>
```

Figura 30. Exemple d'equivalència de propietats (owl:equivalentProperty)

```
<owl:ObjectProperty rdf:ID="Revista">
```

```
<owl:equivalentClass rdf:resource="#Journal"/>
</owl:ObjectProperty>
```

Figura 31. Exemple d'equivalència entre classes (owl:equivalentClass)

```
<owl:ObjectProperty rdf:ID="Journal of Applied Phycology">
  <owl:sameAs rdf:resource="#J.Appl. Phycol"/>
</owl:ObjectProperty>
```

Figura 32. Exemple d'equivalència entre individus (owl:sameAs)

La utilització de les sentències d'equivalència implica lògiques molt complexes i es recomanen l'ús d'altres formats com SKOS, per tal de mapejar vocabularis controlats.

La idea central de OWL, en referència a la resta de llenguatges d'ontologies simples (SKOS o RDF), és que permet formular restriccions amb l'objectiu de descriure el coneixement amb una major precisió, alhora permet validar les descripcions dels conceptes i inferir coneixement (establir deduccions o axiomes).

En resum, el disseny d'una ontologia requereix una presa de decisions important a l'hora d'utilitzar uns o altres formats. Així doncs, el model ontològic a utilitzar només dependrà de la complexitat del model ontològic que es vulgui representar.

4.2.2.2.3. SKOS

SKOS és un format o llenguatge d'ontologia estàndard desenvolupat específicament per donar suport i ser utilitzat en els sistemes d'organització del coneixement (KOS), com per exemple, tesaurus, sistemes de classificació, encapçalaments de matèries i taxonomies dintre el marc del web semàntic. Així doncs, podem incloure SKOS dintre dels estàndards d'enriquiment o valor de dades on estan agrupats els esquemes de codificació i els vocabularis controlats (LCSH, etc.). SKOS és un vocabulari de RDF (*Resource Description Framework*)_especificació del W3C definida com a model de dades de metadades per representar sistemes d'organització del coneixement de manera semi-formal (KOS). Com SKOS està basat en RDF la seva representació és llegible per màquina i pot ser intercanvable entre aplicacions de software i publicat al World Wide Web, de manera que també es podria agrupar dintre dels estàndards per a l'intercanvi de dades.

A diferència d'OWL (usat amb un altre tipus de context), SKOS ha estat dissenyat per proveir un camí de canvi dels sistemes d'organització existents al web semàntic. A més de ser un llenguatge de modelat que facilita el desenvolupament i intercanvi de nous sistemes d'organització del coneixement (KOS), de fet en la seva creació s'han tingut en compte les guies estàndards de construcció de tesaurus monolingües i multilingües (ISO 2788 i BS8723-2), les quals s'aniran analitzant a través de l'aplicació actual que ens ofereixen els estàndards del web semàntic (SKOS, RDF, OWL, etc.). Per aquest notiu, en

aquest estudi és descarta l'ús d'OWL estrictament per al desenvolupament de sistemes d'organització del coneixement en entorns digitals, tot i que en aquest capítol veurem que algunes de les seves funcionalitats permetran enriquir i millorar el llenguatge SKOS. A mesura que s'avança en aquest capítol es va observant l'aplicació pràctica que pot tenir aquest estàndard en qualsevol sistema d'informació que al mateix temps requereix d'un sistema d'organització del coneixement per a la seva indexació i recuperació de la informació.

En un nivell bàsic, podem dir que el format SKOS està format per recursos conceptuals que s'identifiquen amb una URI (*Uniform Resource Identifier*) que s'etiqueten amb nexes lèxics amb un o més llenguatges naturals. També són documentats amb diferents tipus de notes que estan semànticament relacionades entre elles a través d'una jerarquia informal i xarxes associades i agregades a aquests esquemes de conceptes. És a dir, representen perfectament les estructures jeràrquiques i associatives de qualsevol sistema d'organització del coneixement.

SKOS es representat doncs amb 5 elements bàsics: conceptes, etiquetes, relacions, notes i esquemes.

1. Els **conceptes** són representats per una classe anomenada [skos:Concept](#) que permet als implementadors afirmar que un determinat recurs és un concepte. Això es fa amb dos processos:
 - Creant o reutilitzant una URI que identifica de forma unívoca el concepte (vegeu Annex 4.1), i
 - Afirmant en RDF, usant la propietat [rdf:type](#), que el recurs és identificat per una determinada URI del tipus [skos:Concept](#).

Exemple 1. Representació de conceptes en rdf/skos:

```
<http://www.example.com/animals> rdf:type skos:Concept
```

Exemple 2. Representació de conceptes en format Turtle, usant el prefix *ex* definit a sota:

```
ex:animals rdf:type skos:Concept
```

Figura 33. Exemples de representació de conceptes (SKOS)

2. Les **etiquetes** caracteritzen els conceptes i són les expressions utilitzades per referir-se als conceptes en llenguatge natural. SKOS ofereix té 3 propietats bàsiques per fixar les etiquetes dels recursos conceptuals: [skos:prefLabel](#), [skos:altLabel](#) and [skos:hiddenLabel](#).

Aquestes etiquetes són subpropietats de [rdfs:label](#) i són utilitzades per enllaçar a la etiqueta [skos:Concept](#) d'un [RDF plain literal](#), és a dir, són etiquetes associades al concepte i que representen una cadena de

caràcters que van acompanyades d'una etiqueta opcional de llengua i que poden ser usades amb text pla i amb un llenguatge natural. Per exemple: "animals" i "en.US". [Veure: [RDF-CONCEPTS](#)].

Cadascuna d'aquestes tres propietats, implica un estatus específic per a l'etiqueta que introdueix, que pot anar des d'una forta relació, denotació unívoca, a una cadena per ajudar en la cerca. Aquestes propietats són definides formalment com disjunctes per pars. Això significa, per exemple, que és un error que un concepte tingui la mateixa etiqueta literal tant si fa referència a "preferred label" com a la etiqueta "alternative label".

- **Etiquetes "Preferred Lexical Label":** la propietat [skos:prefLabel](#) permet assignar una etiqueta lèxica preferent a un recurs. Els termes que siguin utilitzats com a descriptors en els sistemes d'indexació [[WillpowerGlossary](#)] es representaran mitjançant aquesta propietat.

```
ex:animales rdf:type skos:Concept;
skos:prefLabel "animales".
```

Figura 34. Exemple d'etiqueta "prefLabel" (SKOS)

Aquestes etiquetes preferents poden tenir etiquetes "literals RDF simples", és a dir, són etiquetes opcionals per definir un idioma. SKOS permet d'una forma senzilla l'etiqueta't multilingüe.

```
ex:animales rdf:type skos:Concept;
skos:prefLabel "animales"@es;
skos:prefLabel "animals"@en.
```

Figura 35. Exemple d'etiquetat RDF literal d'etiqueta multilingüe (SKOS)

Cal tenir en compte que en una etiqueta preferent (*prefLabel*) implica que només hi hagi un recurs amb una etiqueta assignada com a preferent, és a dir, només pot haver un idioma preferent. Així doncs, en un sistema d'organització del coneixement (KOS) multilingüe només podem assignar a un idioma una etiqueta preferent, la resta d'idiomes s'hauran d'etiquetar com a "alternative labels". D'altra banda, l'etiqueta preferent d'un concepte també es pot utilitzar per representar de forma inequívoca al mateix concepte dintre del propi sistema i les seves aplicacions. Així doncs, tot i que el model de dades SKOS no obliga formalment a que es faci així, es recomana que no existeixin dos conceptes en el mateix sistema d'organització que tinguin la mateixa etiqueta lèxica preferent per a un mateix idioma. Exemple t'etiquetat SKOS inconsistent:

<Fauna> skos:prefLabel "animals"@en ; skos:prefLabel "fauna"@en .	No poden haver 2 "preferred labels"
--	--

<Fauna> skos:prefLabel "animals"@en ; skos:altLabel "animals"@en .	Una <i>prefered label</i> no pot ser al mateix temps <i>alternative label</i>
<Love> skos:altLabel "animals"@en ; skos:hiddenLabel "animals"@en .	Ni el mateix concepte etiquetat com <i>alternative label</i> potser una <i>hiddenLabel</i>
<Love> skos:prefLabel "animals"@en ; skos:hiddenLabel "animals"@en .	Ni una mateixa etiqueta <i>prefered</i> potser una <i>hiddenLabel</i>

Figura 36. Exemple t'etiqueta't SKOS inconsistent

Exemple t'etiqueta't SKOS consistent:

```
MyResource>
  skos:prefLabel "animals"@en ;
  skos:altLabel "fauna"@en ;
  skos:hiddenLabel "aminals"@en ;
  skos:prefLabel "animaux"@fr ;
  skos:altLabel "faune"@fr .
```

Figura 37. Exemple d'etiquetat SKOS consistent

Les propietats *skos:prefLabel*, *skos:altLabel* i *skos:hiddenLabel* no tenen un domini establert i el més efectiu segons la guia de referència d'SKOS del W3C és establir la classe de tots els recursos: *rdfs:Resource*. En l'exemple següent s'ha etiquetat un recurs de tipus *owl:Class*.

```
<MyClass>    rdf:type    owl:Class    ;
  skos:prefLabel    "animals"@en    ;
  skos:altLabel     "fauna"@en     ;
  skos:hiddenLabel  "aminals"@en    ;
  skos:prefLabel    "animaux"@fr    ;
  skos:altLabel     "faune"@fr     .
```

Figura 38. Exemple t'etiqueta't consistent sobre el domini de les propietats SKOS

- **Etiquetes “*Alternative Lexical Label*”**: la propietat [skos:altLabel](#) permet assignar una etiqueta lèxica alternativa a un concepte. Aquest és especialment útil quan s'assignen altres etiquetes diferents a la etiqueta preferent del concepte. Un exemple habitual es troba en la representació de sinònims.

```

ex:animales rdf:type skos:Concept;
             skos:prefLabel "animales"@es;
             skos:altLabel "criaturas"@es;
             skos:prefLabel "animals"@en;
             skos:altLabel "creatures"@en.

```

Figura 39. Exemple d'etiquetes alternatives o "prefLabel" (SKOS)

També són útils per presentar quasi sinònims, abreviatures, acrònims, etc. [UNE 50-106-90⁴⁰, secció 5.1-b]. Exemple 6.1. Abreviatures representades amb *preflabel*:

```

ex:onu      rdf:type      skos:Concept;
             skos:prefLabel "Organización de
             las Naciones Unidas"@es;
             skos:altLabel "ONU"@es.

```

Figura 40. Exemple d'abreviatures representades amb "prefLabel" (SKOS)

També es possible utilitzar la etiqueta *skos:altLabel* per representar conceptes que agrupen altres conceptes més específics però que no estan inclosos explícitament com a conceptes d'un sistema d'organització del coneixement concret. Exemple 6.2.

```

ex:rocas    rdf:type      skos:Concept;
             skos:prefLabel "rocas"@es;
             skos:altLabel "basalto"@es;
             skos:altLabel "granito"@es;
             skos:altLabel "pizarra"@es.

```

Figura 41. Exemple d'etiquetes alternatives o "altLabel" (SKOS)

SKOS no pretén substituir les directrius existents per al disseny dels sistemes d'organització del coneixement (ISO 2788 i BS8723-2⁴¹), és per això que la indicació que apareix a les guies sobre l'exemple anterior no és recomanable, tanmateix, s'exposa per veure les possibilitats que SKOS ofereix. És a dir, el més apropiat seria incloure un concepte de tipus "skos:Concept" propi per a cada tipus específic de terme que es vol definir. Per exemple, en l'exemple anterior on el terme preferit és "rocas" es definiria cada tipus de roca (*basalto*, *granito* i *pizarra*) i es relacionarien com a termes específics de roca.

- **Etiquetes Ocultes:** una etiqueta lèxica oculta o *skos:hiddenLabel* s'associa a un recurs quan els responsables d'un sistema d'organització del coneixement precisa que determinades expressions estiguin disponibles per aplicacions que realitzen

⁴⁰ Traducció de la ISO 2788-1966. Veure també la ANSI/NISO Z39.19-2005

⁴¹ BS 8723-2:2005: Structured vocabularies for information retrieval. Guide. Thesauri. Published : November 2005 [Status : Current, Project Underway]

processos d'indexació basats en el text complet de documents i operacions de cerca, però no es vol que aquestes etiquetes quedin visibles d'altra forma. Poden ser utilitzades també per incloure variants d'errors ortogràfics d'altres etiquetes lèxiques, etc.

```
ex:animales rdf:type skos:Concept;  
             skos:prefLabel "animales"@es;  
             skos:altLabel "bestias"@es;  
             skos:hiddenLabel "béstias"@es.
```

Figura 42. Exemple d'etiquetes lèxiques amb "prefLabel" (SKOS)

3. Relacions Semàntiques: en tot sistema d'organització del coneixement, les relacions semàntiques desenvolupen un paper fonamental per a la definició dels conceptes. El significat d'un concepte es defineix no només per les paraules de la llengua natural de les seves etiquetes, sinó també pels seus vincles amb altres conceptes del vocabulari. SKOS proporciona tres propietats estàndards i bàsiques en tot sistema d'organització del coneixement, similars a les categories de les relacions jeràrquiques i associatives que es donen en els tesaurus: [skos:broader](#), [skos:narrower](#) i [skos:related](#)

- [skos:broader](#) i [skos:narrower](#) permeten la representació jeràrquica dels termes, representa la relació entre un gènere i les seves espècies específiques, o, depenent de la interpretació, la relació entre un tot i les seves parts.
- [skos:related](#) permet la representació de vincles associatius (no jeràrquics) com la relació entre un tipus d'esdeveniment i una categoria d'entitats que participen en aquesta. Un altre ús de `skos:related` és l'associació de dos categories entre les que ninguna és més general o més específica que l'altra. SKOS:related permet la representació de vincles associatius (no jeràrquics) que també es poden utilitzar per representar vincles de tipus part-tot, el significat dels quals és similar al de les relacions jeràrquiques.

En les relacions semàntiques de tipus genèric/específic s'utilitza la propietat **[skos:broader](#)** per indicar que un concepte té un significat més ampli o general que un altre. En canvi, s'utilitza la propietat **[skos:narrower](#)** per fer l'afirmació inversa, o sigui, s'assigna a aquells conceptes que tenen un significat més específic o concret.


```

ex:animales      rdf:type      skos:Concept;
skos:prefLabel   "animales"@es;
skos:narrower    ex:mamíferos.
ex:mamíferos     rdf:type      skos:Concept;
skos:prefLabel   "mamíferos"@es;
skos:broader    ex:animales.

```

Figura 43. Exemple d'etiquetes "narrower" i "broader" (SKOS)

Cal tenir en compte, al igual que passa en qualsevol sistema d'organització del coneixement, que un concepte SKOS pot associar-se a varis conceptes més amplis al mateix temps. Per exemple: un concepte *ex:perro* pot tenir a *ex:mamíferos* i *ex:animalesDomesticados* com a *conceptes genèrics*.

```

.ex:animales skos:prefLabel "animales"@es.
ex:mamíferos skos:prefLabel "mamíferos"@es;
skos:broader ex:animales.
ex:gatos skos:prefLabel "gatos"@es;
skos:broader ex:mamíferos.

```

Figura 44. Exemple de jerarquies transitives (SKOS)

Les declaracions implícites **skos:broader/skos:narrower** són mútuament inverses. Un concepte X és el genèric d'un altre concepte I, per tant, I és un concepte específic de X. Aquesta afirmació es dona tant en els sistemes d'organització del coneixement tradicionals com en el model de dades de SKOS. Seguint l'exemple dels "mamífers" podem dir que en molts casos trobarem que les relacions jeràrquiques d'un esquema de conceptes siguin considerades transitives (OWL, Transitive Property). Per exemple, si *ex:animales* és més genèric que *ex:mamíferos*, i aquest és més genèric que *ex:gatos*, es pot afirmar que *ex:animales* és més ampli que *ex:gatos*. Però, sí hi ha jerarquies incorrectes, sobretot en casos com els tesaurus que són dissenyats amb uns estàndards i aquesta funció no és apropiada. Considerem doncs que *ex2:vehículos* es més ampli que *ex2:coches*, i aquest més ampli que *ex2:ruedas*. Per tant, no es pot afirmar que "ruedas" és un concepte específic de "vehículos". Però SKOS permet anar definint aquesta transitivitat o intransitivitat i anar definit els problemes que puguin sorgir per tal de solucionar aquesta representació. Per veure més informació al respecte cal consultar la Guia de Referència SKOS (Alistair Miles i Sean Bechhofer ed., 2009).

Per tal de solucionar aquestes problemàtiques de jerarquies, SKOS utilitza la propietat transitiva *skos:broaderTransitive* com una super-propietat de *skos:broader* per a inferir les següents sentències:

ex:gatos	skos:broaderTransitive	ex:mamíferos.
ex:mamíferos	skos:broaderTransitive	ex:animales.

Figura 45. Exemple de jerarquia transitiva amb etiqueta skos:broader

Per tant, es pot fer la declaració següent: *ex:gatos skos:broaderTransitive ex:animales.*

La guia de referència SKOS, informa de que el seu model no indica que *skos:broader* i *skos:narrower* siguin etiquetes amb propietats transitives ni tampoc que siguin intransitives. Per exemple, seguint amb l'exemple de "gats" que és un concepte específic de "mamífers" i aquest és a la vegada un terme específic d'"animals", podem afirmar que "gats" és també un terme específic d'"animals" i al mateix temps es manté compatible amb el model SKOS. La no declaració d'*skos:broader* amb la propietat transitiva implica que no pot inferir-ne la propietat *skos:broader* entre "gats" i "animals" mitjançant els axiomes d'SKOS. No obstant això, no s'impideix que els editors d'un esquema de conceptes SKOS puguin afirmar que les declaracions de propietats jeràrquiques reflecteixen un comportament transitiu a nivell local.

En aquest sentit, SKOS no assumeix que les relacions jeràrquiques siguin irreflexives per defecte, tot i que en casos específics on es permet la reflexibilitat d'algunes declaracions *rdfs:subClassOf* es pot interpretar doncs com a un *skos:broader*. Per exemple, quan es considera la conversió d'una ontologia RDFS/OWL en un esquema de conceptes SKOS, en aquest cas sí és vàlida la declaració de que "rdfs:subClassOf" sigui reinterpretada amb *skos:broader*. Però *rdfs:subClassOf* és una propietat reflexiva, això significa que per cada classe C, la declaració C C:subClassOf C és certa ([OWL](#)). En aquest cas, tots els conceptes serien conceptes genèrics de si mateixos. En les directrius de molts tesaurus està prohibit decalar un concepte com a genèric de si mateix, i pel que fa a SKOS bàsic tampoc es contempen diferents tipus de relacions jeràrquiques com pot ser instància-classe o part-tot, més endavant s'exposen diferents solucions per a fer front a al problema (creació d'especialitzacions de relacions semàntiques).

Pel que fa a les relacions associatives, és a dir, l'establiment d'un vincle associatiu entre dos conceptes, SKOS ho representa amb l'etiqueta [skos:related](#).

ex:pájaros	rdf:type	skos:Concept;
skos:prefLabel		"pájaros"@es;
skos:related		ex:ornitología
ex:ornitología	rdf:type	skos:Concept;
skos:prefLabel		"ornitología"@es.

Figura 46. Exemple de relació associativa o "skos:related"

Com es descriu en la **Guía de referència de SKOS** [[SKOS-REFERENCE](#)], la propietat “skos:related” es [simètrica](#) (OWL). Així doncs, del gràfic RDF anterior es dedueix, per exemple, que ‘ex:ornitología’ es el subjecte d’una declaració “skos:related” que té com a objecte “ex:pájaros”.

Cal tenir en compte que la **(no-)transitivitat** d’*skos:related* en el model de dades d’SKOS, “skos:related”, no es defineix com una propietat transitiva. De fet, una relació skos:related transitiva pot tenir conseqüències no desitjades, veure exemple següent:

ex:medi ambient	skos:related	ex:efectes ambientals.
ex:efectes ambientals	skos:related	ex:selecció natural.
ex:selecció natural	skos:related	ex:competició.
ex:competició	skos:related	ex:competidors.

Figura 47. Exemple de relació transitiva "skos:related"

En l’exemple anterior (figura 47), el fet d’establir relacions de tipus transitiu, s’està afirmant que una propietat transitiva, com per exemple “medi ambient” està relacionada directament (*skos:related*) amb el concepte “competidors”. El resultat és l’establiment de la relació de dos conceptes que no estan relacionats i per tant hi ha una inconsistència entre termes. Si cada declaració individual té sentit, la declaració inferida no té cabuda en les intencions originalment previstes per l’editor del sistema d’organització.

La mescla d’estructures jeràrquiques i associatives fa que la relació transitiva *skos:broader* sigui disjunta a *skos:related*. Si els recursos A i B estan relacionats a través d’una relació “skos:related”, no s’ha d’establir una cadena de relacions *skos:broader* entre A i B, i el mateix passa en casos de representació de *skos:narrower*.

- 4. Notes:** són equivalents a les notes d’abast o definicions que normalment contenen els sistemes d’organització del coneixement. Les relacions semàntiques són molt importants per la definició dels conceptes. Juntament amb aquestes caracteritzacions estructurals els conceptes a vegades tenen que definir-se amb major precisió utilitzant documentació informal. Com ja s’ha explicat anteriorment la guia SKOS està inspirada en les directrius de tesaurus com la ISO2788 i la BS8723-2. Aquesta propietat especifica entre d’altres com “skos:scopeNote”, “skos:definition”, “skos:example” i “skos:historyNote” permet ajustar més els tipus específics de documentació que s’inclouen en un sistema.

- [skos:scopeNote](#) proporciona informació possiblement parcial, sobre el significat d'un concepte. Pot ser una indicació de com es limita l'ús del propi concepte en els processos d'indexació.

ex:hidrologia skos:scopeNote
"Useu-lo per a les obres que tracten de l'aigua com a part de les ciències naturals i considerada en tots els seus aspectes"@es.

Figura 48. Exemple de "Scope note" provinent de la Llista d'encapçalaments de matèries en català (LEMAC)

- [skos:definition](#) proporciona una explicació completa del significat d'un concepte. Exemple 11.1 Definició extreta del GEMET:

ex:atmosphere skos:definition
"The gaseous envelope surrounding the Earth in a several kilometers-thick layer. (Source: UVAROVA)"

Figura 49. Exemple d'etiqueta "skos:definition" (Definició extreta de GEMET)

- [skos:example](#) proporciona un exemple de l'ús d'un concepte.

ex:organizacionesCienciaCultura skos:example
"Académicas científicas, museos generales, ferias mundiales"@es.

Figura 50. Exemple d'etiqueta "skos:example"

- [skos:historyNote](#) descriu canvis en el significat o la forma d'un concepte.

ex:abusoNiños skos:historyNote
"establecido en 1975; previamente se usaba: Crueldad con los niños [1952-1975]"@es.

Figura 51. Exemple d'etiqueta nota "skos:historyNote"

SKOS inclou també dos especialitzacions d'*skos:note* d'utilitat per als gestors o editors de sistemes d'organització: *skos:editorialNote* i *skos:changeNote*.

- [skos:editorialNote](#) subministra informació de manteniment, com per exemple, recordatoris de treball editorial que queda per fer o advertències per al ús de futurs canvis editorials que puguin fer-se.

```

ex:dobleclick
skos:editorialNote "Revisar este término después
de completar la fusión de la compañía"@es

ex:folksonomía
skos:editorialNote "Comprobar la ortografía con
Thomas Vander Wal"@es.

```

Figura 52. Exemple d'etiqueta nota "skos:editorialNote"

- [skos:changeNote](#) documenta aquells canvis molt particulars a efectes d'administració i manteniment.

```

ex:tomate skos:changeNote
"Clasificado previamente bajo 'frutas' i
actualmente bajo 'vegetales' por Horace
Gray"@es.

```

Figura 53. Exemple d'etiqueta nota "skos:changeNote"

És important notar el vincle jeràrquic entre *skos:note* i les seves diferents especialitzacions permet que tota la documentació associada a un concepte sigui recuperada de manera directa. Cada *skos:definition* es una *skos:note*, cada *skos:scopeNote* és un *skos:note*, i així successivament. Com s'exemplifica anteriorment, en les propietats de documentació d'SKOS es poden utilitzar literals RDF simples.

SKOS ofereix altres patrons possibles ja que aquestes propietats no poden limitar-se a literals ([secció 4.2](#) del manual SKOS). Una característica important dels literals simples és la capacitat d'utilitzar etiquetes d'idioma com en el cas de l'etiqueta't de propietats. Per tant, la informació pot estar disponible en diferents idiomes. Veure exemple següent:

```

ex:piñas rdf:type skos:Concept;
skos:prefLabel "piñas"@es;
skos:prefLabel "pineapples"@en;
skos:definition "Fruta de un tipo de plantas de la familia de las
Bromeliáceas"@es;
skos:definition "The fruit of plants of the family
Bromeliaceae"@en.

```

Figura 54. Exemple d'ús t'etiqueta't d'idiomes en etiquetes de notes (SKOS)

A més, altres propietats no pertanyents a SKOS poden servir per a documentar els conceptes. Per exemple, la propietat de Dublin Core "[dct:creator](#)", pot ser utilitzada per indicar la persona que va crear el següent concepte:

```

"ex:águilaPescadoraMadagascar dct:creator [ foaf:name "John Smith" ]"

```

5. Esquemes de conceptes: els conceptes es poden crear i usar com a entitats independents. Però sobretot durant la indexació els conceptes

generalment són associats a un vocabulari cuidadosament recopilat (tesaurus o esquemes de classificació). SKOS ofereix també els mecanismes per a la representació de tals sistemes utilitzant la classe [skos:ConceptScheme](#). A continuació es mostra un exemple amb la definició d'un esquema de conceptes d'un tesaurus i la descripció d'aquest recurs mitjançant les etiquetes propietats de Dublin Core [dct:title](#) i [dct:creator](#):

```
ex:tesaureAGROVOC
rdf:type skos:ConceptScheme;
dct:title "AGROVOC thesaurus ";
dct:creator ex:FAO.
```

Figura 55. Exemple d'esquema de conceptes (SKOS)

Un cop creat l'esquema de conceptes, aquest pot enllaçar-se amb altres conceptes que el continguin, utilitzant la propietat [skos:inScheme](#):

```
ex:mamíferos rdf:type skos:Concept;
skos:inScheme ex:tesauroAnimales.
ex:vacas rdf:type skos:Concept;
skos:broader ex:mamíferos;
skos:inScheme ex:tesauroAnimales.
ex:peces rdf:type skos:Concept;
skos:inScheme ex:tesauroAnimales.
```

Figura 56. Exemple de propietat "skos:inScheme" d'esquema de concepte

Amb la finalitat de proporcionar un accés eficaç als punts d'entrada de les jerarquies de conceptes genèrics/específics, SKOS defineix la propietat "[skos:hasTopConcept](#)". Aquesta propietat permet vincular un esquema de conceptes d'un a molts conceptes de capçalera.

```
ex:tesaureAGROVOC rdf:type
skos:ConceptScheme;
skos:hasTopConcept ex:mamíferos;
skos:hasTopConcept ex:elefante.
```

Figura 57. Exemple de propietat "skos:hasTopConcept" d'esquema de concepte

Els esquemes de conceptes SKOS estan dissenyats per representar vocabularis tradicionals i s'anima als dissenyadors a seguir les directrius existents d'elaboració de sistemes d'organització del coneixement com la ISO2788 o la BS8723-2. Com ja s'ha explicat anteriorment, quan s'elabora un esquema de conceptes basat en SKOS es recomana que no hi hagi 2

conceptes amb la mateixa etiqueta lèxica preferent en un idioma determinat (vegeu secció Etiquetes).

Tot i les similituds amb les directrius de KOS i SKOS hi ha algunes diferències entre els esquemes de conceptes i els sistemes d'organització del coneixement tradicionals, principalment degut al context del web semàntic en el que es desenvolupa SKOS. En la [secció 4.6](#) de la guia de referència d'SKOS ([SKOS-REFERENCE](#)) es poden trobar aquestes casuístiques, però cal destacar que una de les característiques importants d'SKOS és la possibilitat que el mateix concepte estigui associat a diferents esquemes usant la propietat *skos:inScheme*.

Per últim, cal remarcar que el vocabulari d'SKOS només ofereix un recolzament limitat per a la representació de la informació d'un sistema d'organització del coneixement, que es dona a través de les propietats *skos:inScheme* i *skos:hasTopConcept* que permeten vincular conceptes i esquemes de conceptes. Però, encara no existeix ningun mecanisme en SKOS per a deixar constància de que una declaració específica sobre aquests conceptes, com per exemple, una declaració "skos:broader", es refereix a un determinat esquema de conceptes. En canvi, en un sistema d'organització del coneixement generalment si es visualitza una estructura que consta tant de conceptes com de vincles que els defineixen. La guia fa una proposta en referència aquest tema (vegeu [secció 5.3](#), [NAMED-GRAPHS](#)) per solucionar problemes com gràfics associats a una URI. Sí és cert que un esquema de conceptes SKOS pot estar relacionat amb un conjunt de dades [RDF](#), o inclús definint per dit conjunt, el que permet la creació de consultes [SPARQL](#) (llenguatge de consulta per a RDF) les quals s'encarreguen de proporcionar alguna forma de procedència o de contenció.

2.2.2.4. SKOS avançat

Respecte als sistemes d'organització del coneixement no són tan simples, per tant, SKOS proporciona una sèrie d'elements de vocabulari o directius que s'ocupen de les necessitats de representació més avançades, per tant, SKOS és adaptable i compatible amb una ampla gama d'enfocaments de models per a sistemes d'organització del coneixement. Aquests tipus de sistemes són especialment dissenyats per satisfer les necessitats que es plantegen en casos d'ús concrets ([SKOS-UCR](#)):

- Agrupació de conceptes basats en criteris específics,
- Documentació avançada per mitjà de recursos complexos,
- Establiment de relacions entre les etiquetes dels conceptes,
- Creació de conceptes complexos a partir d'altres senzills (coordinació),
- Accés transitiu a relacions jeràrquiques,
- Representació de notacions de conceptes.

SKOS permet definir grups significatius o col·leccions de conceptes. Tals agrupacions són normalment mostrades en els tesaurus.

llet		
<llet	d'origen	animal>
llet de vaca		
llet de cabra		
llet de búfal		

Figura 58. Exemple de col·leccions de conceptes (SKOS)

Aquestes col·leccions es poden utilitzar per representar conjunts de dades en la terminologia de desenvolupament del tesaurus corresponent. Seguint l'exemple anterior (figura 58) el terme "leche de 'origen animal" és l'etiqueta d'un nexe que contempla d'altres conceptes, ja que hi ha un consens en relació a que l'etiqueta d'un node no representa una etiqueta per a un concepte per dret propi, sinó que han d'introduir-se entitats específiques per representar-les.

Col·leccions etiquetades: per poder modelar correctament les estructures de col·leccions de conceptes SKOS introdueix la classe "[skos:Collection](#)". Les instàncies d'aquesta classe agrupen conceptes específics mitjançant la propietat "[skos:member](#)" tal i com es mostra en l'exemple següent:

```
ex:leche rdf:type skos:Concept;  
skos:prefLabel "leche"@es.  
ex:lecheVaca rdf:type skos:Concept;  
skos:prefLabel "leche de vaca"@es;  
skos:broader ex:leche.  
ex:lecheCabra rdf:type skos:Concept;  
skos:prefLabel "leche de cabra"@es;  
skos:broader ex:leche.  
ex:lecheBúfalo rdf:type skos:Concept;  
skos:prefLabel "leche de búfalo"@es;  
skos:broader ex:leche.  
  
<My collection> rdf:type skos:Collection;  
skos:prefLabel "leche de origen animal"@es;  
skos:member ex:lecheVaca;  
skos:member ex:lecheCabra;  
skos:member ex:lecheBúfalo.
```

En aquest exemple, la col·lecció es defineix com un nexe buit, és a dir, no es defineix una URI associada a la col·lecció. Les URIs es poden assignar a les col·leccions però generalment no és necessari. A més, s'utilitza *skos:prefLabel* per assignar una etiqueta lèxica a la col·lecció, ja que aquesta propietat, com altres propietats t'etiqueta't SKOS, es pot utilitzar amb altres recursos no conceptuals.

També podem ordenar les col·leccions (alfabèticament, cronològicament, etc.) a través de la classe [skos:OrderedCollection](#), juntament amb la propietat

[*skos:memberList*](#). Aquesta propietat vinculada a una instància d'*skos:OrderedCollection* a un node de tipus *rdf:List* (possiblement buit), seguint el model que permet la definició de les [col·leccions RDF \(RDF-PRIMER\)](#). Per exemple:

```

ex:bebés rdf:type skos:Concept;
skos:prefLabel "bebés"@es.
ex:niños rdf:type skos:Concept;
skos:prefLabel "niños"@es.
ex:adultos rdf:type skos:Concept;
skos:prefLabel "adultos"@es.

_:b0 rdf:type skos:OrderedCollection;
skos:prefLabel "Personas según la
edad"@es;
skos:memberList _:b1.
_:b1 rdf:first ex:bebés;
rdf:rest _:b2.
_:b2 rdf:first ex:niños;
rdf:rest _:b3.
_:b3 rdf:first ex:adultos;
rdf:rest rdf:nil.

```

Figura 60. Exemple d'ordenació de col·leccions (*skos:OrderedCollection*)

Col·leccions SKOS, relacions semàntiques i presentacions sistemàtiques: cal tenir en compte que segons el model de dades SKOS, les col·leccions són disjunctes dels conceptes. Per tant, és impossible utilitzar les relacions semàntiques d'SKOS (vegeu [secció 2.3](#) de la guia de referència SKOS), per ubicar una col·lecció en una xarxa semàntica d'SKOS. En altres paraules, les agrupacions de conceptes en col·leccions no substitueixen a les declaracions sobre la localització dels conceptes en un esquema de conceptes. En l'exemple anterior sobre la "llet", els diferents tipus de "llet" definits s'han de vincular explícitament a un concepte més genèric *ex:leche* utilitzant la propietat *skos:broader*. Per exemple:

```

ex:lecheVaca skos:broader ex:leche.
ex:lecheCabra skos:broader ex:leche.
ex:lecheBúfalo skos:broader ex:leche.

```

Figura 61. Exemple de vinculació de "*skos:collection*" amb la propietat "*skos:broader*"

Per tant, es possible generar una presentació sistemàtica (jeràrquica) que inclogui l'agrupació de conceptes "llet d'origen animal". La jerarquia *skos:broader* i la informació sobre la pertinença a una col·lecció es pot utilitzar per fer aquesta agrupació, però cal tenir en compte que per aquest procés encara requereix un algoritme especialitzat.

Probablement, l'ús de col·leccions és complex de representar per una aplicació, sobretot en aquells sistemes d'organització del coneixement destinats

principalment a estructures jeràrquiques de navegació, on pareix més intuïtiu representar “etiquetes de nexes” o “termes guia” com a instàncies de *skos:Concept*, i utilitzar les relacions semàntiques normals per a vincular-les a altres conceptes. Exemple:

```

ex3:lecheOrigenAnimal    rdf:type    skos:Concept;
skos:prefLabel    "leche de origen animal"@es;
skos:broader                ex3:leche;
skos:narrower                ex3:lecheVaca;
skos:narrower                ex3:lecheCabra;
skos:narrower ex3:lecheBúfalo.

```

Figura 62. Exemple de vinculació de col·leccions jeràrquiques a través "skos:Concept"

La representació de les dues opcions d'agrupació de col·leccions està oberta depenent de l'aplicació que s'utilitzi. Però l'W3C no recomana l'ús de les col·leccions ja que pot suposar una pèrdua de precisió semàntica. Moltes aplicacions descriptives resulten de naturalesa molt específica i no s'han d'usar en els índexs habituals de manera conjunta amb les etiquetes de conceptes, de manera que la representació de les etiquetes de conceptes no és una pràctica recomanable.

Relacions entre etiquetes: algunes aplicacions requereixen la creació de vincles entre etiquetes associades a conceptes. Per exemple, considerar la relació entre una etiqueta preferent per al concepte “Corporació” i la seva abreviatura, “Corp.” declarada com a una etiqueta alternativa, o un vincle per a denotar la traducció entre dos etiquetes en diferents idiomes: "Vaca"@es i "Cow"@es. L'ús de les propietats SKOS per al etiquetat lèxic, com per exemple *skos:prefLabel*, es limita a utilitzar literals RDF. Per tant, aquestes etiquetes no poden ser objectes d'una declaració RDF, de manera que no es possible especificar una relació directa entre elles. Per a solucionar-ho, s'ha ampliat el vocabulari SKOS amb una extensió opcional per a etiquetes amb l'[SKOS-XL](#) (vegeu SKOS-REFERENCE, SKOS-XL). Aquesta extensió presenta la classe *skosxl:Label* que permet tractar les etiquetes amb recursos RDF de primer ordre. Cal tenir en compte que cada instància d'aquesta classe ha d'ajuntar-se en primer lloc a un literal RDF simple a través de la propietat *skosxl:literalForm*. Per exemple, el concepte "Organización de las Naciones Unidas" que pot estar etiquetat pel seu nom oficial o per l'acrònim de la organització, es pot representar de la següent manera:

```

ex:ONUetiqueta1    rdf:type    skosxl:Label;
skosxl:literalForm    "Organización de las Naciones
Unidas"@es.
ex:ONUetiqueta2    rdf:type    skosxl:Label;
skosxl:literalForm    "ONU"@es.

```

Figura 63. Exemple de propietat "skosxl:literalForm"

Les instàncies de *skosxl:Label* poden estar relacionades amb conceptes utilitzant certes propietats ([*skosxl:prefLabel*](#), [*skosxl:altLabel*](#), [*skosxl:hiddenLabel*](#)) que reflecteixin el tipus de literal d'etiquetat utilitzat. Per últim, aquestes instàncies poden vincular-se entre si mitjançant declaracions especificades amb la propietat [*skosxl:labelRelation*](#):

```

ex:ONU      rdf:type
skos:Concept;
skosxl:prefLabel
ex:ONUetiqueta1;
skosxl:altLabel
ex:ONUetiqueta2.
ex:ONUetiqueta2
skosxl:labelRelation
ex:ONUetiqueta1.

```

Figura 64. Exemple 1 de propietat "skosxl:labelRelation"

La solució anterior no és totalment satisfactòria, ja que una aplicació que sigui "sensible" als acrònims pot perdre la informació real de que entre les dues etiquetes hi ha una relació d'acrònimia. A més, es perdria informació sobre la direcció a la que apunta el vincle. Conseqüentment, es recomana que en cas d'utilitzar SKOS-XL s'especialitzi la propietat *skosxl:labelRelation* amb la finalitat de satisfer les necessitats de certes aplicacions, tal i com es mostra a continuació:

```

ex:esAcrónimoDe rdfs:subPropertyOf
skosxl:labelRelation.
ex:ONUlabel2 ex:esAcrónimoDe ex:ONUetiqueta1

```

Figura 65. Exemple 2 de propietat "skosxl:labelRelation"

El model SKOS-XL és compatible amb l'estàndard SKOS. Si una instància d'*skosxl:Label* s'adjunta a un concepte mitjançant una declaració amb *skosxl:altLabel*, s'entén que a través del model de dades SKOS-XL la forma literal de la instància *skosxl:Label* es relaciona amb el concepte a través d'una declaració estàndard amb *skos:altLabel*. En l'exemple anterior, ex:ONU té associada una etiqueta alternativa en la seva forma literal "ONU"@es".

Coordinació de conceptes: El procés d'indexació usada amb els tesaurus i altres sistemes d'organització molts cops utilitzen la noció de coordinació. La coordinació és la activitat en la que es combinen els conceptes d'aquest tipus de sistemes. En general hi ha dos tipus de coordinació: pre-coordinació i post-coordinació. La diferència fonamental entre els dos conceptes depèn de quan es produeix la coordinació en relació al moment en que es realitza la recuperació de la informació.

La pre-coordinació és dóna abans de la recuperació de la informació, i es produeix per part del gestor del sistema d'organització del coneixement o per un indexador que està utilitzant-lo. Per exemple, si un indexador usa dos conceptes d'un esquema de conceptes com "Bicicleta" i "Reparació", i explícitament els combina amb una sintaxis determinada com "Bicicletes—Reparació" per assignar una matèria única a un document concret.

En canvi, la post-coordinació es dóna durant la recuperació de la informació, per exemple, si s'indexa un determinat document amb dos conceptes diferents "Bicicletes" i "Reparació" i un usuari decideix realitzar una cerca de tots els documents que s'estan indexant amb aquests conceptes. Així doncs, en aquest procés la post-coordinació és una representació indirecta com una consulta SPARQL per accedir a les dades RDF. Per exemple:

```
Donat aquests conceptes:  
ex:bicicletas skos:prefLabel "Bicicletas"@es.  
ex:reparación skos:prefLabel "Reparación"@es.  
  
Consulta SPARQL que retorna únicament docuemnts que  
indexen els 2 conceptes:  
SELECT ?document  
WHERE {  
  ?document dct:subject ex:bicicletas.  
  ?document dct:subject ex:reparación.  
}
```

Figura 66. Exemple de post-coordinació amb consulta SPARQL

Però SKOS no proporciona per si mateix ningun mecanisme per expressar que un determinat concepte és en realitat una coordinació prèvia d'altres. Per suposat, és perfectament factible una coordinació prèvia d'altres conceptes, i que és possible estendre SKOS per establir un model de representació de conceptes coordinats. Per exemple, s'ha suggerit que es pugui definir una nova propietat *ex:coordinationOf*:

```
ex:coordinationOf a rdf:Property;  
rdfs:domain skos:Concept;  
rdfs:range rdf:List.
```

Figura 67. Exemple 1 de coordinació entre termes amb propietat "ex:coordinationOf" (SKOS)

Es pot utilitzar també en afirmacions com:

```
ex:reparaciónBicicletas a skos:Concept;  
ex:coordinationOf (ex:bicicletas ex:reparación);  
skos:prefLabel "Bicicletas--Reparación"@es.
```

Figura 68. Exemple 2 de coordinació entre termes amb propietat "ex:coordinationOf" (SKOS)

El W3C, a través de la guia SKOS, suggereix que OWL es podria utilitzar també per coordinar conceptes. Per exemple:

```
ex:reparaciónBicicletas a skos:Concept;  
owl:intersectionOf (ex:bicicletas ex:reparación);  
skos:prefLabel "Bicicletas--Reparación"@es.
```

Figura 69. Exemple 3 de coordinació entre termes amb propietat "owl:intersectionOf" (OWL)

En canvi els models establerts per a la pre-coordinació de conceptes encara no han sorgit en la comunitat SKOS. SKOS pot estendre's per representar conceptes coordinats, inclús s'ha suggerit una propietat *ex:coordinationOf* que pot ser combinada amb l'ús d'OWL⁴² però encara no s'han explorat suficientment per justificar la seva inclusió en el vocabulari SKOS. W3C està treballant en una extensió del vocabulari SKOS⁴³ que ho permeti fer.

Jerarquies transitives

Com s'ha descrit anteriorment les propietats per representar les jerarquies d'un sistema d'organització *skos:broader* i *skos:narrower* no es defineixen com a transitives. La semàntica no és compatible amb inferències de tipus "animals" és més genèric que "mamífers" i "mamífers" és més genèric que "gats", i no on "animals" és més ampli que "gats". Per aplicacions que precisen de tal semàntica, per exemple, per a realitzar la expansió d'una consulta, SKOS disposa de dos propietats específiques, [*skos:broaderTransitive*](#) i [*skos:narrowerTransitive*](#). Estan definides com a super-propietats transitives d'*skos:broader* i *skos:narrower* (SKOS Reference, 2009). Aquest model permet, usar eines d'inferència per a el web semàntic i accedir a una clàusula transitiva d'una jerarquia expressada amb *skos:broader* i *skos:narrower*. Per exemple:

```
ex:animales skos:prefLabel "animales"@es.  
ex:mamiferos skos:prefLabel "vertebrats"@es;  
skos:broader ex:animales.  
ex:gatos skos:prefLabel "tortugues"@es;  
skos:broader ex:vertebrats.
```

Figura 70. Exemple de jerarquia transitiva (SKOS)

⁴²4.6.2. SKOS Concept Schemes and OWL Ontologies. *SKOS Reference*, 2009. Url: <http://skos.um.es/TR/skos-primer/#secskosowl>

Al llegir les declaracions anteriors un sistema pot fer ús de la definició “skos:broaderTransitive” com una super-propietat ”skos:broader” per inferir les següents sentències:

```
ex:tortugues skos:broaderTransitive ex:vertebrats
ex:vertebrats skos:broaderTransitive ex:animals.
```

Figura 71. Exemple 1 de jerarquia transitiva amb super-propietat "skos:broader"

La transitivitat de la propietat “skos:broaderTransitive” fa que es pugui deduir la declaració desitjada:

```
ex:tortugues skos:broaderTransitive ex:animales.
```

Figura 72. Exemple 2 de jerarquia transitiva amb super-propietat "skos:broader"

L'ús de la super-propietat *skos:broaderTransitive* permet explotar les interpretacions transitives de les xarxes jeràrquiques dels conceptes com millor encaixin, mentre no interfereixin amb la semàntica de la propietat *skos:broader*, la qual no pot aplicar aquesta transitivitat. Es pot interpretar que les declaracions amb *skos:broader* són suficientment explícites per assegurar l'existència de vincles directes amb els conceptes superiors mentre que *skos:broaderTransitive* s'usa per a reflectir de forma més general i possiblement indirecta les relacions amb aquells que tinguin una posició més ancestral.

Una propietat no-transitiva es deriva a partir d'una altra que si és transitiva però no hereta la seva transitivitat. No obstant això, és compatible amb la semàntica d'RDFS/OWL per *rdfs:subPropertyOf* (*OWL Reference*, 2004), on una propietat P és una sub-propietat de Q, i en cas de que P es situï entre dos recursos, llavors Q també es dona entre ells, però tampoc garantirà la herència de la transitivitat.

Notacions: alguns sistemes d'organització del coneixement com la CDU (Classificació Decimal Universal), utilitzen notacions com mitjà principal d'accés als conceptes que contenen. Aquestes notacions són símbols que normalment no són reconegudes com a paraules o seqüències de paraules en qualsevol llenguatge natural, per això són utilitzables independentment de dit context. Són dígitos que són complimentats amb signes de puntuació, caràcters, etc. Per exemple:

```
512 Álgebra
512.6 Ramas especiales del Álgebra
```

Figura 73. Exemple de notacions CDU

SKOS permet també representar aquest tipus de notacions a través de la propietat *skos:notation*, aquesta propietat es pot associar a un concepte literal tipificat en [RDF](#) (literal de tipus explícit, *RDF-Primer*, 2004). El tipus de dades

literal especifica la sintaxis d'un esquema de codificació que s'inscriu en l'ús de notacions en el sistema d'organització en particular. El valor del literal és la notació en ella mateixa, en aquest cas el propi codi de classificació. Per exemple:

```
ex:cdu512 skos:prefLabel "Álgebra"@es ;  
skos:notation "512"^^ex:NotaciónCDU.
```

Figura 74. Exemple de notació CDU amb SKOS

El model anterior no és realment necessari quan els editors consideren a les notacions en si mateixes com un llenguatge senzill d'etiquetes independents. Llavors es pot usar una propietat *skos:prefLabel* sense indicador d'idioma, com:

```
ex:cdu512 skos:prefLabel "Álgebra"@es ;  
skos:notation "512"^^ex:NotaciónCDU ;  
skos:prefLabel "512" .
```

Figura 75. Exemple de notació amb "skos:prefLabel"

En la [secció 6.5.1](#) de la guia SKOS Reference (2009) es poden trobar més detalls sobre com utilitzar aquest tipus de dades.

Especialització del model SKOS

SKOS està pensat com un denominador comú entre els diferents enfocaments de modelat. La especificació actual del vocabulari permet que molts sistemes d'organització del coneixement existents poden ser transferits a el web semàntic. Però hi ha tanta varietat de models d'aquest tipus de sistemes que fan impossible que es puguin captar tots els detalls d'aquests sistemes, i no tots els sistemes que empen SKOS conserven la primera "S" (Simple), és a dir, no tot són sistemes simples d'organització del coneixement. No obstant això, les aplicacions que necessitin un nivell de detall major es beneficien enormement d'SKOS al tractar-se d'un vocabulari pensat per al web semàntic, ja que SKOS es pot ampliar i adaptar-se a les necessitats específiques d'un grup d'usuaris d'un sistema d'organització del coneixement en particular, mantenint la compatibilitat amb les aplicacions que es fonamenten en les característiques bàsiques d'SKOS. Inclús un mateix usuari pot crear les seves pròpies característiques i classes, juntament amb elements estàndards d'SKOS, usant les propietats *dfs:subPropertyOf* i *rdfs:subClassOf* del vocabulari de l'esquema RDF (RDF-Premier, 2004).

Amb la propietat *skosxl:labelRelation* es pot construir una propietat semànticament més rica dedicada a la representació de la vinculació d'acrònims. També són possibles altres usos com la creació de diferents variants de propietats "skos:broader" i "skos:narrower". Les normes sobre tesaurus identifiquen un reduït conjunt de tipus de relacions jeràrquiques, com

per exemple de tipus genèric, part-tot, o instància-classe (ISO2788). L'enfocament d'SKOS permet al dissenyador d'aplicacions crear noves propietats diferenciant la declaració de sub-propietats d'*skos:broader*.

```
ex:broaderGeneric rdfs:subPropertyOf skos:broader.  
ex:broaderPartitive rdfs:subPropertyOf skos:broader.  
ex:broaderInstantive rdfs:subPropertyOf
```

Figura 76. Exemple de propietat "rdfs:subPropertyOf" amb "skos broader"

A més, SKOS pot combinar-se amb altres vocabularis de modelat. SKOS és un vocabulari RDF/OWL que es pot ampliar-se fàcilment per adaptar-se a les necessitats concretes dels usuaris. Les característiques d'SKOS també es poden utilitzar al web semàntic com a complement d'altres vocabularis de modelat. Tal i com s'observen en els exemples següents:

- **Usar propietats d'etiquetat SKOS per etiquetar recursos de tipus *skos:Concept*.** Exemple que descriu a Tim Berners-Lee:

```
<http://www.w3.org/People/Berners-Lee/card#i> rdfs:type foaf:Person;  
foaf:name "Timothy Berners-Lee";  
rdfs:label "TBL";  
skos:prefLabel "Tim Berners-Lee"@en.
```

Figura 77. Exemple 1 de combinació de formats semàntics (SKOS i RDF)

En aquest exemple, veiem com una aplicació pot identificar a "Tim Berners-Lee" com a etiqueta preferent en lloc de tenir que escollir entre 2 etiquetes compatibles, [rdfs:label](#) "TBL" o [foaf:name](#) "Timothy Berners-Lee", aquestes etiquetes són compatibles perquè *foaf:name* és una sub-propietat de *rdfs:label*.

En general, la capacitat de reutilitzar els elements del vocabulari SKOS i altres vocabularis RDF segons es precisi, de fet és el que dona a RDF gran part del seu poder expressiu.

- **Conceptes SKOS i classes OWL**

La Guia de referencia d'SKOS defineix un "skos:Concept" com una classe OWL (SKOS-Reference, 2004):

```
skos:Concept rdfs:type owl:Class.
```

Figura 78. Exemple 2 de combinació de formats semàntics (SKOS i OWL)

Les instàncies *skos:Concept* (ex:Pintura en un vocabulari) són individus en termes de OWL.

```
ex:Pintura rdfs:type skos:Concept.
```


Figura 79. Exemple 3 de combinació de formats semàntics (SKOS i OWL)

Això fa que es plantegi si una instància d'un concepte SKOS com *ex:Painting* pugui tractar-se com una classe per si mateixa. Els usuaris, podrien definir propietats per "painting" com un tipus de suport:

```
ex:soporte rdf:type owl:DatatypeProperty.  
ex:soporte rdfs:domain ex:Pintura.
```

Figura 80. Exemple 4 de combinació de formats semàntics (SKOS i OWL)

Conceptualment, una classe com *skos:Concept* pot ser vista com una metaclasse. Les seves instàncies serien els conceptes que apareguin en un vocabulari. Per tant, és comprensible que els usuaris d'SKOS especifiquin les característiques a nivell de classe de conceptes, per exemple, que les "pintures" tenen un suport o que el producte "formatge" tingui un país d'origen. Cal tenir en compte que SKOS no agafa cap postura al respecte de quina variant OWL utilitzar (OWL-Reference): OWL Lite, OWL Full o OWL-DL. Els usuaris d'OWL Full seran capaços d'utilitzar l'escenari anterior mitjançant el tractament dels conceptes SKOS explícitament com a classes, per exemple, mitjançant la addició de les declaracions del tipus:

```
ex:Pintura rdf:type owl:Class.
```

Figura 81. Exemple 4 de combinació de formats semàntics (SKOS i OWL)

Això és possible perquè OWL Full no requereix que el conjunt de classes i individus sigui disjunt. Les persones que desitgin utilitzar la variant DL d'OWL no poden utilitzar aquest mecanisme de metamodelat, posat que és imprescindible que es compleixi la condició de disjunció entre classes i individus en tota ontologia OWL-DL. Els usuaris d'OWL-DL que estiguin interessats en la vinculació de classes OWL a conceptes SKOS han de mantenir la distinció formal. Però es poden utilitzar les propietats d'anotació OWL per salvar aquesta limitació, sempre que creïn i utilitzin les seves pròpies extensions per SKOS, com:

```
ex:ClassePintura rdf:type owl:Class.  
ex:ConceptoPintura rdf:type skos:Concept.  
ex:ClassPintura ex:correspondenciaConcepto ex:ConceptoPintura.
```

Figura 82. Exemple 3 de combinació de formats semàntics (SKOS i OWL)

En conjunt, OWL i SKOS són formats que evolucionen constantment, i normalment són també usats amb finalitats diferents, no obstant, hi ha tres consideracions importants a tenir en compte sobre les seves característiques:

- Els conceptes d'SKOS són individus d'OWL (per tant, estan formalment relacionats),
- SKOS no agafa cap postura sobre si també ha de ser possible tractar els conceptes d'SKOS amb classes d'OWL (pot haver certa flexibilitat en l'ús del format, però alhora inconsistència d'ús, en una comunitat d'usuaris caldran directius d'aplicació),
- Les restriccions d'OWL-DL eviten que els conceptes d'SKOS es tractin com a classes d'OWL (en cas que s'hagin de combinar els dos formats, caldrà tenir present l'ús d'OWL-DL com a format consistent).

En conclusió, els grups de treball d'OWL i SKOS estan en continu desenvolupament de les pròpies directrius i per això cal anar observant l'evolució dels dos models i la seva relació, sobretot quant a l'ús i aplicació entre conceptes d'SKOS i les classes/individus d'OWL.

4.3. Integració dels formats semàntics en biblioteques

L'aparició de tecnologies del web semàntic, tot i que les primeres iniciatives són bastant recents, han marcat tendència també entre la comunitat bibliotecària. Entre el maig del 2010 i agost del 2011 es crea el grup incubador del W3C sobre *Library Linked Data Incubator Group* o grup de dades vinculades en biblioteques. El terme *linked data* és definit per la pròpia comunitat (vegeu pàgina principal de la *Linked Data Community*) com el fet de connectar dades web relacionades que no estan vinculades prèviament. No obstant això, hi ha un altre terme relacionat que és *Linked Open Data* o LOD (Peset, et al., 2011), concepte relacionat amb l'*Open Data* que fa referència al moviment d'alliberació de dades en les organitzacions, promogut per l'*Opening up Government* (2010), on es fan accessibles dades a través de formats intercanviables i amb una llicència per al seu ús i reutilització, la diferència en el cas de *Linked Open Data* és que es parla de dades web amb format RDF.

La missió del *Library Linked Data Incubator Group* és ajudar a augmentar la interoperabilitat global de dades de les biblioteques al web. L'objectiu és fomentar la cooperació de les persones involucrades en les activitats del web semàntic, sobretot se centren en el *linked data* i la comunitat bibliotecària, conèixer les iniciatives existents i la identificació de vies de col·laboració futures. Els primers indicis d'aparició del *linked data* en la comunitat bibliotecària (Dunsire, G., et al. 2012) es va donar en la publicació de la LCSH (*Library of Congress Subject Headings*) en format SKOS, entre 2008 i 2009 (Policy and Standards Division Annual Report 2009). Paral·lelament també es crea el catàleg LIBRIS de la Biblioteca Nacional Sueca que utilitzava el *linked data* en els seus registres bibliogràfics (Malmsten, M. 2008, LIBRIS, 2009). Posteriorment apareixen d'altres iniciatives de biblioteques nacionals com és el cas de França, Alemanya i la British Library (Regne Unit), fins arribar al 2011 amb la difusió del portal Europeana que ja tenia una primera versió al 2009 (Europeana version 1.0. January 2009 - September 2011). En el cas

d'European, no només s'integren dades de biblioteques, sinó també s'inclouen dades de museus i arxius. La iniciativa més recent en el camp de biblioteques, segons la comunitat *W3C Libray Linked Data*, es dona l'estiu del 2012 amb OCLC, amb la iniciativa de la *Library of Congress* que està treballant amb la transició del model bibliogràfic tradicional del catàleg al model basat en *linked data*. Tot i que molts projectes bibliotecaris relacionats amb el web semàntic es concentren en l'any 2011, no és fins a l'estiu del 2012 quan la majoria d'aquests projectes comencen a ser visibles. Els darrers informes publicats del grup incubador del *W3C Libray Linked Data* incloent casos d'ús és l'octubre del 2011⁴⁴.

És difícil determinar l'expansió de formats/estàndards del web semàntic en biblioteques i/o el nombre de projectes de *linked data* en biblioteques, però a través de la *Library of Congress* on es registren les activitats externes a la *Library of Congress* sobre ús de LCSH amb el format SKOS (Policy and Standards Division Annual Report, 2010) i el *W3C (Library Linked Data Incubator Group, 2010)*, es poden citar els casos registrats recentment per aquestes organitzacions:

Projectes Desenvolupats sobre <i>linked data</i> i biblioteques	Anys
LCSH – SKOS (LC, 2010)	2008-2009
Europeana Version 1.0	2009
HIVE project (Metadata Research Center, University of North Carolina): aplicació automàtica de la terminologia LCSH a la informació bibliogràfica.	2010
<i>Library of Congress</i> Subject Headings (ZVON.org); robust LCSH searching geared toward human users	2010
EUROVOC (EU); supporting a developing, interoperable union list of vocabularies in the EU	2010
National Diet Library (Japan); using both LCSH in English and Japanese headings	2010
STW web Services (German National Library of Economics); using economics headings from LCSH applied to bibliographic information	2010
TELplus WP3.1 (The European Library); full-text search engine prototype using LCSH subject heading vocabulary	2010

⁴⁴ Nota: darreres consultes fetes al Desembre del 2013.

Browsing LCSH (Universitätsbibliothek Braunschweig); browser	2010
LC Subjects.org; provides a mirror site to id.loc.gov and allows institutions to create their own URIs and do experiments with LCSH data.	2010
<u>Bibliographic Network</u> : <i>linked data</i> techniques would allow bibliographic records to be described as an information graph using web standards to facilitate users' discovery requirements.	2011
<u>AGRIS</u> : The AGRIS (International Information System for the Agricultural Sciences and Technology) <i>linked data</i> strategy focuses on exploiting the semantic richness of its data by creating an open dataset in agricultural sciences.	2011
<u>Community information service</u> : A <i>linked data</i> approach to community information services could provide the data with an open license and using open standards which allows its re-use and lowers the barrier to participation in information curation and sharing.	2011
<u>BnF</u> : <i>linked data</i> technologies could help Bibliothèque nationale de France (BnF) to bring together data from several sources with a scalable and interoperable data model and to improve the publication of resources in the online catalog.	2011
<u>Identification and deduplication of library records</u> : The application of <i>linked data</i> to library records could help to develop automated matching algorithms for library records so that only one record exists for each intellectual item.	2011
<u>Linked data and legacy library applications</u> : The addition of <i>linked data</i> applications to the library information systems creates a challenge for system architects on how to adapt legacy systems to make use of new <i>linked data</i> applications.	2011
<u>Open Library data</u> : The Open Library is a large bibliographic database. <i>Linked data</i> technologies can be used to easily reference specific manifestations in the Open Library database.	2011
<u>Regional catalog</u> :The publication of <i>linked data</i> from German regional library services could help to create a German central catalog more easily.	2011
<u>Pode</u> : Pode use case concentrates on converting library data to RDF, converting it to FRBRized library data, and linking data	2011

to individual instances in other Linked Open Data datasets.	
Polymath virtual library : The use of <i>linked data</i> will benefit the Polymath Virtual Library in improving the process of obtaining enllaços from different sources and spreading the type of these sources.	2011
Talis Prism 3 : Talis Prism 3 is a next-generation OPAC/search and discovery interface. Prism 3 is powered by the Talis Platform, a hosted <i>linked data</i> service which offers both SPARQL querying and powerful full text search capabilities.	2011

Taula 19. Projectes Desenvolupats sobre *linked data* i biblioteques

A més, la *Library of Congress* va experimentar al 2010 amb diferents versions de la LCSH en múltiples idiomes, com per exemple el *link* entre LCSH i les llistes d'encapçalament RAMEAU de la Biblioteca Nacional de França, descargables ja al 2012 per a qualsevol biblioteca usuària. Altres iniciatives similars han estat dutes a terme amb la Universitat de Laval (Canadà) per establir enllaços entre la versió francesa i canadenca de la LCSH, amb les biblioteques nacionals de Xile i Espanya per enllaçar les versions respectives de la LCSH. I també van començar a establir contacte amb el projecte de la Biblioteca Alexandrina per treballar amb l'àrab. No hi ha publicats encara els resultats d'aquests experiments.

Altres iniciatives existents de l'aplicació de formats semàntics a les biblioteques és la conversió o mapeig del format MARC21 a RDF per a dades bibliogràfiques, tot i la flexibilitat del format (permet mapeig amb ISBD, FRBR i RDA) s'ha demostrat que hi ha gran pèrdua d'informació pel que fa al mapeig semàntic, ja que la conversió dels grafs complexes que suposen les dades bibliogràfiques fa que es produeixi incoherència semàntica de les dades, és a dir, hi ha un baix nivell d'equivalència entre les classes i les propietats d'ambdós formats (Dunsire, et al., 2012). És per això que la majoria de projectes se centren en la conversió dels vocabularis controlats ja existents (LCSCHE) i/o aplicació d'ontologies com a vocabularis controlats. Un bon exemple de cas d'ús és el catàleg de la biblioteca nacional sueca, LIBRIS, on la conversió de MARC21 a RDF s'ha centrat en una equivalència de classes simple, s'han basant en les descripcions simples dels recursos relacionats i els recursos s'ha relacionat amb les entitats de "primera classe".

4.4. Discussió

Aquest estudi mostra com el web semàntic i en particular els vocabularis controlats com són el cas de les ontologies, estan esdevenint un valor afegit per a tot tipus de sistemes d'informació i, sobretot, la seva influència en la

millora dels processos de recuperació de la informació. Els resultats obtinguts mostren que la flexibilitat és la característica fonamental dels llenguatges ontològics basats en RDF (Méndez 1999, W3C References 2004-2009, Danskin, 2013, etc.) i, en conseqüència, la seva interoperabilitat i semàntica proporcionen una major eficàcia, eficiència i rellevància en la recuperació de la informació (major representació dels conceptes i les seves relacions).

En els sistemes d'informació del camp de les biblioteques i centres de documentació s'aposta per l'ús del format SKOS, ja que s'utilitza per a la creació de vocabularis controlats de tipus ontologies, que són la base per a la indexació de la informació/documentació d'un sistema i per a la seva posterior recuperació de la informació. Al mateix temps, s'utilitza aquest format per convertir sistemes d'organització del coneixement tradicionals (tesaurus, llistes d'encapçalaments de matèries, etc.), ja existents en sistemes d'informació i documentació, en ontologies, permetent que aquests siguin utilitzats en el web semàntic, per tant, és un model d'adaptació sostenible. De fet, SKOS segueix fidelment la normativa internacional per construir tesaurus (ISO 2788, la versió espanyola UNE 50-106-90 i UNE 50125 per tesaurus multilingües), de manera que és un llenguatge estàndard que permet adaptar els vocabularis controlats (procés d'skosificació o conversió d'un llenguatge controlat amb una ontologia elaborada amb llenguatge SKOS). Tot i aquesta flexibilitat i interoperabilitat entre diferents tipus de llenguatges, la interoperabilitat dels sistemes d'informació tradicionals amb el web semàntic depèn de molts altres factors, tal i com s'estudia en el capítol d'Interoperabilitat semàntica.

Cal destacar també que SKOS és un llenguatge flexible que pot combinar-se amb altres vocabularis de modelat com per exemple RDF/OWL. No obstant això, l'alta varietat de models de formats semàntics fa que sigui impossible captar tots els detalls d'aquests tipus de sistemes. Així doncs, encara existeixen algunes limitacions de l'ús de vocabularis combinats semàntics com és el cas de OWL-SKOS (per exemple, vinculació de classes OWL amb conceptes SKOS). Paral·lelament, trobem accessibles un major nombre de conjunts de metadades basades en el model RDF (ontologies de diferents dominis), i, tot i que hi ha directius sobre bones pràctiques d'ús i aplicació d'aquests estàndards, la proliferació de recursos i eines basats en el web semàntic pot arribar a esdevenir un caos, de manera que és necessari establir estratègies per a la publicació, mapeig, descoberta i avaluació d'ontologies per permetre la gestió i desenvolupament dels sistemes d'informació semàntics i/o de les pròpies ontologies d'una forma més eficient.

És important destacar, que els formats semàntics tenen també una adaptabilitat amb els formats i estàndards emprats en biblioteques i centres de documentació (Dublin Core/RDF, MARC/RDF, MADS/RDF, SKOS i la nova ISO de tesaurus 25964 de 13/12/2012). Conseqüentment, la cohesió dels sistemes tradicionals amb els futurs sistemes semàntics es preveuen amb una gran

adaptabilitat. Així mateix, es presenten com una gran oportunitat, ja que s'estan desenvolupant sistemes amb major flexibilitat i interoperabilitat, a més, els professionals de l'àmbit de la biblioteconomia i documentació tenen ja una llarga trajectòria amb l'ús de formats i estàndards en sistemes d'informació i documentació. Tanmateix, encara no hi ha un consens amb la conversió de registres bibliogràfics amb formats semàntics, per exemple, en el cas d'Europeana, han elaborat el seu propi model de metadades, l'EDM (Europeana Data MODEL), al igual que altres comunitats d'àmbits científics (FAO, etc.). De manera que amb les conclusions d'aquest estudi es corrobora la gran diversitat de sistemes i ús de diferents tipus de metadades produeixen un entorn molt complex que desemboca en l'elaboració de models propis per satisfer les necessitats de cada comunitat.

L'aparició de tecnologies del web semàntic estan marcant tendència entre la comunitat bibliotecària, arxivística, etc., tot i que les primeres iniciatives són bastant recents, i consegüentment la tecnologia encara està en fase de desenvolupament i/o en constant canvi, s'observen pocs resultats visibles en aplicacions reals (no prototipus), i la majoria d'aplicacions no estan totalment complertes, com és el cas del cercador semàntic d'Europeana que està en fase beta i no indexa totes les col·leccions existents a Europeana.

De fet, la comunitat bibliotecària que treballa més activament amb el camp del web semàntic i el *linked data*, Library *Linked Data* Incubator Group del W3C, és de recent creació (maig del 2010 i agost del 2011). El mateix passa amb el desenvolupament de projectes relacionats amb biblioteques i el web semàntic que apareixen majoritàriament al 2011 (vegeu taula 19). En aquest estudi realitzat sobre l'estat de l'art dels formats i estàndards en sistemes d'informació d'àmbit especialitzat, es constata doncs, que la comunitat bibliotecària té un treball pendent present i futur sobre l'estudi de la interoperabilitat i adaptabilitat dels diferents formats, elaboració de directius comunies i estàndards emprats en els sistemes d'informació actuals i futurs (semàntics). A més, de fomentar la col·laboració en la recerca interdisciplinària sobre els formats semàntics i l'estudi de les seves capacitats en els sistemes d'informació i documentació i en la recuperació de la informació en general (web semàntic).

En conclusió, els resultats obtinguts mostren uns sistemes d'informació futurs amb una major semàntica i interoperabilitat, i això es tradueix amb un major accés i recuperació de la informació de qualitat (major pertinença en resultats dels processos de recuperació de la informació). Però, com exposa l'W3C i concretament els grups de treball dels formats OWL i SKOS, els llenguatges estan en continu desenvolupament de les seves pròpies directrius, de manera que és necessari estudiar l'evolució de tots els models semàntics i la seva relació i interoperabilitat, sobretot quant al seu ús i aplicació en sistemes d'informació.

4.5. Conclusions

Els estàndards en sistemes d'informació digitals (distribuïts o no) serveixen principalment per proporcionar interoperabilitat entre la informació continguda en el sistema d'informació i els serveis d'informació que siguin integrats en una plataforma digital.

Es poden trobar formats i estàndards en tots els nivells d'interoperabilitat semàntica (estructures de dades, dades categòriques i dades factuais). Els estàndards generalment també es poden agrupar en quatre grups: estàndards d'identificació i localització (exemples de formats: DOI, URI, etc.), estàndards de llenguatge de marcat (HTML, XML, RDF, MARC, etc.), estàndards de protocols d'intercanvi d'informació (Z39.50, SRU, OAI-PMH, etc.), estàndards de metadades (Dublin Core, MODS, METS, etc.). Principalment s'estudien els estàndards del grup 4 i 2, perquè són les que fan referència al nivell de descripció i recuperació de la informació en els sistemes digitals, i la resta d'estàndards responen a la tecnologia pròpiament (protocols d'intercanvi d'informació, identificadors, etc.).

El context de les metadades és molt heterogeni i poden existir tants esquemes de metadades com àmbits d'aplicació. En l'àmbit de les ciències aquàtiques i marines, els estàndards i formats d'estructures de dades més utilitzats són els formats més habituals coneguts en l'àmbit de les biblioteques especialitzades, els formats MARC i Dublin Core, no obstant, s'identifiquen fins a uns 62 estàndards d'esquemes de metadades relacionats en l'àmbit de les ciències aquàtiques i marines. Cal destacar també que ambdós formats internacionals identificats com els referents (MARC i Dublin Core), normalment no es troben en el mateix sistema d'informació perquè són creats amb objectius diferents i cadascun inclouen col·leccions de diferent tipus (catàlegs, repositoris digitals, etc.). Els estàndards coneguts en el camp bibliotecari sempre inclouen la premissa de la interoperabilitat, però no estan preparats per als nous usos demandats en l'escenari actual del web semàntic, no obstant, les seves característiques com a estàndards els permeten ser adaptables als nous sistemes d'informació semàntics.

Actualment s'està produint un canvi en els entorns digitals, on es desenvolupen sistemes d'informació més interoperables i flexibles que permeten tant l'ús de diverses col·leccions com diferents formats de descripció i accés a la informació. L'escenari global que envolta la informació digital mostra un usuari que no només té accés i consum de la informació de forma universal a través d'Internet, sinó que també és gestor i creador de la informació. Conseqüentment, l'escenari del creixement de la informació és constant i desmesurat de manera que també augmenta la informació desestructurada i heterogènia.

En síntesi, la tendència actual i els nous sistemes d'informació van cap a la construcció d'una web més estructurada (web semàntic o web 3.0), amb nous estàndards i formats semàntics (RDF, OWL, SPARQL, SKOS, etc.) que permetin una recuperació de la informació de qualitat, aportant un major significat a la informació continguda en la xarxa i essent més comprensible per a les màquines. Aquesta millora de la recuperació de la informació digital és dona gràcies a l'aplicació d'ontologies, llenguatges o vocabularis controlats semàntics (representats amb formats com OWL i SKOS), que proporcionen una major riquesa al contingut objecte de ser indexat i/o a la representació dels conceptes que conformen una col·lecció (sistema o web) i una major vinculació amb termes relacionats que altres vocabularis controlats emprats en el camp de la biblioteconomia i documentació (tesaurus, llistes d'encapçalaments, etc.)

En el camp de les biblioteques també canvia el concepte de col·lecció com a un conjunt d'objectes, però no necessàriament físics, ja que una col·lecció pot ser un conjunt de metadades que a punta a recursos d'informació distribuïts de diverses col·leccions seleccionades en la xarxa per la rellevància que ha definit la pròpia biblioteca digital (per exemple, similitud temàtica o cooperació bibliotecària), o a través de la consulta/interrogació d'un usuari a un sistema d'informació *online*.

Es conclou, que per tal d'assegurar la interoperabilitat dels sistemes d'informació digital és primordial utilitzar metadades elaborades conforme a estàndards d'ús comú en el seu àmbit d'aplicació, que siguin flexibles i interoperables. I per tant, fer possible que recursos descrits mitjançant diferents esquemes puguin ser recuperats independentment de si pertanyen a diferents sistemes d'informació i/o responen a informació distribuïda en el World Wide Web. Recollint tot el que s'ha dit, ressaltar que és necessària la col·laboració interdisciplinària entre diferents professionals de les ciències de la informació i les tecnologies de la informació i s'ha de treballar cap a la interoperabilitat dels diferents sistemes d'informació existents en un mateix ecosistema.

CAPÍTOL 5. INTEROPERABILITAT SEMÀNTICA EN SISTEMES D'INFORMACIÓ DE QUALITAT

5.1. Introducció

La recerca actual sobre interoperabilitat es divideix en múltiples contextos amb un a objectiu comú, la integració, gestió i accessibilitat de la informació.

En els sistemes d'informació digital més complexos, com ara els sistemes distribuïts (*subject gateways, etc.*), la interoperabilitat semàntica és una de les parts més importants de la seva arquitectura de la informació, ja que com la seva pròpia definició i context indiquen, són sistemes que s'envolten de múltiples organitzacions, i, per tant, de diferents sistemes d'informació que s'han de poder comunicar entre ells. A més d'incloure àmbits multidisciplinaris diferents i molts cops multilingües. Conseqüentment, en un escenari múltiple com aquest, la major problemàtica i objectiu principal és fer front a la heterogeneïtat de la informació (sistemes, recursos, llengües, vocabularis, etc.) que es produïda en aquests tipus d'entorns digitals i poder compartir, reutilitzar i recombinar la informació i serveis per poder proporcionar informació rellevant i de qualitat a l'usuari final. Principalment es poden identificar quatre tipus de heterogeneïtat que corresponen a potencials tipus de interoperabilitat:

1. A nivell de **sistema** o incompatibilitats de software
2. A nivell **sintàctic**, és a dir, les diferències entre codificacions i representacions a nivell informàtic (algorismes, metadades, etc.)
3. A nivell **estructural** o diferències entre models de dades, estructures de dades i esquemes (algorismes, metadades i mapeig principalment)
4. **Semànticament** trobem diferències d'inconsistència en la terminologia i el significat.

Tot i que en qualsevol sistema d'informació digital s'ha d'assolir la interoperabilitat en tots els nivells, també és veritat que des del camp de la biblioteconomia i la documentació la interoperabilitat és estudiada majoritàriament des del tercer i quart nivell d'interoperabilitat. Sobretot, el nivell interoperabilitat semàntica (SI) és el que s'ha considerat sempre com a punt clau d'estudi en el món de les biblioteques digitals, i és el que es tindrà en compte al llarg de tot aquest estudi.

En el context tradicional, la biblioteconomia i la documentació tracta la interoperabilitat semàntica des de la indexació i l'accés a les matèries per donar suport a la cerca, fent que els sistemes siguin més precisos alhora de la recuperació de la informació (IR). Un altre exemple, són les bases de dades bibliogràfiques en línia, que han aplicat funcionalitats de cerca creuada, per tal d'accedir a les matèries que indexen les bases de dades que són multidisciplinàries i multilingües.

Un altre context més recent és la necessitat d'integrar a l'heterogeneïtat sobretot dels recursos d'informació a Internet que s'incrementen desmesuradament en el web Això també implica oferir una arquitectura del web basada en serveis, el resultat del qual és un increment de múltiples sistemes d'informació digitals: biblioteques digitals, *subject gateways*, *data warehouses*, intercanvi electrònic de dades (metadades), arquitectures *peer to peer*, diferents sistemes de gestió del coneixement (web semàntic, ontologies, etc.), serveis e-Government, etc. A més a més de l'accés a la informació, la *Navegació* esdevé el punt més important de tots aquests sistemes d'informació digitals, sobretot la navegació per matèries.

L'altre escenari que es troba en la interoperabilitat semàntica són les activitats basades en la gestió del 'cicle vital de la informació', dintre la disciplina de la gestió del coneixement. Aquí podem trobar sobretot institucions governamentals (serveis e-Government), museus, empreses, etc.

També en la disciplina de la biblioteconomia i la documentació es contempla el concepte del 'cicle de vida de la informació' des de la perspectiva dels processos de gestió i arxiu de la informació en un entorn digital (*subject gateways* o biblioteques digitals): la creació del producte d'informació, l'*adquisició* (desenvolupament de la col·lecció), catalogació, indexació, organització i representació del coneixement (metadades, control d'autoritats, vocabularis controlats, identificació i nomenclatures, etc.), la integració (interoperabilitat semàntica i sintàctica, enllaços, etc.), mediació (les interfícies d'usuaris, personalització, serveis de referència, recomanacions, transferència de la informació, etc.), l'accés (difusió), ús d'aplicacions i recursos compartits (terminologia), manteniment, preservació i arxiu (emmagatzematge).

Així doncs, pel que fa a les *subject gateways*, la interoperabilitat semàntica, s'ha d'assegurar en tots els passos del 'cicle vital de la informació', en l'ús de vocabularis, les relacions semàntiques i en els processos de mapeig de la informació.

- En la creació i/o modificació de la informació, els recursos terminològics i la creació de metadades pels documents o objectes digitals són essencials per afavorir la SI. Les mesures de SI aportaran més control a nivell terminològic entre les diferents disciplines i comunitats científiques. Els creadors o autors contribueixen també al desenvolupament terminològic així com ajudar en el desenvolupament i actualització dels vocabularis.
- La publicació, els productors d'informació (biblioteques, editorials i altres institucions rellevants o intermediàries), poden ajudar també en altres processos del cicle de vida de la informació com és la *Creació i modificació* de la informació, la *Catalogació*, *Mediació* (interfícies

d'usuaris, personalització, referències, recomanacions, transferència d'informació, etc.), *Manteniment* de la informació, i l'*Arxivament* i la *Preservació* de la informació. Aquesta fase no és tan rellevant per a la SI, però, aporta informació temàtica i semàntica, i en molts casos des de llenguatges controlats.

- L'*Adquisició* i desenvolupament de la col·lecció, des de la perspectiva de la SI pot utilitzar els vocabularis controlats per determinar la inclusió o no d'un document o objecte digitals. Les activitats d'interoperabilitat permetran incloure o no documents en un sistema dependent de la equivalència terminològica que pugui trobar-hi.
- *Mediació*, determinarà la qualitat d'un sistema d'informació, ja que basant-se en vocabularis d'interoperabilitat semàntica s'oferirà a l'usuari un major suport a la navegació i accés a la informació.
- Es recursos terminològics i les mesures de interoperabilitat semàntica milloraran l'ús i avaluació de la informació.
- El procés de *Manteniment* en un entorn digital, se centra bàsicament en l'actualització i accessibilitat d'enllaços, així com la seva relació en la informació semàntica i els vocabularis controlats que incorpori, els quals s'han de poder adaptar als canvis per tal d'assegurar la seva perdurabilitat.
- En la *Preservació* de la informació també es té en compte la informació semàntica: vocabularis, mètodes i eines, les dades incloses en els documents (metadades), i tecnologies que puguin estar associades i siguin també perdurables.

Quant a la interoperabilitat semàntica, les activitats més rellevants dintre del 'cicle de vida de la informació' són: la catalogació (indexació, organització i representació del coneixement, etc.), la integració de la informació i l'accés a la informació (cerca i descoberta de recursos).

5.2. Antecedents de la interoperabilitat semàntica: problemàtiques i solucions

Els problemes d'interoperabilitat semàntica fan referència principalment a l'heterogeneïtat dels recursos d'informació ja que la majoria d'entorns digitals ofereixen un entorn de cerca creuada que integra dades heterogènies.

Tècnicament, la SI està relacionada amb tasques complexes que envolten els múltiples nivells i funcions dels sistemes d'informació. No obstant això, els nivells d'interoperabilitat semàntica resumeixen amb l'estudi de tres tipus de dades (Doerr, 2004):

- Dades estructurals o estructures de dades (referent a metadades i descripció de dades)
- Dades categòriques (referent a vocabularis controlats i estàndards), i
- Dades factuais (per exemple: autoritats i noms geogràfics)

Avui dia la majoria de solucions per poder fer front a aquestes problemàtiques passen per la línia de recerca del web semàntic. És a dir, s'intenta donar un major significat al web a la web, i, per tant, una major semàntica, per tal de solucionar problemes relacionats amb la cerca i recuperació de la informació a través d'una infraestructura comuna, amb la qual es poden compartir, processar i transferir informació fàcilment (W3C). Bàsicament, això és possible mitjançant la utilització de llenguatges estàndards (*RDF, OWL, SPARQL*, etc.), que fan possible compartir i reutilitzar dades i documents entre diferents tipus de sistemes i usuaris.

Els diferents estudis analitzats sobre biblioteques digitals especialitzades, mostren clarament dues línies de treball per assegurar la interoperabilitat semàntica entre els sistemes: la estandardització (metadades i formats ontològics, dades estructurals), i la integració (*crosswalks* i tècniques de mapeig). Algunes de les iniciatives que s'han dut a terme són:

- Ús de llenguatges controlats (tesaurus, ontologies, llistes controlades de paraules clau, etc.) que fan un control de sinonímia. A més, suporten la cerca creuada entre bases de dades, són capaços d'agregar contingut a les dades i/o construir amb eines de software per presentar la informació a l'usuari final (un portal, un repositori *OAI*, etc.).
- Per tal de millorar la precisió a la cerca, hi ha projectes que estableixen relacions semàntiques entre els termes claus definits, per exemple en un tesaurus jeràrquic, una ontologia formal, taxonomies, etc. Projectes relacionats amb portals temàtics que suporten la cerca creuada. Com ara les *subject gateways* com *INTUTE*, *CARMEN* o el portal *Vascoda*.
- Mapeig o traducció de termes és necessari també per tal de millorar la cerca i que hi puguin haver documents multilingües.
- Quant a la navegació, la interoperabilitat semàntica ha de poder garantir la navegació estructurada per col·leccions i subcol·leccions, de manera que és necessari incorporar estructures de categories com els sistemes de classificació.
- Un altre exemple es dona en diferents col·leccions de documents que es relacionen amb períodes diferents o tenen diferents perspectives culturals o polítiques, i la interoperabilitat semàntica ha de poder mesurar els diferents contextos que s'incloguin en una col·lecció. Per tant, la SI haurà de contemplar un o varis vocabularis apropiats a les col·leccions que integri un sistema, és a dir, s'haurà de produir un procés de mapeig perquè existiran diferents vocabularis especialitzats i diferents grups d'usuaris.
- Si ens trobem casos de documents ben estructurats i agrupats en un mateix sistema però que provenen de diferents llocs, les operacions de interoperabilitat semàntica s'han d'aplicar de dues formes: decidir la equivalència de la definició semàntica de les dades, dels elements de

metadades, dels camps bibliogràfics, de les propietats, atributs, etiquetes o similars que s'hauran de mapejar o bé partir d'estàndards.

- En els casos on els documents de o les seves metadades provinguin de diferents Sistemes d'Organització del Coneixement (KOS), o de diferents esquemes o vocabularis de codificació (encapçalaments de matèries i sistemes de classificació), podrien ser també mapejats (o enllaçats via *crosswalk*) per tal de millorar la seva SI.

5.3. Models d'interoperabilitat semàntica

Com s'anunciava anteriorment, tots aquests nivells d'interoperabilitat fan referència a les problemàtiques referents a la semàntica i funcions diverses existents dintre els sistemes d'informació. Tot i que en aquest estudi es centra en la interoperabilitat lligada amb les metadades i els sistemes d'organització del coneixement (KOS), en qualsevol sistema d'informació digital de qualitat s'han de tenir en compte tots els nivells d'interoperabilitat semàntica existents (Doerr, 2004):

1. *Estructura de dades*. Descriuen possibles estats de la qüestió i suporten informació de control i funcions de gestió que recauen en el sistema no en l'objecte, on hi ha els fitxers, les taules i els registres de dades. És on es troben les *metadades*, el contingut de les dades, les dades de gestió de la col·lecció, els serveis de descripció de dades, etc. Des d'un punt de vista ontològic, l'estructura de les dades pot estar lligada a un domini universal però no a particulars. Codifiquen les relacions més rellevants en un domini perquè es relaciona directament amb les funcions vitals per construir el sistema d'informació, és a dir, proporciona la informació abstracta del concepte individual com el recurs, l'agent (creador), data, tipologia, etc. Per tal de garantir la interoperabilitat semàntica entre sistemes, és necessari el mapeig de les estructures de dades, i per tant, els estàndards de metadades han de ser esquemes globals, orientats a objectes, flexibles i interpretables. A més a més, els algorismes de mapeig depenen de les dades categòriques i, aquestes han d'estar estandarditzades en el KOS per tal de fer les equivalències entre les dades.
2. *Dades categòriques*. Són dades *universals*, com són les *classificacions*, terminologies o estàndards. L'estandardització comença amb els *vocabularis controlats*. Mentre que el nivell d'estandardització és molt alt, a nivell global és pobre comparant-ho amb les estructures de les dades. La terminologia no va lligada només a dades sinó també representen comunitats (cultures, ideologies, etc.). Així que el KOS sobre dades categòriques no poden estandarditzar-se sempre. Un mateix recurs, malauradament, no sempre té representat el mateix nivell de classificació, i una equivalència exacta entre termes no és sempre possible.

3. *Dades factuais o particulars.* Part de les dades factuais es refereixen al fet que només poden aparèixer un cop dintre d'un entorn digital, com per exemple la relació d'un autor específic, un lloc o una data de publicació. Tot i que el mateix autor, lloc o data poden tornar a aparèixer ha de ser possible poder-ho diferenciar. Al contrari que les dades categòriques, poden diferenciar-se, exceptuant els noms geogràfics, per regles de codificació de dades (identificar ítems que es repeteixen) i amb la descripció de particulars, és a dir, per mitjà un estàndard artificial identificador que s'ha d'assignar a cada objecte (per exemple estàndards *d'autoritats*). Per dur-ho a terme s'han de tenir en compte diversos algorismes que detectin duplicats, sistemes de '*data cleaning*', tècniques de '*data mining*' que serviran d'instrument d'indexació i suport per al mapeig de dades.

Resumint, la interoperabilitat semàntica d'un sistema està formada per l'esquema o estructura de dades d'un sistema (metadades, fitxers de dades, registres, taules, etc.), les dades categòriques o universals (classificacions i vocabularis controlats) i dades factuais o particulars (controls d'autoritats).

El primer pas doncs per assegurar la interoperabilitat del sistema és el mapeig de la pròpia estructura de dades que hi ha dintre dels diferents sistemes d'informació implicats. És a dir, s'han de fer processos associatius entre uns tipus d'elements i uns altres i buscar la relació entre els esquemes que integren el sistema. En aquest mapeig de dades o metadades la tècnica més estesa és el *crosswalk* (MMI, 2007), o mapeig semàntic i tècnic de l'estructura de metadades a altres estructures de metadades, per exemple, fer la equivalència dels elements i atributs de les metadades:

Marc Fields	Dublin Core Elements	NOAA FGDC
245\$a (Title) 245.10 \$a Aquatic plant book	<DC:title>Aquatic plant book</>	1.1.8.4 (Title) Aquatic plant book
100,110,111,710,711 (Author) Ex.: 100.10 Cook, Christopher D.K.	<DC:creator>Cook, Christopher D.K. </>	1.1.8.1 (Originator) Cook, Christopher D.K.
260\$a (Publication Place) Amsterdam	<DC:publisher>Amsterdam</>	1.1.8.8.1 (Publication Place) Amsterdam
260\$b (Publisher) SPB Academic Publishing \$c 1996	<DC:publisher>SPB Academic Publishing	1.1.8.8.2 Publisher SPB Academic

		<i>Publishing</i>
260\$c (Date) 1996	<DC:data> 1996</>	1.1.8.2 (Publication date) 1996
650\$a (Subject); 650 2/653 (Subject); Ex: 650.04 <i>Freshwater plants Identification</i> \$x	<DC:subject> <i>Freshwater Plants, Identification</i> </>	1.6.1.1 (Theme Keyword) <i>Freshwater plants – Identification;</i> 1.6.1.2 (T.K. thesaurus/term uncontrolled)

Taula 20. Exemple de Crosswalk: mapeig entre diferents formats de metadades (Font: Elaboració pròpia)

El mapeig semàntic pot ser automatitzat a través d'algorismes que transformaran una estructura de dades A en una estructura de dades B i a l'inrevés, sempre preservant el significat original de la informació obtinguda.

La interoperabilitat semàntica, a més de la integració i interoperabilitat de les estructures de metadades necessita també una integració del coneixement i la terminologia de les àrees temàtiques que conformen el sistema d'informació. Així mateix, les metadades no sempre inclouen suficient informació sobre matèries, i molts cops són els propis recursos els que no disposen de suficient informació, d'aquesta manera caldrà tenir en compte les dades categòriques i factuais. Els darrers anys la navegació per matèries a Internet ha estat cada cop més rellevant i sobretot en entorns com les *Subject Gateways* o les biblioteques digitals, on els catàlegs cada cop són més especialitzats i inclouen matèries més interdisciplinàries que poden indexar i classificar amb menys profunditat. La dificultat de donar accés als diferents nivells de granularitat que trobem en entorns multidisciplinaris, el manteniment dels propis llenguatges i els canvis de l'ús dels sistemes d'indexació de matèries dibuixen un escenari heterogeni que cal afrontar de diverses formes des del punt de vista de la interoperabilitat (Koch, T. 2006 i Chan, L.M. 2002):

1. Ús de sistemes de classificacions comuns, reclassificació/reindexació de tots els documents (manualment o automàticament): la indexació per paraules clau, usant llistes de vocabularis controlats com *Subject Headings* o *thesaurus* poden servir de sistemes interoperables d'accés a les matèries. Però, crear un accés integrat, amb un mapeig controlat o incontrolat de *keywords* cap a un altre sistema de *keywords*, és metodològicament molt difícil i molt costós, i la majoria de sistemes que opten pel mapeig utilitzen les classificacions com a llenguatge de mapeig numèric és més senzill que el terminològic o semàntic.

2. Derivació/Modelització. Un vocabulari controlat especialitzat que es desenvolupa amb un d'existent més comprensiu o adequat.
3. Mapeig (procés intel·lectual): un sistema mapejat consisteix bàsicament en l'establiment d'equivalències entre termes de diferents vocabularis controlats o entre termes verbals i números de classificació, i evidentment, seria en el, o entre el mapeig de les seves estructures i relacions jeràrquiques i associatives. El resultat és un sistema comú i adequat per fer front a la heterogeneïtat d'un sistema multidisciplinari i multilingüe. Però s'ha de tenir en compte que el mapeig de sistemes amb classificacions comporta pèrdua de precisió en la recuperació de la informació, ja que els mapejos són fets de forma bilateral i en una única direcció, des d'un sistema comú a cadascun dels sistemes participants. D'altra banda, les classificacions en les *Subject Gateways* (Renardus, etc.) i molts altres entorns digitals s'utilitzen com a estructures comunes de navegació i accés multilingüe als recursos d'informació. L'aplicació d'aquest tipus de sistemes és llarga, costosa i al darrera sempre hi ha esforços intel·lectuals ja que l'automatització dels processos no ha donat bons resultats.
4. Traducció/Adaptació. Un vocabulari controlat és desenvolupat a través de termes traduïts d'un altra llengua diferent amb o sense modificació.
5. Suport en la indexació i la classificació en els diferents sistemes per tal de proveir una superposició i interoperabilitat parcial. Si no es vol optar pel mapeig hi ha altres passos a seguir que impliquen tenir un protocol OAI:
 - Enriquiment de metadades, autoritats, índexs o classificacions, etc.
 - Incloure un procés de classificació semiautomàtic per l'accés a les matèries.
 - Ús de sistemes d'indexació comuns (per disciplines).
6. No fer res. Dependrà de la indexació del text complet, de la indexació de cites des de sistemes d'accés online (bases de dades bibliogràfiques, Google, xarxes socials, etc.).

5.4. Tendències: estandardització i interpretació

Els diferents estudis realitzats entorn les biblioteques digitals especialitzades mostren clarament dues línies de treball per garantir la interoperabilitat semàntica, l'estandardització i la interpretació (Patel,M.;Koch,T; Doerr,M. 2004).

L'estandardització és proactiva, tothom pot compartir i accedir a les dades amb l'ús d'un estàndard comú. En aquest sentit, l'estandardització es refereix a la gnificat de les metadades, el contingut del seu esquema, compartir els mateixos conceptes de KOS, ús de controls d'autoritats, noms geogràfics i

identificadors comuns, etc. L'estandardització pot donar estabilitat a llarg termini però alhora es pot convertir en un sistema poc flexible i és per això que només seria adequada en aplicacions que no fossin molt diverses.

La interpretació és més reactiva i es basa en la traducció, mapeig o correlació de metadades, esquemes de continguts (*crosswalk*), i de llenguatges o vocabularis controlats (*subject gateways*) com a eines suficients per interpretar la informació de diversos entorns. La interpretació és un sistema més flexible i selectiu podria aplicar-se en ambients amb alt nivell de diversitat però, a la vegada és poc estable al llarg del temps perquè la probabilitat de canvi i actualització és més elevada.

	ESTANDARITZACIÓ	INTERPRETACIÓ
AVANTATGES	<ul style="list-style-type: none"> - Immediatesa en la comunicació de la informació (integrada, unida, etc.) i sense transformació, ni alteració - Informació potser guardada de forma singular - Informació dels recursos seleccionats potser funcionalment completa per proveir un servei integrat 	<ul style="list-style-type: none"> - Els recursos d'informació, com per exemple dades legals, no necessiten adaptació - Els recursos poden servir com a funcions locals addicionals - Només necessita interpretació les parts rellevants de l'aplicació (indexació i navegació creuades) - La <i>interpretació</i> potser optimitzada per múltiples funcions - Els intèrprets s'adapten als canvis fàcilment
DESAVANTATGES	<ul style="list-style-type: none"> • L'esforç de produir un estàndard com terminologies potser molt alt • Els recursos d'informació s'han d'adaptar als estàndards • L'estàndard ha de preveure els usos futurs. Introduir elements nous que s'adapten al temps i poden causar problemes de comptabilitat • Un estàndard és per al seu domini. No potser òptim per totes les aplicacions. La selecció a vegades és una decisió política 	<ul style="list-style-type: none"> • <i>Interpretació</i> necessita temps per processar durant la comunicació • Alt esforç manual per produir la base del coneixement per a l'intèrpret, per exemple, creació de taules de correlació per terminologies • Pèrdua d'informació ja que es necessiten sistemes d'informació molt precisos • Mecanització de processos encara no està resolta

Taula 21. Models d'estandardització i interpretació (Font: Elaboració pròpia)

Les tendències en la interoperabilitat semàntica dels sistemes d'informació digitals apunten cap a l'estandardització com a model més eficaç per assegurar la seva heterogeneïtat. Però trobem que a nivell de biblioteques digitals i *subject gateways*, és fonamental utilitzar tècniques d'interpretació (*crosswalk* i *mapeig de KOS*) quant als llenguatges documentals es refereix. La rellevància i la precisió de la recuperació de la informació són tan vitals com l'accés a la informació, i, per tant, és necessari l'ús, intercanvi i consistència terminològica per assegurar una plena interoperabilitat semàntica.

D'altra banda, cal diferenciar entre diferents tipus de sistemes d'informació distribuïts com ara per exemple les biblioteques digitals i les *subject gateways*. En aquest cas, tot i que a nivell d'interoperabilitat de sistema segueixen el mateix model, quant a la seva interoperabilitat semàntica són diferents, ja que utilitzen diferents sistemes d'organització del coneixement (KOS). Les biblioteques digitals tendeixen més cap a l'ús i desenvolupament de metadades i ontologies⁴⁵ (*web semàntica*) i les *subject gateways*, usen metadades i altres vocabularis controlats com a *thesaurus*, sistemes de classificacions i *subject headings*.

En ambients digitals ecològics (*subject gateways*), predominen les tècniques interpretatives, centrades principalment amb el mapeig de KOS i tècniques de *crosswalk*. La rellevància i la precisió en la recuperació de la informació són tan importants com l'accés a la informació, l'ús, l'intercanvi i una terminologia consistent per tal d'obtenir una interoperabilitat semàntica plena.

Probablement seria necessària la combinació de les dues aproximacions per assegurar la interoperabilitat semàntica, d'un sistema d'informació complex, multidisciplinari i multilingüe com el que envolta una *subject gateway*. De fet, una *subject gateway* pot contemplar sistemes estàndards per a la descripció de recursos, accés i/o navegació (Metadades, AACR2, classificacions,) i sistemes d'interpretació per a la indexació, accés i/o navegació (mapeig de vocabularis controlats, principalment *thesaurus*).

5.5. Estudis de casos: subject gateways (models distribuïts) i linked data (models ontològics)

“La interoperabilitat és la capacitat de comunicar, executar programes, o transferir dades entre diferents unitats funcionals de tal forma que a l'usuari final se li requereixi poc o cap coneixement de les característiques úniques d'aquestes unitats”. (*ISO/IEC 2382 Information Technology Vocabulary*).

⁴⁵ Ontologies: són esquemes conceptuals exhaustius dintre d'un domini, contenen una estructura de dades típicament jeràrquica amb entitats rellevants i les seves relacions i regles dintre d'un domini específic. No obstant també estableixen relacions associatives dintre d'una mateixa ontologia o amb d'altres

Els diversos factors que afecten al funcionament de la interoperabilitat en aplicacions de sistemes de recuperació de la informació en xarxa que cal tenir en compte són els següents (Moen, 2001b):

- Múltiples i diversos sistemes de maneig i recuperació de la informació,
- Múltiples protocols,
- Múltiples esquemes de metadades,
- Múltiples formats de dades,
- Múltiples idiomes i tipus de caràcters,
- Múltiples vocabularis, ontologies i disciplines.

A continuació s'exposaran diferents casos d'estudi rellevants que permetran observar i analitzar totes les característiques mencionades anteriorment

5.5.1. Les subject gateways com a model d'interoperabilitat semàntica en sistemes d'informació distribuïts de qualitat

En aquesta secció s'estudia com els factors esmentats anteriorment representen la línia de base de les passarel·les temàtiques que es tracten i com l'arquitectura de la informació d'aquestes és un model per a la interoperabilitat semàntica.

Actualment les *subject gateways* són conegudes com a webs basades en mecanismes avançats per poder accedir a una col·lecció d'alta qualitat. Són infraestructures d'informació acadèmica i/o de recerca que permeten que permeten a varies institucions d'àmbit nacional o internacional unir-se per crear un espai de col·laboració entre comunitats científiques que tenen unes mateixes necessitats d'informació i documentació. Fet que les caracteritza com a sistemes d'informació d'alta qualitat, gràcies a la implementació de les premisses de *Cooperació*, *Coordinació* i *Sostenibilitat* (evitar duplicitat d'esforços, etc.):


- Forta **cooperació** i **coordinació** entre diferents països sota un mateix objectiu. Donar cobertura temàtica i en la majoria dels casos també geogràfica a una comunitat d'usuaris, facilitant així la millora dels serveis oferts per les passarel·les.
- Recuperació de la informació de forma sostenible. Donar accés lliure a la informació que estigui amb uns criteris de selecció comuns per a que tant la recuperació com el manteniment dels sistemes sigui sostenible.
- Duplicació d'esforços. S'estableixen mecanismes de cooperació i coordinació per tal d'establir uns criteris comuns que evitin la duplicació d'esforços entre les diferents passarel·les integrants d'una xarxa, sobretot quant al desenvolupament de criteris comuns per a la selecció de recursos (Rodríguez-Yunta, 2004)

- Normalització de sistemes. Llenguatges documentals, codis específics (país, dates, etc.), esquemes de metadades, etc. Facilitant així la interoperabilitat entre els sistemes.
- Dissenyar uns principis d'accés a la informació així com la compatibilitat en el disseny d'estructura dels camps utilitzats.
- Determinar condicions sobre drets de propietat intel·lectual.
- Avaluar i aconsellar en la provisió d'aspectes tècnics com l'arquitectura dels sistemes.

5.5.1.1. Model d'arquitectura d'informació d'una Subject Gateway

Encara que l'arquitectura de la informació pot ser estudiada a través de tres perspectives (l'usuari, el contingut i el context), només el contingut s'estudia quan volem veure el nivell d'interoperabilitat en qualsevol sistema d'informació. Per tant, els principals factors per a l'estudi són l'autor (reputació i control de la informació), el format (tipus de contingut), estructura (documents i integració de la informació), les metadades (ús de normes de descripció i el nivell de qualitat i consistència), *update* (actualització o dinamisme del contingut) i la integritat (volum d'informació).

No existeix una definició única del que és l'arquitectura d'informació, però es podria entendre com una disciplina que consisteix en una combinació de sistemes d'organització del coneixement per a l'accés i recuperació d'informació en un entorn digital, que dóna suport a la usabilitat i recuperabilitat, definició proposada per *Morville i Rosendfeld* (2006).

Semantic Interoperability (Doerr, 2004)	Semantic Interoperability (Moen, 2001b)	 In conjunction	Information Architecture (circles of information ecology; Morville&Rosefeld, 2006)
Data structures model	Data formats	Types of content	Content: formats
	Metadata schemes	Label systems (metadata) and use of description standards	Content: description and structure
	Controlled vocabularies	Organization systems (classification)	Content: organization structures

Categorical data model		schemes)	
	Languages, controlled vocabularies, character sets, ontologies, etc.	Controlled vocabularies (thesaurus, SH, etc.), ontologies	Content: description schemes or metadata
Factual model (only for authority records)		Authorities	Authorities (ownership)
<i>[This is not implicit in the Doerr model]</i>	Information retrieval systems and protocols	Browsing systems (cross-browsing)	Context: (system) and user perspectives (information access)
<i>[This is not implicit in the Doerr model]</i>	Information retrieval systems and protocols	Search systems (cross-searching and cross-indexing)	Context: (system) and user perspectives (information access)

Taula 22. Relació entre la interoperabilitat semàntica i l'arquitectura de la informació (Font: Sebastià-Salat, M.; Reverté, C., 2011)

Cal assenyalar que l'estudi de la interoperabilitat, especialment a nivell semàntic, té dos factors claus correlacionats que són difícils de separar quan s'estan estudiant casos reals: el contingut, que es troba dintre d'un sistema d'informació, i el context, centrat amb la recuperació d'informació del sistema i la seva arquitectura de la informació.

L'exhaustivitat (cobertura) i especificitat (precisió) del contingut està relacionada amb el context (l'ús de vocabularis controlats, indexació i eines de navegació). Com més cobertura, d'un tema d'interès, hi ha en un sistema d'informació, major és la recuperació d'informació/documentos rellevants per part de l'usuari, i com més gran sigui l'especificitat d'un terme, és menys probable recuperar els resultats no rellevants que poden haver en un sistema.

5.5.1.2. Requisits generals d'interoperabilitat semàntica en una SG

Els mecanismes que fan possible la coexistència dels diferents sistemes d'informació que integren una passarel·la temàtica són la indexació creuada, la navegació i cerca creuada. A part, evidentment, de la utilització de la resta de mecanismes d'interoperabilitat, es poden trobar en altres tipus de sistemes d'informació, els protocols, el llenguatge d'interrogació i la sintaxis dels registres, els esquemes de metadades, els vocabularis controlats i les regles de catalogació entre d'altres estàndards.

A partir dels projectes més representatius europeus i internacionals (SG australianes, projecte RDN, Intute, Desire, Renardus, Vascoda, etc), s'identifiquen diferents requisits i normes relacionades amb la interoperabilitat semàntica (consultar projectes analitzats en Annex 5.1):

- A nivell de sistema

- Nivell de sistema (sistemes, programari i protocols): HTTP, SOAP, Z39-50, OKBC, JDBD, OAI-PMH (Open Archives Initiative Metadata Harvesting) and CIP (Common Indexing Protocol)
- Nivell sintàctic (sintaxis: es refereix a la capacitat de càrrega de dades de múltiples fonts en una eina comuna): XML, HTML, Zthes, DTD, SRW i SKOS-Core (esquema RDF)
- Nivell estructural (estructures, sistemes i models de dades): RDF, OWL (web Ontology Language), UML
- Semàntica (diferències de terminologia i significat per ser comprès tant per la gent com pels ordinadors): MARC, Dublin Core, IEEE LOM, CIDOC CRM, MPEG-7,

- A nivell d'interrogació i accés a la col·lecció: cerca creuada i navegació creuada:

- La *Cerca Creuada* permet que d'una sola interfície l'usuari pugui accedir a diferents col·leccions provinents de diferents sistemes d'informació que provenen d'institucions i països diversos. Aquest procés és possible principalment a l'aplicació del protocol d'intercanvi d'informació Z39.50, que fa possible la integració de recursos impresos (provinents d'un OPAC, per exemple) i dels recursos web de la xarxa.

D'altra banda, la navegació creuada complementa la *cerca creuada*, permetent l'accés al contingut dels diferents sistemes d'informació a través d'una estructura jeràrquica de matèries (navegació per matèries). Aquesta es produeix gràcies als vocabularis controlats que hi ha dintre dels sistemes d'informació, com per exemple les *subject headings* (INTUTE Project),

classificacions (Renardus Project) o tesaurus (*Australian Subject Gateways*). És a dir, hi ha un procés de mapeig de les estructures de navegació locals, identificant les relacions i la sintaxis (algorismes de recuperació de la informació i metadades). Però també s'han de tenir en compte la interacció de formats i/o esquemes de descripció (indexació creuada).

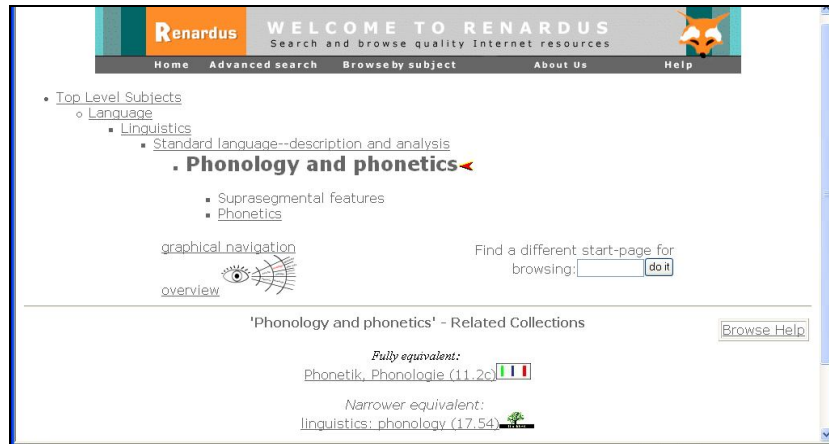


Figura 83. Exemple 1: Navegació creuada (Font: Renardus)

- **A nivell de descripció, interrogació i recuperació de la informació: *indexació creuada***

Quant a la *Indexació Creuada*, juntament amb estàndards de metadades (Dublin Core majoritàriament), s'utilitza per a la descripció de recursos i com a llenguatge de recuperació de la informació. Per tant, serà important avaluar els esquemes comuns de KOS que són integrats en cadascun dels sistemes, conèixer les eines i sistemes de catalogació, i les normatives de descripció utilitzades.

5.5.1.3. Arquitectura de la informació de les *subject gateways*

Un usuari (a través d'una interfície) s'ha d'adreçar a diferents col·leccions digitals utilitzant només un llenguatge d'indexació i recuperació de la informació, per mitjà del qual és *mapegen* altres llenguatges documentals, vocabularis especialitzats, i inclús altres versions dels llenguatges utilitzats. L'element bàsic de l'arquitectura de la informació d'un sistema d'aquestes característiques per aconseguir la interoperabilitat semàntica s'engloba dintre del sistema de recuperació de la informació del propi sistema (*cross-searching*, *cross-browsing* i *cross indexing*). Els 2 processos bàsics a tenir en compte són:

- La relació entre la interrogació feta per l'usuari i el document recuperat.
- El procés d'equivalència dels termes: identificació de termes, conceptes i relacions jeràrquiques entre els diferents llenguatges documentals existents en el mateix sistema d'informació.

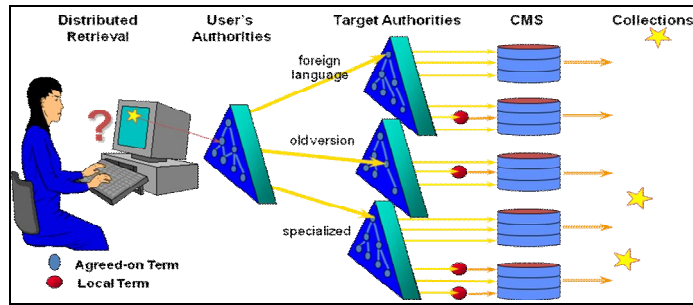



Figura 84. Interoperabilitat entre vocabularis controlats (procés de mapeig, Doerr, 2000)

Així doncs, l'ús dels vocabularis controlats són la clau del sistema de recuperació de la informació d'un sistema distribuït com són les *subject gateways*. Les principals problemàtiques associades a aquests tipus de sistemes sorgeixen quan s'intenta utilitzar un únic vocabulari controlat a través de múltiples dominis o sistemes, o bé, quan es vol fusionar múltiples vocabularis controlats. Una de les possibles solucions per fer-hi front és el mapeig de vocabularis.

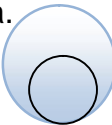
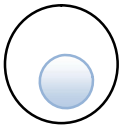
El resultat final del *mapeig* ha de ser la consistència de la indexació i recuperació de la informació provinent d'una o més col·leccions. Per a que això es produeixi hi ha d'haver una equivalència entre els diferents llenguatges documentals que integren la col·lecció. És a dir, equivalència tant dels termes com dels conceptes i les seves relacions.

1. *Equivalència Exacta: A i B són coincidents*




2. *Equivalència Parcial: A és parcialment equivalent a B amb els BT o amb els NT*

➤ En aquest casos són millors les equivalències parcials entre els BT que les equivalències parcials de NT, perquè hi haurà una major consistència.

Cas BT  Cas NT 

3. *Equivalència inexacta ("±"):* A i B tenen una major o menor coincidència

Major o Menor



Taula 23. Exemple de nivells d'equivalència entre termes d'un descriptor d'un Tesauro A a un descriptor del Tesauro B (Font: Elaboració pròpia)

És a dir, una equivalència **BT** de **A** hauria de ser mínima, per exemple amb un altre concepte de **B** que pot ser un **NT** o un **BT** d'un tercer tesauro (**C**).

O bé, una equivalència **NT** ha de tenir una coincidència màxima, és a dir, una combinació de **B** és **NT** d'un concepte **BT** de **A** o **C**.

Com més vocabularis s'hagin de mapejar més augmentaran les dificultats en el mapeig i en conseqüència, en la consistència dels llenguatges.

La situació més òptima seria que les col·leccions indexades utilitzessin protocols comuns de tesauro o que altres vocabularis controlats es poguessin integrar en la SG i al mateix temps desenvolupar-se i allotjar-se de forma separada en cadascun dels seus sistemes. Les especificacions de formats d'intercanvi per a tesauro més conegudes són: format d'autoritat MARC 21, XML basat en formats com Zthes DTD [Zthes] i RDF representacions com SWAD-Europe's o l'esquema SKOS-Core.

Problemàtiques del mapeig de vocabularis controlats

Encara que el mapeig de llenguatges documentals o vocabularis controlats sigui contemplat com a element clau per garantir la interoperabilitat semàntica d'una *subject gateway*, una Biblioteca Digital o un sistema d'informació distribuït i a més a més multidisciplinari i multilingüe, s'han d'afrontar diverses problemàtiques encara no resoltes:

1) Multilingüisme

- L'accés a la informació és multilingüe però la navegació i la cerca potser monolingüe. La problemàtica principal succeeix quan hi ha un llenguatge comú i el control del vocabulari emprat és un vocabulari traduït.
- Llenguatges creuats. Cercar i recuperar informació basada en traduccions, poca precisió, no hi ha sempre equivalència de termes ni tampoc en la cobertura temàtica.
 - Grau de consistència baix (diferents cobertures i nivells d'especificitat)
 - Problemes de polisèmia i sinonímia
 - Problemes per interrogar a múltiples col·leccions, la selecció, el filtratge, el rànquing dels documents recuperats, etc.

2) Heterogeneïtat

- Diversitat cultural. El fet d'integrar vocabularis de diferents llengües comporta també problemàtiques a tenir en compte des del punt de vista conceptual.

- Heterogeneïtat entre les estructures jeràrquiques dels vocabularis controlats:
 - Dificultats en les relacions d'equivalència entre els llenguatges: poca profunditat a nivell temàtic, grau de precisió i consistència baix, etc.
 - L'ús de traduccions (thesaurus multilingües) pot comportar pèrdua d'informació ja que les estructures conceptuals del coneixement són diferents també en la llengua (per exemple món occidental i oriental).
 - Diferències estructurals entre els llenguatges afecta també al mapeig de les estructures que tindran nivells d'especificitat diferents i també diferències semàntiques, de sintaxis, lèxic, diferents nomenclatures, etc.
 - Granularitat/Especificitat. Cada vocabulari pot tenir diferències amb l'especificitat dels termes utilitzats per indexar els continguts, de manera que els resultats obtinguts poden no ser òptims.
 - Diferents formes d'escriure sinònims i quasi-sinònims: Els termes disponibles poden no satisfer les necessitats dels usuaris, o bé, un usuari pot no localitzar els termes necessaris perquè hi manquin referències creuades de diverses classes incloses en el vocabulari.

3) Funcionalitat

- Conèixer les diferències entre metodologies de cerca per les bases de dades cercades en un mateix sistema d'informació: Fins ara, les biblioteques han abordat aquest problema mitjançant l'ús de Z39.50, protocol comú de cerca per cercar mitjançant múltiples catàlegs d'accés públic en línia (OPAC). No obstant això, l'ús d'un protocol comú és insuficient, la interpretació dels termes de cerca i els resultats de cerca que retornen els sistemes podrien ser completament diferents en funció de si una base de dades espera una cerca pre-coordinada o post-coordinada. Per exemple, un vocabulari controlat pot haver estat desenvolupat per ajudar a la navegació web, mentre que un altre vocabulari pot haver estat construït per proporcionar condicions per a consultes post-coordinades mitjançant operadors booleans.
- El propòsit de cadascuna de les bases de dades o sistemes implicats: Les finalitats o audiències poden ser diferents, per consegüent, els resultats poden no satisfer als diferents usuaris. Per exemple, els resultats obtinguts d'una base de dades generalista (catàleg d'una biblioteca pública) en contra posició d'una base de dades d'alt nivell científic/ tècnic, com potser *web of Science*.

- Desenvolupament del vocabulari (organització, literatura, etc.): cada vocabulari integrat en un sistema pot tenir termes comuns amb significats diferents i no disposar de qualificatius per diferenciar-los.

Per tant els factors claus a estudiar per fer front a aquestes problemàtiques d'interoperabilitat entre vocabularis controlats són:

1. La semblança entre 2 o més vocabularis: similaritat del contingut temàtic d'interès en els diferents dominis que s'agrupen en un mateix sistema.
2. Ús de diferents vocabularis controlats utilitzats per indexar un contingut des de dominis similars. Per exemple, *NAL Agricultural Thesaurus (USDA)* o *Medical Subject Headings (US National Library of Medicine)* i *LC Subject Headings (Library of Congress)*.
3. El grau d'especificitat o granularitat dels vocabularis controlats utilitzats per indexar diferents dominis d'un contingut o bases de dades.
4. Les diferències estructurals entre els vocabularis: per exemple, conèixer la forma emprada per descriure sinònims i quasi-sinònims en cadascun dels vocabularis controlats utilitzats.
5. Conèixer la finalitat i ús de cada vocabulari controlat d'interès.

No hi ha solucions absolutes, però, l'efectivitat i la consistència entre els llenguatges documentals integrats en un sistema multidisciplinari i multilingüe, com és el cas de les ciències aquàtiques, passa per utilitzar estàndards primerament, i a poder ser, sistemes d'indexació multilingües per tal que tinguin una cobertura lingüística europea. A més a més, és cabdal que tots els vocabularis controlats compleixin amb els factors d'interoperabilitat necessaris per a ser mapejats.

A nivell de cerca, existeixen 3 enfocaments sobre el mapeig de vocabularis que caldrà tenir en compte:

- Vocabularis controlats combinats. Vocabularis controlats individuals que s'integren per formar un vocabulari controlat més gran que és aplicable a tots els àmbits específics.
- Micro vocabularis. S'utilitza l'estructura d'un vocabulari controlat genèric com a guia i model per a la creació de múltiples vocabularis controlats i altament específics. Aquest és el cas més habitual emprat per les *subject gateways*, la majoria d'elles han utilitzat estàndards de classificació (DDC o UDC) com a llenguatge de mapeig, tot i que aquesta darrera solució té un baix nivell de profunditat a nivell de matèries però, útil per a sistemes d'informació multidisciplinaris.
- Llenguatges d'intercanvi. Utilització de diversos mapejos per traduir termes que són continguts en un vocabulari per cercar les equivalències en un altre vocabulari.

A nivell d'indexació es recomana només l'ús d'un vocabulari controlat. Es podria utilitzar un altre vocabulari controlat si es compleixen tots els factors que afecten a la interoperabilitat citats anteriorment.

Si els vocabularis controlats, que es volen fusionar són molt diferents en la seva especificitat, poden haver moltes problemàtiques, llavors és indispensable un esforç major, ja que podria ser necessari crear noves jerarquies i altres relacions entre els termes, per aconseguir un vocabulari amb una estructura consistent, a més de requerir professionals (lexicògrafs, etc.) per tal d'abordar problemes associats amb la semàntica i la forma, la especificitat, etc. i sobretot amb la polisèmia o l'homonímia.

Finalment, afegir que la majoria dels vocabularis controlats desenvolupats compleixen qualsevol de les normes ISO 5964 i la norma ANSI/NISO Z3919, no obstant, les diferències i la variabilitat en la terminologia i la sintaxi encara pot causar problemes en diversos dominis o amb la interoperabilitat entre sistemes. No obstant això, hi ha molts treballs al respecte per tal de fer front a la interoperabilitat entre els vocabularis on s'expliquen diferents mètodes de mapeig (Annex D ANSI/NISO Z39.19-2005):

- Modelització. Hi ha un vocabulari com a base que serveix de model de desenvolupament o del qual es deriven altres vocabularis controlats per tal de ser interoperables (la terminologia, estructura, jerarquia i la forma dels vocabularis derivats fomenten la coherència) quant a la),
- Traducció/adaptació. Alguns nous vocabularis controlats s'han creat mitjançant la traducció dels termes d'un vocabulari controlat existent o l'adaptació de termes en un vocabulari controlat existent per al seu ús en una versió traduïda,
- Vocabulari satèl·lit. Similar al mètode de modelització, però es construeix un nou vocabulari com a superestructura que inclou vocabularis controlats que són més apropiats a especialitats concretes d'interès (permet que els vocabularis de nova construcció puguin incloure condicions específiques).
- Node. El vocabulari controlat d'origen es vincula a altres vocabularis controlats més detallats que s'apliquen en un sol node, o enllaços que van a la jerarquia superior dels vocabularis,
- Mapeig Directe. Procés d'establiment de relacions d'equivalència i altres termes entre vocabularis controlats diferents, comporta un procés manual/intel·lectual.
- Mapeig de co-ocurrències, Utilitza la co-ocurrència de termes inclosos en les bases de dades subjacents per trobar famílies de termes relacionats i construir un vocabulari controlat.

- Switching: procés de traducció de termes equivalents d'un vocabulari controlat a un altre, pot no ser un procés directe, sinó que es pot utilitzar un llenguatge de connexió per facilitar el procés.
- Vinculació temporal a través d'una llista d'unió. En aquest enfocament, diferents vocabularis controlats estan units a través d'un programari de mapeig de termes que no són equivalents conceptualment però estan estretament relacionats d'alguna manera. Aquesta llista de termes es realitza de forma dinàmica en resposta a una consulta d'un usuari específic, i els resultats es mostren en una llista d'unió temporal.

Vinculació amb servidors de vocabularis controlats. Enfocament addicional per a la interoperabilitat que implica la transmissió d'una consulta a un o més servidors de vocabularis controlats. Les consultes de l'usuari s'envien al servidor central, que passa a les consultes dels vocabularis registrats, es recullen els diferents resultats, i es presenten d'una forma consolidada per a l'usuari.

5.5.2. Estudis de casos en ciències aquàtiques i marines

Tant en el camp de les *subject gateways* com en les biblioteques i centres de documentació especialitzats en ciències aquàtiques, predomina l'ús dels tesaurs especialitzats com a llenguatges d'indexació i recuperació de la informació. En l'Annex 5.1, hi ha una taula representativa de les *subject gateways* més destacades on queda reflectit l'ús dels tesaurs, però, també cal destacar altres SGs que usen classificacions (Renardus) i *subject headings* (INTUTE) en la indexació, navegació i recuperació de la informació.

D'altra banda l'anàlisi d'aquests recursos, per dificultats d'accessibilitat als documents interns, no ens permet aprofundir a nivell tecnològic en l'arquitectura de la informació que apliquen els sistemes d'indexació i recuperació de la informació i la seva efectivitat. Però sí podem diferenciar les classificacions que s'empren en SGs d'àmbit més multidisciplinari i d'àmbit nacional (Renardus), entre d'altres coses, perquè són llenguatges estàndards i és per això que són sistemes més globals que poden representar millor aquesta cobertura multidisciplinària (vegeu capítol 3). A més, l'estructura numèrica de les classificacions fa que siguin més fàcils de mapejar que no pas els tesaurs o les *subject gateways*. Per contra poden ser sistemes de recuperació de la informació menys precisos i poden ocasionar pèrdues d'informació.

D'altra banda, els *tesaurus* com a sistemes d'indexació són més complexos però, també són més pertinents per entorns especialitzats (INTUTE). A més a més poden fusionar tesaurs, metatesaurus i construccions de concordances creuades. El nivell de profunditat i cobertura temàtica és també més gran.

5.5.2.1. Anàlisi de l'entorn dels sistemes d'informació en ciències aquàtiques i marines

En l'àmbit de les ciències aquàtiques es va dur a terme un estudi quantitatiu amb qüestionaris realitzats via correu electrònic a la xarxa MedLibs (Mediterranean Marine and Aquatic Libraries and Information Centres Network), formada per 45 membres aproximadament, però, només es van rebre 15 respostes (Annex 5.2). El mateix qüestionari es va formular a la xarxa EURASLIC (xarxa Europea de biblioteques i Centres de Documentació en ciències aquàtiques i Marines), però, es van obtenir el mateix nombre de respostes i participants, per tant, l'estudi de mostra escollit és el d'àmbit Mediterrani. A causa de la baixa participació i per corroborar que la mostra seleccionada era l'adequada, també es va procedir a fer una anàlisi qualitativa, basat en la consulta i observació de les pàgines web dels membres participants de la xarxa EURASLIC. Llavors s'analitzaren els sistemes d'informació (opacs, repositoris, etc.) disponibles a nivell web. El període d'estudi i anàlisis dels qüestionaris i de les webs equival al 2007-2008, però durant el 2011 i 2012, tot i revisar les pàgines web de totes les biblioteques de la xarxa EURASLIC, per veure els canvis i possibles innovacions produïdes en el transcurs dels anys, no es varen detectar canvis significatius. Al maig de 2013, es va consultar la literatura científica produïda en el darrer congrés d'EURASLIC (EURASLIC 15th), on apareixien les línies de recerca de les organitzacions participants de l'àmbit de les ciències aquàtiques i marines, en aquest cas, tampoc s'observen diferències significatives. L'evolució cap a nous sistemes d'informació és lenta, i els projectes presentats sobre sistemes d'informació se centren en el desenvolupament de repositoris institucionals. De fet, tal i com s'exposa al llarg de tot el projecte de tesi, la FAO, institució referent tant per temàtica com per ser innovadora en l'aplicació de noves tecnologies relacionades amb sistemes d'informació semàntics (*skosificació* de tesaurs especialitzats i aplicació del *linked open data* en els seus sistemes d'informació, etc.). El sistema d'informació referent de la FAO és AGRIS, aquest encara disposa d'un sistema d'informació semàntic en fase de prototipatge (vegeu OpenAgrisBeta, 2011-).

En aquest estudi, fet a través d'enquestes realitzades a biblioteques i centres de documentació en l'àmbit de les ciències aquàtiques i marines, s'ha avaluat principalment l'ús i la diversitat del llenguatge documental en l'entorn de treball de cadascun dels participants i també si la seva aplicació era dintre o fora dels entorns digitals. Hi ha una segona part de l'estudi centrada en la utilització de formats i estàndards en els sistemes d'informació com, per exemple, ús de protocols d'intercanvi de dades, ús d'estàndards de catalogació i descripció de dades, etc.

Els resultats del estudi mostren clarament com dintre del camp de les biblioteques especialitzades en ciències aquàtiques predomina l'ús dels tesaurus com a llenguatge d'indexació. Veure gràfic 1:

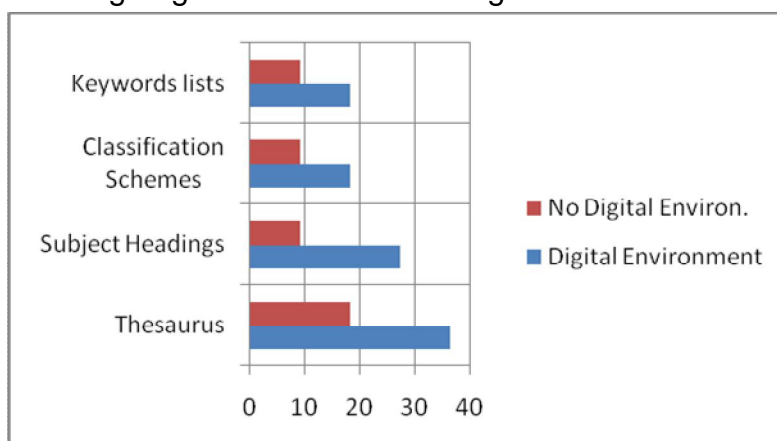


Figura 85. Tipologia de vocabularis controlats usats en biblioteques en ciències aquàtiques (Font: Elaboració pròpia)

Els resultats del qüestionari fet a la xarxa MedLibs també mostren 5 tesaurus representatius de vocabularis controlats en ciències aquàtiques tal com es mostra en la *Taula 20*, però, cal destacar el paper que juguen les *Subject Headings* de la *Library of Congress* (LCSH) i altres vocabularis de representació local. Amb els propis tesaurus i la necessitat de crear eines pròpies d'indexació com llistes de paraules clau, per part dels diferents centres participants en aquest estudi, reflecteixen la gran heterogeneïtat del propi camp científic. A continuació, *Taula 20*, es poden veure les grans disciplines que hi ha dintre les ciències aquàtiques: biologia, agricultura, medi ambient, desenvolupament sostenible, etc.

Titol	Institució	Àrea coneixement	Ontologia (SKOS&RDF)	Llengües	Estructura ⁴⁶
ASFA Tesaurus	FAO (International)	Aquatic Science & Fisheries Abstracts	Si	Multilingüe (Anglès, Francés i Espanyol)	Jeràrquica i Associativa BT, NT, RT, UF, SN
AGROVOC Tesaurus	FAO & Agriculture Organization of the United Nations (International)	Agricultural tesaurus with Fisheries and Aquaculture part	Si	Multilingüe (17 llengües)	Jeràrquica i Associativa BT, NT, RT, UF, SN
GEMET Tesaurus	EEA (Europe)	Environmental Tesaurus	Si	Multilingüe (22 languages)	Jeràrquica i Associativa BT, NT, RT, UF, SN,

⁴⁶ Abreviacions: BT: "broader term", NT: "narrower term", UF: "use for", SN: "scope note", SC: "subject category". All of them use standard structures relations.

					Groups and Themes
NBII Tesauros	CSA; U.S. Geological Survey's Biological Informatics Office (USA)	Biological, Ecological and Environmental Science	Si	Monolingüe Anglès; Inclou SOAP web Services amb diferents tesauros (ASFA, Life Sciences Tesauros; CSA Pollution Tesauros; CSA Sociological Tesauros; CERES/NBII Tesauros)	Jeràrquica i Associativa BT, NT, RT, UF, SN, SC
OECD Tesauros	United Nations (International)	Economic and Social Development	Si	Multilingüe	Jeràrquica i Associativa BT, NT, RT, SN

Taula 24. Tesauros rellevants en ciències aquàtiques (Font: Elaboració pròpia)

Com es pot veure en el camp temàtic d'estudi, els tesauros en ciències aquàtiques s'elaboren amb els estàndards internacionals i compleixen un dels requisits per a la interoperabilitat semàntica, que són multilingües. Només hi ha un cas, el tesauro ASFA consta sols de tres idiomes i probablement es necessitarien fer traduccions si la *subject gateway* proposada vol incloure la resta d'idiomes existents a nivell europeu.

Quant a l'apartat de l'ús de formats i estàndards, la majoria de biblioteques responen a l'ús de formats i estàndards habituals (vegeu capítol 2.4), MARC, MARC21 (amb exportació a XML), com a formats de contenidors i/o de codificació de les dades bibliogràfiques, i ús d'AACR2 com a regles de catalogació. Pel que fa a protocols d'intercanvi d'informació, la majoria de sistemes utilitzen el protocol Z39.50 i un 20% de les institucions entrevistades el OAI-PMH per als repositoris institucionals, ja que en aquell moment de la recollida de les dades (2007-2008), la majoria de biblioteques no utilitzaven repositoris d'accés obert. Cinc anys després, la situació no ha variat significativament, perquè encara hi ha institucions que estan desenvolupant els seus repositoris institucionals d'accés obert (15th Euraslic Conference, 2013). Cal afegir que és possible que moltes de les institucions d'aquest àmbit mai arribin a tenir un repositori propi d'accés obert, ja que estan utilitzant altres

repositoris col·lectius nacionals o especialitzats on publiquen les seves dades institucionals (per exemple: AquaticCommons o OceanDocs).

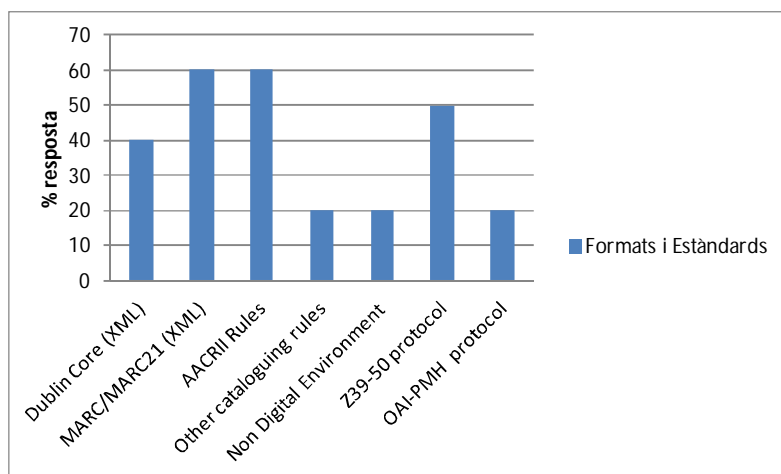


Figura 86. Resultats ús de formats i estàndards en institucions de ciències aquàtiques i marines (Font: Elaboració pròpia)

5.5.2.2. Tendències: mapeig de vocabularis controlats en ciències aquàtiques i marines

Actualment, la integració de vocabularis controlats com a eines de gestió, arxiu i indexació de la informació i la documentació per a la seva futura recuperació en un entorn digital, segueix una clara línia d'investigació, les ontologies. Sobretot quan es tracten de sistemes d'informació multilingües que inclouen col·leccions heterogènies.

El terme ontologia, apareix juntament amb el web semàntic i vindria a ser una terminologia o vocabulari controlat representat mitjançant 'un esquema de dades similar als models de metadades estàndards ja existents (Dubin Core). AGROVOC defineix el terme com "el sistema semàntic que conté termes, definicions d'aquests termes i especificacions de les relacions entre els termes". Per tant, és un llenguatge amb major grau d'interoperabilitat semàntica entre els sistemes d'informació, ja que, tant un intercanvi de dades com el seu mapeig són molt més ràpids i senzills.

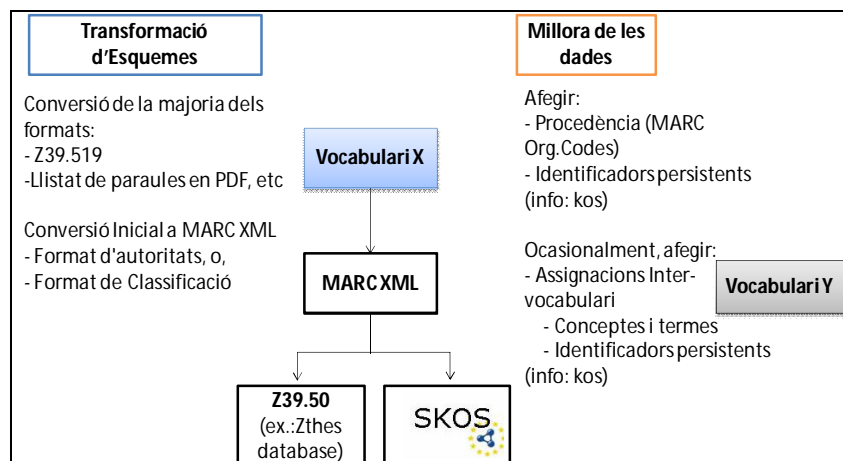


Figura 87. Exemple de procés de mapeig a través de metadades (Font: Elaboració pròpia)

Tanmateix, cal estudiar amb profunditat la situació que hi ha en l'àmbit de les ciències aquàtiques, sobretot, a nivell de biblioteques i centres de documentació existents per poder determinar les dimensions d'un projecte futur de sistema d'informació i establir les línies d'estudi més adequades per executar-lo.

Gràcies als estudis previs realitzats sobre el context i l'audiència es van aportar dades de la heterogeneïtat del camp temàtic i la disgregació de les fonts d'informació existents. D'altra banda, es detecten comunitats científiques i xarxes bibliotecàries internacionals en l'àmbit de les ciències aquàtiques amb la necessitat i interès comú per accedir i difondre la informació científica. A més a més de la manca d'eines avançades de gestió i organització de la informació al web del camp de les ciències aquàtiques que també justifiquen l'esforç d'un projecte comú de *subject gateway*.

L'àmbit del projecte està pensat perquè sigui a nivell europeu per això no només es tenen en compte els resultats obtinguts a la xarxa MedLibs, sinó que s'han avaluat també projectes europeus i/o internacionals més rellevants quant a interoperabilitat semàntica i integració de vocabularis controlats del àmbit de les ciències aquàtiques i ciències afins (agricultura, ecologia, medi ambient, etc.).

El primer exemple és el *CERES/NBII Taurus Partnership Project* (California). Aquest projecte va desenvolupar un protocol general estàndard per a la comunicació i distribució de tesaurus. Van utilitzar eines que eren capaces de crear metadades i construir interrogacions al web. A més a més, van desenvolupar un protocol HTTP utilitzant un tesaurus RDF (XML) com a format de representació que seguia l'estàndard NISO Z39.19. El perfil Zthes Z39.50, basat en la ISO 2788, va ser un model abstracte per a la navegació de tesaurus, i per representar i buscar tesaurus. Es va intentar que el model fos

suficientment general perquè es pogués utilitzar en una comunicació entre tesaurus. Subseqüentment Zthes ha estat utilitzat com a part del ZING, 'Z39.50-International: Nex Generation' dintre de SRW Search/Retrieve web Service protocol (SRW), que inclou els mecanismes d'accés SOAP i URL.

Un altre projecte a destacar és el que du a terme l'AGROVOC (FAO), que converteix tot el tesaurus a ontologia, a través de la conversió de les dades del tesaurus en fitxers RDF, amb el format estàndard OWL (web Ontology Language) que fa possible la representació i mapeig dels diferents llenguatges i terminologies emprats.

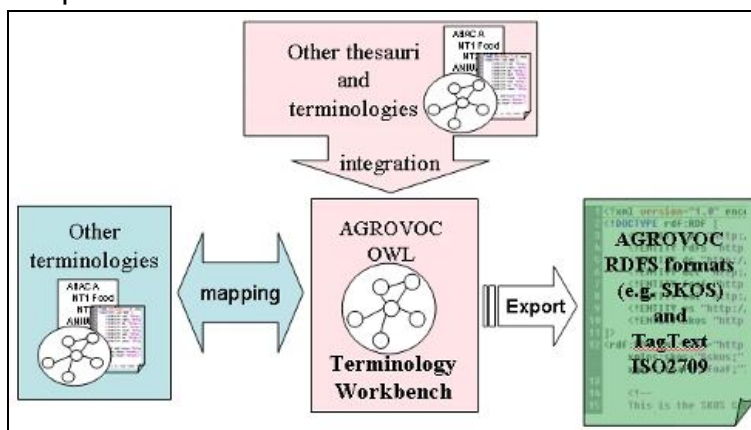


Figura 88. Model d'Arquitectura de la informació AGROVOC (Font: AGROVOC Architecture)

I el darrer projecte a destacar és el de *Marine Metadata Interoperability de la National Science Foundation*. Aquest és el més recent i es basa també en la conversió de tesaurus en fitxers RDF (esdevenint fitxers de metadades) que posteriorment serveixen per al mapeig dels diferents vocabularis integrats en el sistema. Referencia uns 27 vocabularis controlats entre ontologies i tesaurus, però no s'especifica quin està mapejant o no. Cal destacar que l'objectiu del projecte inclou només la part temàtica de ciències marines i descarta els àmbits aquàtics. Finalment, cal afegir el projecte NeOn, un projecte d'ontologies d'àmbit marí però a nivell europeu, en el qual també col·laboren AGROVOC i MMI. És un dels projectes on s'ha fet un mapeig entre dos dels vocabularis d'interès per aquest estudi, l'ASFA i l'AGROVOC, Tot i que Però, els resultats, a nivell de mapeig, no han estat molt optimistes perquè la cobertura temàtica és ben diferent (Lauser, B.; Johansen, G.; Caracciolo, C. 2008).

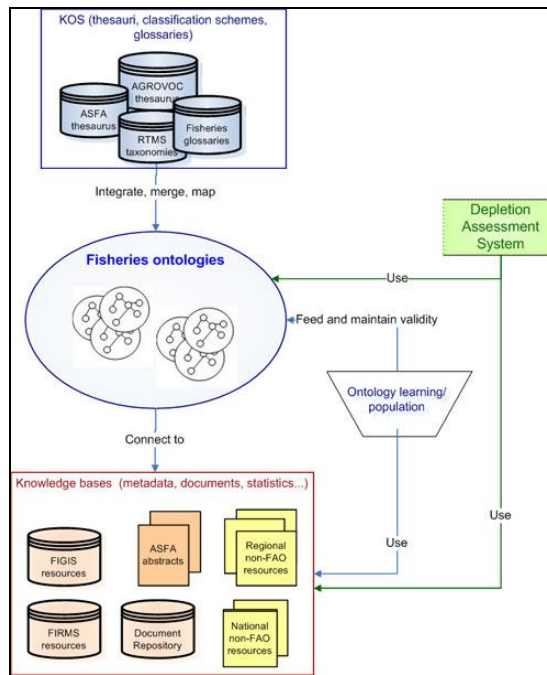


Figura 89. Model d'Arquitectura de la informació del projecte NeOn (Font: AGROVOC architecture)

5.5.3. *Linked data*: model d'interoperabilitat semàntica

Linked data forma part d'una de les activitats que es desenvolupen en la “web semàntica”. La web semàntica vindria a ser una extensió de la web actual dotada de significat, és a dir, un espai on la informació té un significat ben definit, i més fàcilment interpretable tant pels ordinadors com pels humans. El desenvolupament del web semàntic requereix la utilització d'altres llenguatges, com el llenguatge estructurat XML (Extensible Markup Language) i el llenguatge RDF (Resource Description Framework) facilitant als recursos web d'una lògica i un significat, que permeten als sistemes computacionals conèixer i gestionar la informació no només per fer-la accessible a l'usuari, sinó també per a què pugui ser integrada i reutilitzada per altres sistemes o usuaris. L'estandardització dels formats emprats en el web semàntic es gestionen a través del Consorci del World Wide Web (W3C), on hi ha un grup de treball que construeix un document de bones pràctiques *Linked Data Working Group*, a més a més d'existir un grup específic sobre biblioteques, *Library Linked Data Incubator Group*, amb la missió d'augmentar la interoperabilitat global de dades, de la biblioteca a la web.

Les característiques de la pròpia web semàntic i del *linked data*, tenen implícits la interoperabilitat semàntica en si mateixa. Un dels objectius del *linked data* és definir com la utilització de la xarxa per connectar dades relacionades que no s'han relacionat anteriorment, o bé, utilitzar la web per reduir les barreres per a

la interconnexió de les dades vinculades amb altres mètodes (*linkeddata.org*). Wikipedia, defineix *linked data* (LD) com "un terme utilitzat per descriure unes bones pràctiques recomanades per exposar, compartir i connectar peces de dades, informació i coneixements sobre el web semàntic utilitzant URIs (identificador uniforme de recursos) i RDF (Marc de Descripció de Recursos del World Wide Web)".

Segons Tim Berners Lee (2006), "el web semàntic no és només posar les dades a la web. Es tracta de fer connexions, de manera que una persona o una màquina poden explorar les dades existents al web. Amb el *linked data* les dades existents al web es poden trobar d'altres dades enllaçades". Fins ara, els documents de l'hipertext i els seus ancoratges relacionats eren escrits en HTML, però ara, amb el *linked data* les dades estan disponibles amb enllaços que van a conceptes arbitraris descrits en RDF, aquests conceptes s'identifiquen mitjançant adreces URLs del web semàntic. Tim Berners-Lee descriu quatre regles per a l'aplicació de *linked data* al web:

1. Identificació de coses (objectes o conceptes) amb adreces URLs (símbols universals),
2. Utilitzar HTTP i URL's perquè les persones puguin buscar noms, ja que cada cop més augmenta la utilització de nous esquemes URI (i subesquemes dintre l'esquema URL), com per exemple LSIDs, XRIs i DOIs, etc. Això succeeix quan es construeix alguna cosa sota un control independent.
3. Proporcionar informació al web contra una Url. En general, s'han d'observar les propietats i les classes que es troben en les dades i obtenir informació des de l'RDF, RDFS i ontologies OWL, incloent les relacions entre els termes de l'ontologia. El format bàsic és RDF/XML, però, també hi ha l'opció popular alternativa de serialització N3 o *Turtle*. No obstant això, els grans conjunts de dades han de ser proporcionats per un servei de consulta SPARQL, encara que les dades enllaçades bàsiques han de ser també proporcionades.
4. Per crear enllaços a altres llocs. Cal connectar les dades que hi ha en una web, una web il·limitada en la qual es poden trobar tot tipus de dades, igual que l'hipertext web que s'ha aconseguit construir.

Segons el model de qualitat de la informació proposat en aquest estudi, un sistema digital ha de garantir l'autenticitat i la identitat de la informació i els documents, perquè els usuaris puguin accedir a informació de qualitat per fer accessible aquesta informació s'han d'establir diferents metodologies que s'exposaran a posteriori en el capítol d'Informació de qualitat. Bàsicament, hi ha un primer pas on la informació de qualitat és avaluada amb uns criteris preestablerts per l'avaluador del sistema (per exemple: accessibilitat,

rellevància i exhaustivitat), i, posteriorment, aquesta informació s'ha de preservar (exemple de criteris de preservació: accessibilitat, integritat i representació de les dades). Aquestes característiques també són vàlides en el moviment del *linked data*.

Actualment els projectes sobre *linked data* i el web semàntic van creixent molt ràpidament, no obstant, en el camp de les biblioteques, on hi ha bases de dades amb contingut ric i descripcions de les dades relacionades (un document es relaciona mitjançant la seva descripció a altres bases de dades com l'autoritat i bases de dades de matèries, etc.), encara no hi ha un desenvolupament estès. Per tant, el canvi o l'evolució dels sistemes d'informació de qualitat tradicionals cap sistemes semàntics ha de ser tipus de sistemes existents, cal adaptar les dades bibliogràfiques i els seus formats als formats que ja es duen a terme a el web semàntic. Per exemple, els fons bibliogràfics i sistemes d'autoritat accessibles a través d'HTTP han de ser capaços de lliurar el format de descripció de recursos (RDF) que descriu els vincles entre ells. Un exemple d'això és la conversió de tesaurus en formats ontològics (per exemple, tesaurus NBI i tesaurus AGROVOC). Els beneficis d'emprar el web semàntic són molt alts per moltes raons:

1. Intercanvi d'informació. Un major intercanvi d'informació és una major difusió de la col·lecció.
2. L'augment de la funcionalitat. El fet que les dades són intercanviables amb sistemes d'informació externs permet una major retroalimentació al sistema i pot contribuir al desenvolupament de la col·lecció interna.
3. L'augment d'usuaris. Les dades enllaçades ofereixen nous serveis als usuaris, per exemple, la millora dels motors de cerca i els sistemes de recuperació d'informació amb un major grau de comprensió de les interrogacions que realitza l'usuari al sistema, i, per tant, ofereix una major qualitat en la recuperació de la informació.

De fet, les dades es poden mantenir en la seva forma de base de dades relacional, però, en el web semàntic els formats de dades i els protocols d'intercanvi d'informació són diferents. El coneixement està representat en una estructura de tripletes (subjecte-predicat-objecte) que permet una major flexibilitat a través del format RDF, que pot ser expressat per diferents formats, com ara la sintaxi RDF / XML.

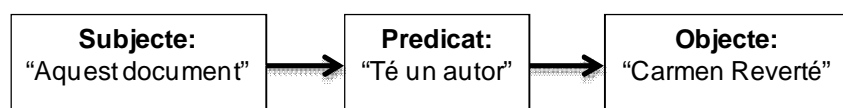


Figura 90. Exemple de representació de les dades amb estructura de tripleta
(Font: Elaboració pròpia)

Desafortunadament, tot i el gran volum de projectes relacionats amb el web semàntic i *linked data*, com s'ha exposat en el capítol d'evolució dels sistemes d'informació, pocs s'implementen en sistemes d'informació distribuïts aplicats al món de les biblioteques i centres de documentació.

L'interès en aquest capítol se centra en la interoperabilitat semàntica dels sistemes d'informació que apliquen *linked data* com a una tendència present i futura (casos d'estudi exposats en els capítols 1 i 2). El *linked data* en si mateix és un facilitador de la interoperabilitat semàntica. En tots els casos estudiats, la interoperabilitat semàntica presenta els següents requisits bàsics:

1. Intercanvi d'informació per diferents sistemes d'informació que poden ser interpretats tant pels ordinadors com per les persones (accessibilitat, reutilització, etc.)
2. Assegurar la precisió del significat de l'intercanvi de la informació i que aquesta sigui entesa per un altre sistema o aplicació no desenvolupada per aquest propòsit (completesa, autenticitat, etc.)
3. Cooperació entre organitzacions sense fissures entre sistemes d'informació (dades interpretables, estandarditzades i codificades)

I les eines que ho fan possible són els estàndards i les tecnologies del web semàntic, juntament amb l'ús de metadades i ontologies que ajuden a la descripció de les dades, accés a la informació/coneixement, recuperació de la informació, intercanvi de dades, reutilització, etc.

Sistemes distribuïts no semàntics	Sistemes d'informació semàntics (<i>linked data</i>)
Intercanvi de dades utilitzant llenguatges de marcatge: XML, MARC, MARC21, etc.	Intercanvi de dades utilitzant llenguatges semàntics: RDF (Resource Description Framework), OWL, etc.
Esquemes de marcatges per validar dades: XML Schema (XSD), etc.	Utilització de regles per validar les dades: SPARQL
Estructura basada en la reutilització (metadades, vocabularis controlats, etc.): XSD, Dublin Core, MARC, etc.	Reutilització de vocabularis i dades de referència per tal d'aconseguir un significat comú (ontologies OWL/SKOS)
Dades emmagatzemades en una base de dades relacional	Dades emmagatzemades amb grafs RDF
Serveis web per fer possible l'accés a	Serveis de dades que fan possible

les dades: SOAP, etc.	l'accés de la informació
Integració de la informació basada amb el sistema	Integració de la informació basada amb interconnexió de dades
Dades i esquemes complerts (món tancat)	Dades i vocabularis incomplerts (Món Obert)
Cada entitat de dades ha de tenir un únic identificador a nivell de sistema d'informació	Cada entitat de dades té en comú identificadors interconnectats, a través de sistemes d'informació a nivell web (DBPedia, AGROVOC tesaurus, VIAF, etc.)

Taula 25. Característiques d'interoperabilitat semàntica entre diferents tipus de sistemes (Font: Elaboració pròpia)

Es conclou doncs, que l'objectiu del *linked data* és compartit per l'objectiu tradicional de les biblioteques públiques, ésser passarel·les de coneixement entre les persones i la informació. Les biblioteques públiques ofereixen accés a la informació i a la documentació per a tothom, i les col·leccions i els serveis no han d'estar subjectes a cap forma de censura ideològica, política o religiosa, ni a pressions comercials, etc. En aquest sentit, el *linked data* també segueix el moviment d'*open linked data*, amb l'objectiu que les dades siguin lliures i disponibles per a tothom.

La diferència és que en un sistema tradicional les dades no eren interpretades pel sistema a nivell de significat. En canvi, amb el *linked data*, la interoperabilitat semàntica és plena i permet que els sistemes arribin a un consens sobre el significat dels elements de les dades i les relacions entre elles. Per tal d'assegurar la interoperabilitat semàntica, és necessari el desenvolupament de vocabularis comuns que descriguin entitats de dades i assegurar que aquestes s'entenguin de la mateixa manera en sistemes diversos.

5.6. Consideracions per implementar un prototipus de sistema d'informació de qualitat a nivell d'interoperabilitat

Les dimensions d'interoperabilitat en un sistema d'informació distribuït i de qualitat, i a més, especialitzat en ciències aquàtiques i marines, haurà de contemplar tres aspectes imprescindibles: la interoperabilitat organitzacional, la interoperabilitat semàntica i la interoperabilitat tècnica.

A nivell d'interoperabilitat organitzacional, com ja s'ha explicat anteriorment implicarà l'establiment d'objectius comuns entre les institucions implicades en el projecte, a més de diferents estructures i processos interns. També s'hauria de comptar amb la col·laboració i suport de les administracions de cadascun dels

països que formin part del projecte. Aquesta part ajudarà a assegurar la interoperabilitat de l'accés dels serveis orientats al usuari.

A nivell d'interoperabilitat tècnica, no s'ha aprofundit molt en aquest tema, però, cal tenir-ho molt en compte perquè, a més d'anar lligada a la interoperabilitat semàntica, té l'objectiu de cobrir a nivell tecnològic la interoperabilitat entre els sistemes i serveis que s'incloguin en qualsevol sistema d'informació distribuït.

La interoperabilitat semàntica és necessària per assegurar la precisió de significat i que l'intercanvi d'informació sigui entès per altres aplicacions. A nivell general, i com a característica principal i element de qualitat, és imprescindible l'ús d'estàndards recomanats per la W3C (RDF, OWL, SKOS, etc.). A nivell específic, en un projecte de cobertura europea i multidisciplinària també s'haurà de tenir en compte la unió de la interoperabilitat semàntica amb el multilingüisme, ja que caldrà assegurar l'accés i la recuperació de la informació en diverses llengües. En aquest sentit, cal afegir que la tendència en la majoria de sistemes d'informació distribuïts és l'ús de vocabularis controlats multilingües, fet que facilitarà molt més la interrogació i mapeig de la informació. Evidentment que s'haurà d'estudiar més detingudament i de forma col·lectiva per totes les institucions participants en un mateix sistema d'informació, per tal de prendre decisions com per exemple, el nombre de llengües que s'hauran d'incorporar en el sistema d'informació, etc. Un altre punt a considerar, tal i com s'ha exposat al llarg d'aquesta tesi, és que l'ús de vocabularis controlats facilita la interoperabilitat semàntica entre els sistemes (Morshed, A., 2013). Un dels vocabularis controlats més estesos en el web semàntic són els tesaurus, d'aquí que un dels processos més viables per millorar la interoperabilitat semàntica en un sistema d'informació seria mapejar els diferents tesaurus que siguin d'interès, és a dir, augmentar la semàntica del sistema. Com s'assenyala en l'Annex 5.1, tots els tesaurus estudiats tenen ja la conversió feta a un esquema RDF, excepte el de OECD que és una versió en paper. Dues opcions són possibles per al mapeig de tesaurus:

- D'una banda, es pot partir dels esquemes RDF i mapejar tots els tesaurus per convertir-los en llenguatges semàntics (conversió a format SKOS o *skosificació*), i, per tant, podran ser fàcilment intercanviables, a més a més de ser un procés automatitzat.
- D'altra banda, una altra opció seria fer un mapeig manual en primera instància i a posteriori automatitzat. Aquesta opció, que hauria de ser aprovada sota consens, oferiria la possibilitat també d'incloure altres vocabularis controlats d'interès que no fossin tesaurus, com les *Library of Congress Subject Headings* (LCSH), i construir així un llenguatge de mapeig entre els diferents sistemes que hagin d'integrar un sistema d'informació distribuït. S'ha de tenir present que en moltes biblioteques i centres de documentació, la LCSH segueix sent un llenguatge

d'indexació molt estès entre els catàlegs i, d'altra banda, les classificacions també continuen coexistint, i poden ser també elements molt importants de cara la interoperabilitat semàntica del sistema. En aquesta línia cal seguir el projecte Europea o HILT. En aquest darrer s'han mapejat tant tesaurus, LCSH, Classificacions (DDC) com altres vocabularis controlats específics.

Els resultats de mapeig entre tesaurus i llistes d'encapçalament de matèries (LCSH), és bastant complex i comporta molts problemes d'interoperabilitat semàntica, principalment perquè les estructures d'ambdós sistemes d'organització del coneixement (KOS) són molt divergents. També existeixen diferències de cobertura temàtica, diferències amb les relacions entre els conceptes, els sistemes de control de sinònims i homònims, diferències semàntiques a nivell de frases preposicionals i adjectives, subdivisions i qualificadors, etc. Una altra aproximació que es va contemplar en el projecte HILT, va ser l'ús del mapeig existent ja entre LCSH i el sistema de classificació DDC (projecte fet per OCLC al 2011, encara sota desenvolupament). Aquesta seria una opció senzilla i viable però actualment en el camp de les ciències aquàtiques no s'utilitza aquest estàndard. Els protocols que s'utilitzen són SOAP i Z39.50 per a la interrogació i Zthes per als fitxers de dades. Els estàndards proposats per a estructurar les dades i la informació en els tesaurus són SRW (que necessita d'altres protocols per funcionar com el Z39.50) i l'SKOS-Core (llenguatge de marcatge per terminologies com els tesaurus), en aquest projecte també es contempla la possibilitat d'usar altres formats com el MARC o Zthes però no se'ls considera tan pertinents com l'esquema RDF SKOS-Core.

5.6.1. Model desitjat d'interoperabilitat semàntica

Usuari/Comunitat interroga al **sistema** i aquest a la **col·lecció** dipositada en diverses bases de dades que són **mapejades** segons un criteris estàndards d'**interoperabilitat** que contrasten la informació per a que la seva posterior **recuperació de la informació** sigui pertinent i precisa.

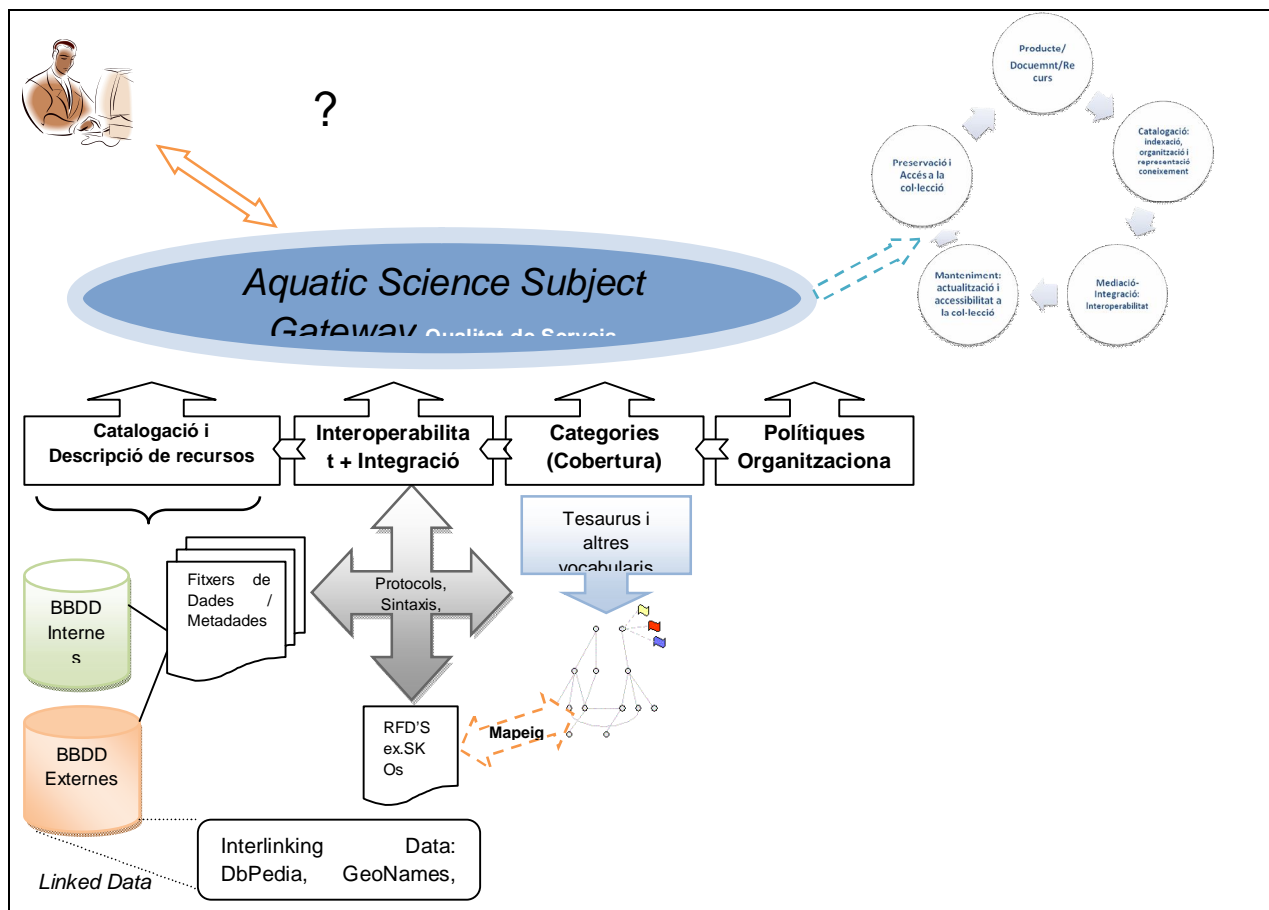


Figura 91. Model d'interoperabilitat semàntica en sistemes d'informació de qualitat (Font: Elaboració pròpia)

Així mateix, perquè sigui un sistema de qualitat ha de complir amb el cicle de vida de la informació. És a dir, des del procés d'adquisició del document o recurs d'informació fins que es converteix en un producte de qualitat per al usuari.

5.7. Conclusions

L'estat de l'art de la interoperabilitat semàntica en sistemes d'informació digitals de qualitat, mostra un context de dades heterogènies, amb una gran desagregació dels recursos d'informació web que han de ser gestionats per grans infraestructures d'àmbit nacional, europees o internacionals.

Es mostra també, un escenari professional actiu i interdisciplinari amb l'objectiu de fer front a l'excés d'informació a Internet i fer possible l'accés, intercanvi i gestió de la informació de qualitat disponible al web. D'una banda, hi ha comunitats i xarxes científiques internacionals en biblioteconomia i documentació (especialitzades o no) i xarxes i comunitats científiques de l'àmbit del web semàntic i *linked data* (principalment provinents de l'àmbit

informàtic). Per tant, un dels primers reptes pendents és la cooperació i la coordinació entre diferents institucions i professionals per tal d'augmentar la interoperabilitat global de dades al web, a més d'analitzar i millorar els actuals i futurs desenvolupaments semàntics.

Com s'ha anat exposant al llarg del capítol, en el món de les biblioteques i centres de documentació no es troben moltes aplicacions i/o serveis semàntics en funcionament, per tant, es pot afirmar que encara no s'ha assolit el web semàntic com una tecnologia més d'aquest àmbit. No obstant això, any rere any creixen els desenvolupaments de prototipus i projectes relacionats amb el *linked data* i les biblioteques i repositoris digitals. Així doncs, un repte que cal tenir en compte és fomentar i constituir la filosofia del *linked data* com una part essencial de la tasca d'experts en ciències de la informació.

Al llarg del capítol s'han exposat els avantatges del *Linked Data* en el camp de les tecnologies de la informació (TI) i les ontologies, on el *linked data* pot establir-se com una eina útil per garantir l'intercanvi de la informació i la integració de recursos heterogenis entre si. El paper dels professionals de la informació està canviant amb els nous avenços tecnològics, el que significa és que cal adaptar-se als canvis i utilitzar aquests canvis tant importants com a noves oportunitats per al desenvolupament de nous projectes.

Com a professionals de l'àmbit de la biblioteconomia i la documentació, el nostre paper dintre l'escenari de la interoperabilitat semàntica es centra en les següents línies:

- Normalització/Estandardització. Des de la definició dels termes, utilització de normes, vinculació de dades i metadades, seguint senzilles instruccions que faran que les dades del sistema siguin accessibles per a molts més usuaris, proporcionant informació de qualitat, etc.
- Manteniment dels sistemes d'organització del coneixement (KOS), i garantir la interoperabilitat entre ells, amb l'ajuda dels usuaris, a través d'adoptar polítiques i normatives relacionades amb els vocabularis i el contingut. La responsabilitat de preservar els estàndards inclou mantenir les versions, l'accés a les versions actuals i anteriors, i donar ràpida resposta i al temps adequat, a les sol·licituds d'informació dels usuaris, mantenint sempre el contingut actualitzat. En la majoria dels casos estudiats, aquestes premisses encara no s'han aconseguit fins ara i, per aquesta raó, és necessari seguir treballant per als desenvolupaments del web semàntic com és el cas del mapeig de vocabularis controlats, assenyalat com a un factor clau d'interoperabilitat semàntica per als desenvolupaments presents i futurs de sistemes d'informació semàntics (importantíssim el mapeig de les col·leccions locals amb la resta de col·leccions web, per tal d'enriquir la col·lecció, i facilitar l'accés i l'intercanvi d'informació entre diferents sistemes i usuaris).

El futur de la interoperabilitat semàntica és paral·lel a l'evolució i desenvolupament del web semàntic. En aquest sentit, l'objectiu general per a fer front a la interoperabilitat semàntica, des del punt de vista del professional de la informació, rau en poder millorar els vocabularis controlats (conversió a ontologies/llenguatges més comprensibles per màquina) i les eines relacionades amb l'alineació per aconseguir un alt grau d'interoperabilitat semàntica.

Finalment, es conclou que la interoperabilitat semàntica ha de ser vista com un element de sostenibilitat per als professionals de la informació i/o experts en gestió del coneixement (intercanvi, reutilització de la informació, enriquiment de les col·leccions amb documents/informació relacionada, etc.).

CAPÍTOL 6. INFORMACIÓ DE QUALITAT: PROPOSTA D'UN MODEL D'IQ PER UN SISTEMA D'INFORMACIÓ DISTRIBUÏT I ESPECIALITZAT

6.1. Introducció

En la literatura científica, i segons des de la disciplina que es tracti el tema, podem trobar diferents perspectives del concepte *Informació de Qualitat* (IQ)⁴⁷ el context conceptual del qual és fonamental per al disseny de sistemes d'informació de qualitat. En la massa crítica publicada sobre aquest tema bàsicament destaquen les àrees del Màrqueting, Auditories de la informació i la Biblioteconomia i la Documentació.

En el context disciplinar de la Biblioteconomia i la Documentació fins fa poc només hi destacaven dues línies de recerca que aborden el concepte d'IQ i que són en primer lloc, els estudis d'usuaris centrats en "necessitats d'informació"; i en segon lloc, l'estudi dels "sistemes d'informació". Tanmateix, en els darrers anys ha sorgit una tercera línia d'estudi que té com a nucli els objectes o documents digitals i els analitza com a indicadors d'informació de qualitat en sí mateixos en el context dels sistemes d'informació digitals de qualitat.

Per aquest motiu, en aquest apartat s'han analitzat els treballs més representatius de l'àrea d'IQ i les contínues revisions que podien aportar noves dades, a més de treballs específics centrats en l'avaluació de la IQ en entorns digitals com poden ser biblioteques digitals, pàgines web i *quality subject gateways*.

La majoria dels estudis sobre la *Informació de Qualitat* analitzats proposen els seus models d'IQ amb els quals s'avalua un sistema o context determinat. Cadascun dels models són representats per unes dimensions d'IQ, i cada dimensió té els seus atributs⁴⁸ que no són més que característiques específiques que defineixen la pròpia informació de qualitat. Tant els models d'IQ com els seus atributs es basen majoritàriament en els conceptes teòrics de la literatura científica del context empresarial, les auditories de la informació, el disseny de sistemes d'informació i/o serveis, així com també els estudis de satisfacció d'usuaris. Tanmateix, hi ha altres estudis que s'analitzaran i que també contempen la IQ des de la perspectiva de l'usuari i les seves necessitats.

⁴⁷ IQ: són les sigles del concepte Informació de qualitat o *Quality Information* (terme en anglès) que utilitzarem al llarg d'aquest capítol perquè són ja d'un ús extensiu en tota la literatura científica d'aquest àmbit.

⁴⁸ Cal destacar les diferents nomenclatures que es troben en la literatura per aquest mateix concepte: atributs, indicadors, factors, propietats o característiques entre d'altres.

Les definicions d'IQ més generalitzades i destacades són recollides en el treball de revisió d'Eppler (2000) on sistematitza la *Informació de Qualitat* com:

- IQ és informació “fit for use”. Aquesta conceptualització d'informació ja “preparada per utilitzar” per l'usuari final la contempen la majoria d'autors estudiats (Wang, 1996).
- IQ és el resultat de les expectatives dels usuaris. Aquesta conceptualització és també compartida per la majoria d'autors (Bovee 2003; Kim 1994).
- IQ és informació que coneix les especificacions o requeriments. Per exemple, les metadades. Aquest és un concepte més actual i no hi ha molts autors que el mencionin des de la perspectiva de la IQ (Naumann, 2000), però que està cobrant un gran protagonisme.

IQ és el nivell d'informació que té contingut, forma i característiques amb un valor específic per als usuaris, com per exemple el temps (Kahn, 1998; Stvilia, 2007).

La IQ es dóna quan la informació és d'alt valor afegit per als usuaris, la informació serveix per a la presa de decisions (Davenport, 1997; Kahn 1998).

- La IQ també es pot definir com la diferència, d'una banda, entre la informació requerida i determinada per un objectiu, i, d'altra banda, en la informació obtinguda (Miller, 1996).
- IQ és la característica de la informació que permet conèixer els requeriments funcionals, tècnics, cognitius, estètics dels productors, administradors, consumidors i experts d'informació (Eppler, 1999; Lee, 2002).

6.2. Models d'IQ que s'apliquen des de la perspectiva de l'usuari

El model d'IQ més conegut i estès és el de Wang (1996) que realitza un estudi empíric per analitzar les dades que realment són importants per als usuaris, i determinar així les característiques emprades per avaluar si les dades que s'obtenen són adequades a les seves necessitats.

El primer pas és identificar les necessitats dels usuaris, després identificar quina estructura jeràrquica (o escala de valors) tenen per a l'usuari i finalment mesurar cadascuna d'aquestes necessitats. Els autors elaboren 2 experiments amb usuaris. En el primer demanen als usuaris que els elaborin llistats amb atributs que corresponguin a IQ, i en el segon, s'avalua la importància de cadascun dels atributs que s'havien representat. A partir d'aquí es pot dibuixar ja un marc d'IQ preliminar on s'agrupen les dimensions (tant les que vénen donades pels usuaris com les conegudes per la literatura seguint la filosofia de “fitness for use”).

Finalment, el model de Wang queda representat per 4 categories que contempen els conceptes d'**accessibilitat** de les dades, **interpretativitat** de les dades, **rellevància i precisió** o **exactitud** de les dades.

En aquests estudis exploratoris es van extraure 179 atributs que primerament acaben agrupats en 20 dimensions, i posteriorment amb 15 (vegeu Annex 6.2). Els atributs més importants per als usuaris inicialment van ser:

- *Precisió,*
- *Fiabilitat,*
- *Nivell de detall,*
- *Identificació d'errors i la correctesa de les dades*

Finalment, per construir el model d'IQ canvien les 4 categories que s'havien seleccionat prèviament i són etiquetades amb conceptes més genèrics i que representen millor el contingut de les seves dimensions (vegeu Annex 6.1):

- *Intrínseca,*
- *Contextual,*
- *Representacional i*
- *Accessibilitat.*

Wang (1996), com molts altres autors, segueix una perspectiva empresarial. El seu model d'IQ mostra una visió del màrqueting on es busquen els "atributs del producte" per atreure el client. Així com el fet de buscar les agrupacions en famílies de factors que també és una estratègia que ens ve donada des del màrqueting.

Molts d'altres autors han analitzat la IQ al llarg de la història de la literatura especialitzada fins als nostres dies, i han seguit l'exemple de *Wang* tot estudiant els models d'IQ basats en la perspectiva de l'usuari. Malgrat repetir les mateixes dimensions categoritzades però, distintament, cal destacar l'aparició d'un nou concepte que alhora s'allunya de la literatura científica i és la *representativitat de les dades*. Probablement, aquest sigui un concepte que ve marcat per l'usuari, no obstant això, no ho podem afirmar categòricament perquè només en la biblioteconomia i la documentació la *representació de les dades* és una disciplina en si mateixa (catalogació descriptiva), i, està lligada forçosament amb l'accessibilitat (catàlegs en línia).

Un altre concepte que apareix el 2003, dins dels models d'IQ basats en l'usuari és el terme d'*integritat* (*Bovee, Sirvastava i Mak, 2003*). A més, la perspectiva empresarial canvia i no dóna tan valor al "producte", sinó a la seva *Accessibilitat i Interpretació*. En aquest context, si en un model d'IQ d'entrada falla ja la dimensió d'*Accessibilitat*, i, la informació és inaccessible per a l'usuari, la resta de valors representats en el model d'IQ ja no són aplicables. El mateix passa quan la informació a la que accedeix l'usuari és intel·ligible, el model d'IQ tampoc funciona.

Una altra perspectiva que canvia és el concepte de *Rellevància*, on deixa de ser un sol atribut (per *Wang*, el concepte era representat dintre la dimensió

Contextual) i passa a ser una dimensió principal que agrupa altres atributs com la *volatilitat*, *frequència de canvi* d'un recurs o d'*actualització* i l'atribut de *data*, (Bovee, 2003). Des d'aquest enfocament, la informació és rellevant quan ho és en el domini de l'usuari i segons els seus propòsits o necessitats d'informació, per conseqüent, no tota informació rellevant per a un usuari ho ha de ser per a la resta.

De fet el concepte de *rellevància* és fonamental en un camp com són les passarel·les temàtiques, ja que la seva tipologia d'usuaris són experts (àmbit de recerca) i, a més a més, engloben camps temàtics diversos, diferents tipus d'organitzacions i diferents necessitats.

Els estudis experimentals dels models d'IQ existents no contemplen grups d'usuaris de diferents organitzacions, però, sí contemplen diferents grups d'usuaris d'una mateixa organització. Així doncs, aquesta podria ser la resposta al per què la rellevància en els models d'IQ no sempre està dibuixada com una dimensió d'IQ. Pel que fa a la rellevància com a dimensió de la IQ està relacionada amb el punt de vista de l'usuari però també amb la matèria, les necessitats de l'usuari, els factors cognitius i els factors de situació (Kim, T.P., 1994). És evident que normalment la rellevància com a model d'IQ s'estudia dins el camp de la RI (Recuperació de la informació), mitjançant la interrogació del sistema de RI i el comportament de l'usuari amb aquesta interrogació

Un altre estudi a destacar és el model d'AIMQ (Lee, et al. 2002), basat en el model d'IQ de Wang (1996), però, en aquest cas, aquest model contempla les avaluacions d'IQ a través de tres components:

1. Un model 2X2 que té en compte els consumidors i els gestors de la informació. Aquest model depèn de si la informació és considerada com a producte o com a servei, i, si les millores poden ser avaluades contra una especificació formal o expectació de l'usuari.
2. El segon component és un qüestionari per mesurar la IQ entre les dimensions d'IQ importants per als usuaris i gestors. Diverses dimensions juntes mesuren la IQ per cada quadrant del model 2x2. Aquest model pot servir també per avaluar la qualitat de la informació en les organitzacions.
3. El tercer component consisteix en dues tècniques d'anàlisi per interpretar les avaluacions recollides a partir dels qüestionaris.

Aquest model està enfocat cap a l'avaluació de la IQ en una organització i no pas cap a entorns digitals, tot i que les tècniques i mesures basades en l'usuari podrien ser perfectament emprades per construir un model d'IQ en qualsevol àmbit disciplinar i informacional.

6.3. Models d'IQ no basats només amb l'usuari

Hi ha molts altres models d'IQ (per exemple, Naumann i Rolker, 2000) que no es basen només en l'usuari, evidentment és un factor clau en tot estudi d'IQ, però cal tenir en compte que la percepció de l'usuari és subjectiva, i que es basa en la seva perspectiva personal, expertesa i el coneixement adquirit. No obstant això, el concepte d'usuari i el seu comportament segueix tenint una gran influència en el marc de l'avaluació de la IQ, perquè és l'usuari qui acaba decidint si la informació és o no de qualitat per a les seves necessitats.

Així doncs, altres factors claus a estudiar en un model d'IQ són els propis documents, el sistema i/o altres factors externs que puguin afectar a la IQ com ara l'actualització o la volatilitat, ja que tot model d'IQ basat en un entorn digital, com són les *subject gateways*, ha de contemplar l'adaptació als canvis i si és possible poder-s'hi avançar com a sistemes proactius.

6.4. Model IQ basat en els documents

Els principals autors que donen referència aquest model d'IQ són Nauman (2000) i Stvilia (2008). El model IQ basat en els documents fa referència a un recurs d'informació on per si mateix proporciona IQ, és a dir, el recurs/document proporciona *metadades* incrustades que serveixen per avaluar la pròpia qualitat del document (autoria, data de publicació, etc.). Malauradament, no poden ser l'única característica del sistema d'avaluació perquè tot i estar estandarditzades hi ha moltes pàgines web, que no usen *metadades*. Altres factors que es poden avaluar en els documents o informació inclosa en un sistema poden ser la *completesa*, el seu *cost*, la *fiabilitat*, l'*actualització*, els *formats*, les *versions*, etc. És a dir, la valoració del recurs es dona per la mateixa informació que proporciona. En aquest tema tampoc hi ha molts autors que siguin explícits, i no aprofundeixen en l'ús de metadades com a mesura d'IQ, però, en aquest sentit cal afegir que tot model d'IQ ha d'incloure metadades estàndards. Així, n el cas de les *subject gateways*, com en la majoria de biblioteques digitals i Repositoris s'usarà l'estàndard Dublin Core, ja que no només permet la descripció de recursos d'Internet sinó que la seva flexibilitat fa que pugui codificar-se en diversos formats (HTML, XML, RDF, etc.), i per això el DC suporta un nivell d'interoperabilitat major que altres tipus de metadades.

Les metadades Dublin Core (Annex 6.4), es componen de quinze elements bàsics de descripció de documents electrònics (*title, creator, subject, description, publisher, contributor, date, type, format, identifier, language, source, relation, coverage, rights*), els quals faciliten informació bàsica sobre el document mateix.

Si es volen utilitzar metadades més específiques que ajudin a donar més valor al recurs descrit s'han d'utilitzar Metadades dublin core qualificades (*qualified Dublin Core metadata*). El Dublin Core qualificat és una extensió del Dublin

Core, en què alguns elements compten amb un qualificador que els fa més restrictius, però, que incrementar l'especificitat o precisió de la metadata. A més a més, les *qualified Dublin Core Metadata* inclouen esquemes de codificació que poden incloure "vocabularis controlats", notacions formals o regles de *parsing* entre d'altres. Tres dels elements més característics del dublin core qualificat són: l'Audiència, Procedència o Provenença i Propietari o Titular dels Drets d'Autor.

En un entorn digital com són les *subject gateways* és evident que les metadades són un dels elements claus del model d'IQ. Les metadades donen valor afegit a un document o recurs digital perquè són elements de descripció que aporten informació addicional del document (*copyright*, accessibilitat o versió, data de creació, modificació, etc). Però són també importants en la recuperació de la informació, de manera que juguen un paper clau dins d'un sistema d'informació/documentació digital.

6.5. Model d'IQ basat en el sistema d'informació

Tot seguint a *Naumann* (2000) sorgeix el model d'IQ basat en l'avaluació de la recuperació de la informació i/o per l'avaluació del *procés d'interrogació del sistema*. L'accés a la informació és un dels indicadors més importants en un model IQ basat en sistema, sobretot, quant al temps de resposta que és un dels factors a tenir més en compte des de la perspectiva d'un usuari.

Es pot dir que les avaluacions centrades en el sistema i les seves interrogacions és un mètode d'avaluació d'IQ més objectiu, però alhora és temporal perquè aquest tipus d'avaluació no pot oferir-nos un model d'IQ fix, ja que és canviant, els processos de cerca d'informació (ISP) poden canviar d'interrogació a interrogació. El propi *Naumann* (2000) no dissenya cap model, sinó que proposa només criteris per avaluar la IQ, ja que la majoria de models d'IQ existents es basen en la definició d'atributs i no en com s'han d'avaluar aquests. En estudis posteriors, *Naumann* i *Sattler* (2006) proposen tècniques d'avaluació basades en qüestionaris (criteris subjectius i no funcionals), tècniques de mesura per vectors (on s'avaluen dimensions com la *completesa*, *precisió*, *consistència* o *actualització*), i mètodes de rànquing entre d'altres.

Una altra perspectiva molt diferent és la que dóna *Stvilia* (2007) és el model d'IQ basat en la tipologia de problemes que es poden donar en la dimensió d'IQ. A partir de les problemàtiques detectades en un sistema d'informació es desenvolupen els models d'IQ més adequats per cadascuna de les situacions donades. A diferència d'altres autors, *Stvilia* estudia dos tipus de col·leccions amb les seves problemàtiques d'IQ, una basada en objectes (metadades Dublin Core), i l'altra en articles (una enciclopèdia online). Identifica quatre variables principals:

- El mapeig,

- Els canvis d'entitat d'informació,
- Els canvis en l'entitat o condició destacada i Els canvis de context.

Aquestes variables volen afrontar la problemàtica d'un entorn tan canviant com és Internet, per quan l'espai informacional és actiu i això afecta també a la temporalitat dels objectes. Els tipus de problemes més comuns identificats per *Stivilla* són:

- La *Incompletesa*,
- La *Redundància*,
- L'*Ambigüitat*,
- La *Inconsistència* i
- L'*Exactitud* o *Precisió*.

Les dimensions en aquest model (Annex 6.1, figura 3) són anomenades taxonomies i són agrupades en tres categories:

1. *Intrínseca*. Avaluen atributs interns o característiques de la informació en relació a estàndards de referència, com per exemple els llenguatges d'etiquetatge o les metadades. La informació ha de poder ser representada per estàndards (validació d'HTML), formateig de dades, actualització, cohesió, consistència semàntica i estructural, etc.
2. *Relacional*. Aquesta categoria permet mesurar la relació entre informació i el seu ús en el context. Les mesures d'avaluació estan relacionades amb mesures de reputació o autoria que poden ser desplegades per una cultura o comunitat determinada. Com per exemple, en l'àmbit de la recerca, on l'ús del *Factor d'Impacte* és una mesura d'avaluació de les revistes científiques. Dins d'aquesta dimensió hi ha la subclasse de *Representació*, com es representa la informació a l'usuari. Aquestes dimensions permeten veure com es reflecteixen les entitats d'informació en algunes condicions externes que es donen en un context determinat (adequació, precisió, exactitud, completesa, naturalesa, complexitat, redundància, rellevància, consistència, volatilitat de les dades, etc.). Es té en compte que si les entitats i condicions externes canvien, les característiques contextuais o relacionals també poden canviar. El context en aquest model es refereix al context d'una activitat (xarxes socials, comunitats, accions, guies de bones pràctiques, normes, convencions, estàndards, etc.).
3. *Reputació*. Té l'indicador o atribut d'*Autoritat*. Serveix per mesurar la posició d'una entitat o objecte d'informació en un àmbit o estructura cultural o d'activitat. Es determina pel seu origen, d'on provenen les dades.

Tot construint les taxonomies de les dimensions, en aquest model es contemplen també 41 tipus generals de mètriques de funció, que estan implementades amb Java, i es poden reutilitzar segons el context específic que s'origina. Aquestes mètriques són possibles perquè un dels contextos que és avaluat per l'autor són les *Metadades*, preses com a mesures indirectes d'IQ. Les metadades són un coneixement molt valuós per conèixer plenament una entitat d'informació perquè ens permeten obtenir informació ràpidament sobre *qui* o *què* va fer el document, i *d'on*, *quan*, *com* i *perquè*. Com bé apunta *Stivilia*, aquesta informació que es pot usar per a molts propòsits i sobretot com a indicadora d'IQ respon a les següents demandes:

- El *Què* en permetre avaluar la qualitat de les contribucions individuals i les transaccions.
- El *Quan*, *On*, *Com* i *Per* *què*, que informen del context del document o entitat d'informació, ajudant a interpretar les dades i evitant així casos d'ambigüitat.
- El *Com* i el *Per* *què*, per quant contribueixen a l'acumulació de coneixement, intercanvi i aprenentatge. I cal no oblidar que aporten IQ relacionada amb el context (per exemple, estàndards, normes i versions que s'apliquin al document).
- El *Qui*, que contribueix en les avaluacions de qualitat de forma indirecta ja que l'autoria és un factor clau en la IQ.

Un marc conceptual basat en les problemàtiques d'IQ es construeix principalment amb la identificació i anàlisi de les activitats que realitza aquell objecte o entitat d'informació, i per tant, s'ha d'analitzar també la dimensió amb la que està relacionada. *Stivilia*, identifica quatre grups d'activitats amb problemàtiques d'IQ:

1. Activitat de **Representació dependent**. Activitats que depenen de com es representa un objecte o entitat de informació, o bé, com una entitat representa a una altra o alguna condició. Per exemple: el procés d'elaboració del document i la inclusió de metadades són dues activitats claus a tenir en compte per representar i visualitzar la informació. Un cas clar és la informació que s'extreu del format del document com la dimensió, la seva estructura, si requereix de cap aplicació específica per ser llegit, etc.
2. Activitat de **Descontextualització**. Activitats on s'usa la informació fora del seu context original de creació. Per exemple, eliminar informació i posar-la en una nova col·lecció per donar suport a altres necessitats d'informació específiques. Cal pensar que en un sistema d'informació

s'ha pogut incloure informació rellevant per un usuari, tot i esdevenir soroll documental per un altre.

3. Activitat d' **Estabilitat dependent**. Activitats que depenen de com estable és la informació i de la importància que té la entitat d'informació. Pot haver-hi canvis d'entitats d'informació o canvis temporals, sobretot, quant a aspectes *intrínsecs* o en les mesures de *representació*. Qualsevol llacuna entre la demanda i els nivells de qualitat de qualsevol dimensió d'IQ lligades amb les categories *intrínseques* i *representacionals* poden portar a un problema d'IQ. Amb aquest punt també fa èmfasi *Bovee (2003)*, ja que si la representació del coneixement o de la informació que arriba a l'usuari no és l'adequada la resta de valor ja no és important i, aquell document o entitat d'informació ja és descartada per l'usuari. El mateix passa amb l'*Autoria*, d'on prové la font o entitat d'informació pot ser el començament d'un problema d'IQ. Per exemple: indicar la font d'informació o cita bibliogràfica d'on s'ha extret el recurs o document recuperat per l'usuari. D'aquesta informació no tan sols extraïem informació de l'autoria sinó que aquesta mateixa informació en proporciona dades de la veracitat de la informació i l'accessibilitat.
4. Activitat de **Comprovació dependent**. Activitats que depenen de la qualitat de les metadades de la informació verificada. Es poden trobar problemes d'IQ en relació a l'autoria o la reputació del recurs, la font i del propi objecte o entitat d'informació. Per exemple: si no s'inclou informació sobre l'autoria i els drets d'autor de l'obra ens podem trobar que aquesta pot està sotmesa a qualsevol ús, reproducció i explotació externa no controlable.

En el model d'IQ de *Stivilia*, les dimensions de *Reputació* i *Relacionals* no són directament avaluades en relació a les problemàtiques d'IQ. D'alta banda, *l'accessibilitat*, *la seguretat* i *la verificabilitat* si són dimensions mesurables i ajuden a avaluar i preveure els nivells d'IQ de les dimensions *Intrínseca* i *Representacional*.

En tots dos estudis que realitza *Stivilia* proposa possibles mètriques per mesurar les problemàtiques trobades, que es basen principalment en estudis estadístics. Dels dos estudis que duu a terme, només en el cas 1 (lligat a les metadades) proposa possibles mètriques per avaluar cadascuna de les problemàtiques trobades, la majoria amb tècniques produïdes amb processos automatitzats.

Problemàtica	Dimensió	Tipus de problemes	Possibles mètriques
<i>Ambigüitat</i>	Intrínseca	Valors contradictoris d'elements iguals. Per exemple: repetició de dates	Comptar els casos dels elements iguals que tenen valors diferents
<i>Imprecisió</i>	Relacional	Enllaços trencats	Comptar enllaços trencats
<i>Incompletesa</i>	Intrínseca	Elements o etiquetes de metadades buits	Comptar les etiquetes buides; comptar els valors incomplets; i el número dels diferents elements
	Relacional		FRBR índex de suport per esqueses DC. Elements requerits: Títol, Creador, Matèria, Descripció, Data, Format, Identificador i Drets (propietat intel·lectual)
<i>Inconsistència</i>	Intrínseca	Formats inconsistents o representació dels mateixos elements	Comptar els casos d'elements iguals que usin formats diferents
	Relacional (semàntica o estructural)		Comptar els casos d'elements "mal usats"; Comptar les instàncies dels elements amb formats no recomanats per les guies
<i>Redundància,</i>	Intrínseca	Elements repetits amb els mateixos valors (duplicats)	Comptar els casos dels esquemes repetits, Soroll d'informació [Conté=1– (mida del terme o vector agafat després de contenir i parar)/(mida del objecte abans de ser processat)]

Taula 26. Model IQ per Stivilla (2007)

A través de l'anàlisi de les problemàtiques es desvela que hi ha quatre dimensions considerades les més importants a tenir en compte: la *Precisió*, la *Completesa*, la *Consistència* i la *Redundància*.

6.6. Altres tipus d'avaluacions de la IQ

En la literatura científica trobem també molts d'altres autors i sistemes d'informació que avaluen la IQ amb simples **llistes d'atributs o criteris** que consideren significatius per dur a terme les tasques d'avaluació de la IQ dels seus entorns d'informació. En la majoria dels casos no es presenta cap model d'IQ perquè s'estudien entorns digitals i difícilment es pot crear un model d'IQ genèric i atemporal. Aquesta mateixa situació també es dona quan es parla de l'avaluació de webs o recursos d'informació digitals. Així doncs, en la majoria de casos apareixen llistats i llistats de criteris d'avaluació on difícilment podem diferenciar la subjectivitat i l'objectivitat de la selecció dels criteris.

Quant a la IQ basada en l'avaluació de recursos d'informació digitals i tot seguint la literatura científica, destaquen principalment les propostes de dos autors: *Miller* (1996) i *Alexander* (1999). *Alexander*, des d'una visió basada en la literatura científica més aplicada, desplega l'avaluació dels webs en 6 criteris d'IQ fonamentals que apareixen al llarg de tota la literatura científica: **autoritat** (informació validada i visible per a l'usuari), **exactitud o precisió** de les dades (fiable i lliure d'errors), **objectivitat**, **actualització** (quant al contingut), la **cobertura** i **audiència**. Tots aquests criteris són aplicats tant per a la selecció de recursos com per al disseny d'interfícies, l'accessibilitat i la usabilitat web. Sobretot, preocupa a aquest autor l'avaluació d'aquells aspectes que poden afectar l'orientació i navegació de l'usuari en un entorn web. *Miller*, en canvi, des d'una perspectiva empresarial, defensa les múltiples dimensions d'IQ, i també descriu la IQ orientada cap a l'usuari, tal com anomena: "IQ segons com la informació és percebuda i utilitzada per l'usuari". A més, l'autor identifica dues fases en tot procés d'IQ, per una banda, subratlla quins atributs són importants, i per altra banda, determina com aquests atributs afecten a les qüestions dels consumidors. També destaca que les dimensions d'IQ poden tenir múltiples usos tals com millorar l'efectivitat d'un sistema d'informació o per desenvolupar estratègies d'informació de qualitat dins de les organitzacions.

Miller aposta per 10 dimensions d'IQ: *Rellevància*, *Precisió* o *Exactitud*, *Actualització*, *Completesa*, *Coherència*, *Format* (com es presenta la informació a l'usuari), *Accessibilitat*, *Compatibilitat*, *Seguretat*, *Validació*. També cal destacar que és l'únic autor que assenyala la **Rellevància**, com a factor clau per a la IQ, ja que defensa que cal vetllar per l'usuari i les seves necessitats d'informació. I si en un sistema es troba informació no rellevant per a un usuari, no voldrà dir que hi ha poca qualitat, sinó que aquell recurs, document o informació probablement va adreçada a un altre tipus d'usuaris.

Fins ara, tots els models d'IQ que s'han revisat corresponen a contextos genèrics, com pot ser una organització o qualsevol entorn o sistema d'informació. Tanmateix, la resta d'estudis que tracten la IQ en els entorns digitals i principalment en el web no construeixen models d'IQ, sinó llistats de criteris que serveixen per avaluar la IQ i que són extrets de la literatura especialitzada existent i que hem estudiat. En aquest context cal destacar els treballs de dos autors més recents, *Codina* (2003 i 2006), i *Gonzàlves* (2007).

Codina (2003), elabora una metodologia d'anàlisi i avaluació de recursos digitals seguint la línia metodològica d'elaborar llistats d'atributs que són els indicadors que mesuren els recursos d'informació digitals, per tal de parametritzar el nivell d'IQ. L'autor descriu un llistat de 55 indicadors (agrupats en 9 categories genèriques, veure Annex 6.2, taula 5) que són representats en 4 dimensions (Contingut, Navegació, Visibilitat i Usabilitat).

En aquest procés metodològic s'aplica l'ús de fitxes descriptives on cada recurs és sotmès a una puntuació, segons les necessitats definides, i si compleix o no amb els diferents atributs que es marquen. . Aplicar aquesta metodologia per a l'avaluació de recursos digitals pot semblar un procés molt fàcil, però, vertaderament la subjectivitat de tot el procés en aquesta avaluació, no sempre és controlable si els avaluadors no són experts en la matèria. A més, aplicar tants atributs pot ser també un procés molt costós i l'èxit de l'avaluació dependrà del nombre de recursos dels quals disposi i de l'escenari en el qual es trobi. En aquest sentit, l'autor també descriu dos processos o fases a tenir en compte segons el nivell de profunditat d'anàlisi al que es vulgui arribar. Un primer nivell identifica i explora el recurs només a nivell d'Autoria, Objectius, Paraules Clau i Públic a qui va destinat. El segon nivell explora de forma sistemàtica la font d'informació, i és el recomanable en un entorn de recerca, on inclouríem tots els atributs, tot i que segons l'entorn o escenari on es faci l'anàlisi els indicadors poden variar.

El resultat de l'estudi de les fonts ofereix un seguit de resultats: les fitxes analítiques i de conclusions on apareix l'*indicador* i el *paràmetre* a analitzar, l'*examen* o pregunta que fa l'analista per a la presa de decisions de si el recurs és o no de qualitat, els *exemples* de recursos que ja compleixen amb aquests indicadors, el *procediment* o orientació de com s'ha d'avaluar l'indicador i la *puntuació* o escala de valors que s'assigni a cada indicador.

Pel que fa a la proposta de *Gonçalves* (2007), es planteja un nou model d'IQ dissenyat per a biblioteques digitals. El model conceptualitza les dimensions centrades en els elements claus que conformen una Biblioteca Digital:

- Catàleg,
- Col·lecció,
- Objectes Digitals,
- Metadades,

- Repositoris i
- Serveis

Cadascun d'aquests sis conceptes contempla un total de 16 dimensions d'IQ (Annex 6.2, taula 4). *Gonçalves* proposa un nou model de Qualitat, a diferència d'altres models d'IQ que estan basats en el contingut, els serveis, els sistemes d'informació, etc. Aquest nou model està basat en les '5S'⁴⁹, és a dir, la IQ es desplega en:

- Grups específics (d'habilitats),
- Estructures,
- Espai,
- Escenaris i
- Societats.

Cadascuna de les dimensions i atributs que proposa per l'avaluació de la IQ estan representades dintre de les '5s'.

En aquest context, un altre concepte el d'objectes digitals (OD) té com a dimensió principal l'*accessibilitat*, i, això fa que vagi lligada a una altra dimensió com són les Metadades, ja que el fet que un document incorpori Metadades serà el que el farà més accessible o identificable per un sistema o un motor d'indexació i recuperació de la informació. Cal destacar que el concepte d'accessibilitat té diverses consideracions en altres entorns web com és l'accés als continguts webs per persones amb discapacitat.

D'altra banda, els OD, es veuen afectats per restriccions tecnològiques, ja que sovint es necessita *hardware* o *software* específic per accedir a un document i, a més, s'han de tenir en compte els canvis de versions i les actualitzacions. La *preservació* és una altra de les característiques que s'ha de tenir present quan parlem d'objectes digitals, així com també les possibles tècniques de preservació digitals existents (migració de continguts, emulació, *wrapping*, *refreshing*, etc.). En relació a la migració de continguts, s'haurà de tenir cura de com s'han creat els documents, és a dir, com el seu origen afecta directament a la preservació i conservació dels documents, depenent de les versions que s'utilitzen en la creació dels documents o objectes digitals aquests poden ser o no recuperats o migrats.

Curiosament dintre de la conceptualització dels OD trobem també les dimensions de *Pertinença* o *Rellevància* que poden estar integrats dins del concepte de *col·lecció* (com passa amb d'altres models d'IQ). *Gonçalves*, com d'altres autors, destaca la *Rellevància* i la *Pertinença* com a factors socials importants perquè estan relacionats amb les necessitats d'informació d'un usuari. La *Pertinença*, dependrà sovint de l'expertesa o coneixement de l'usuari, també de la tasca que realitzi en aquell moment, i finalment d'altres

⁴⁹ 5S: Streams, Structures, Spaces, Scenaris i Societies

factores externs. Quant a la *Rellevància*, en oposició a la *Pertinença*, es defineix com la relació entre la representació d'un document i la representació de la necessitat d'informació o cerca. La *Rellevància* aporta també possibles mesures de tipus estadístic (estimacions o prediccions) que es poden dur a terme amb els propis objectes i per mitjà de la interrogació del sistema, així com també amb els models de vectors que poden representar la distància entre els objectes i la interrogació, i amb altres vectors també coneguts en RI, que permeten mesurar la freqüència d'un terme en un document o la freqüència en una col·lecció, en la mida del document, en l'estructura del document (índexs inversos), etc.

Altres atributs dels OD que cal mesurar són l'Actualització, el *Significat* (transcendència o importància del OD), la *Similaritat* o "*Cerca per exemples*" (la relació que pugui haver entre 2 o més OD), i que puguin ser mesurables els 2 com ja explicava Stvilia (2007), comptant l'ús i la freqüència de paraules dintre del contingut de l'OD, o bé, a partir de les cites i l'estructura o organització interna, i l'ús d'enllaços, etc.

A més d'incloure els darrers atributs com a elements de metadades, també es contempla el propi concepte de Metadades dintre la dimensió de *Precisió*. La *Precisió*, és mesurable a través de la distància del recurs, la propietat o atribut de metadades i el valor, és a dir el rang considerat més correcte entre l'atribut i el recurs. Les *Metadades* també s'inclouen dins del concepte d'estructura, ja que en si mateixes són estructures de dades que donen forma a un document o objecte digital. Altres dimensions que afecten als documents o objectes digitals són la *completesa* existent a totes les metadades que demanen els estàndards, i també la *conformitat* perquè els atributs de metadades tinguin les dades apropiades.

Referent a la *Col·lecció*, *Repositoris* i *Catàleg*, és habitual englobar-les sota el mateix escenari, ja que en tots els casos es requereixen els mateixos atributs: *Completesa* i *Consistència*.

Finalment, i com a darrer concepte característic en les biblioteques digitals trobem els *Serveis*, que fins ara cap model anterior mencionava aquest concepte, i que les dimensions que el representen poden ser internes o externes. Externes, quan aporten informació sobre la satisfacció dels serveis segons l'ús i la percepció de l'usuari, així les dimensions externes són l'*efectivitat* (satisfacció dels serveis) i la *eficiència* (rapidesa, mesurable entre el temps de demanda i temps de resposta). Les dimensions internes, estan relacionades amb la construcció i operacions necessàries per obtenir la funcionalitat requerida. Per exemple, la construcció del sistema, disseny o implementació.

En conclusió aquest model de *Gonçalves*, s'apropa molt més al model d'IQ que es busca en les *subject gateways*, i cal destacar també el paper clau que juguen les Metadades dintre la IQ d'un entorn digital.

6.7. Model IQ en un entorn digital (*Subject Gateways*, Portals, biblioteques digitals)

Finalment, en aquest apartat s'inclou també l'estudi de la IQ en l'entorn digital de les *subject gateways*. Les SGs, de la mateixa manera que les biblioteques digitals inclouen models d'IQ basats en els usuaris i les seves necessitats d'informació, però, no exclusivament, i és per això que s'han inclòs en aquest darrer apartat. De fet, les SGs no representen per si mateixes models conceptuals d'IQ, però, si aporten les seves pròpies perspectives quant a dimensions i criteris d'avaluació de la IQ. A més, tenen les seves pròpies polítiques de qualitat clarament definides i això ens permet veure i comparar els models d'IQ estudiats i dissenyar o millorar un model d'IQ per a les *subject gateways*.

Les dues *subject gateways* incloses en aquest capítol són Intute i el projecte europeu de passarel·la Desire, ja que d'una banda són les SGs més representatives, i d'altra banda són les que tenen accessibles els seus models d'IQ o "Guies d'avaluació d'IQ", com són anomenades per les *Intute Health and Life Sciences Evaluation Guidelines*.

El model d'IQ de la passarel·la britànica Intute està basat en tres categories de factors d'IQ (Contextual, Contingut i Format), cadascuna d'aquestes categories té les seves dimensions i criteris d'avaluació d'IQ.

El **factor Contextual** té a veure amb els conceptes de rellevància i cobertura de la informació, tenint en compte les necessitats dels usuaris. Com a factor inclou tres dimensions:

- *Autoria* i la *Reputació* de les fonts o recursos d'informació. Normalment s'avaluen des de la perspectiva de l'expertesa i coneixement de l'autor, i també des de la reputació de l'organització a la qual pertany o des de la que s'extreu la informació. És a dir, l'autoria va lligada al recurs i farà que sigui un recurs útil o valuós. Si no hi ha autoria la veracitat de la font és baixa o nul·la. S'estudien i avaluen d'on provenen els recursos, si són accessibles en d'altres suports (paper, bases de dades, Cd-roms, en línia, etc.) i si continuen essent interessants.
- *Reputació* com a criteri és més confús perquè sovint la reputació o la popularitat no ha d'anar lligada amb la qualitat, i tampoc l'expertesa d'un autor ha d'implacar necessàriament la qualitat d'un document. Així doncs, necessitem la conjunció de tots els factors per arribar a la IQ.

- *Comparació de recursos.* Es contempla com un factor que pot ajudar a avaluar la IQ ja que molts aspectes que afecten a la qualitat poden ser comparats entre diferents materials per crear un propi judici sobre el valor i utilitat d'un recurs determinat. Es poden avaluar aspectes com l'accessibilitat, formats, versions d'un mateix recurs, comparativa de costos, el valor entre materials similars, etc.

La següent categoria que es contempla com a factor d'IQ és el **Contingut**, en molts altres models vindria lligada al **entorn Contextual**. Inclou les dimensions de *Cobertura, Precisió i Actualització*:

Cobertura. S'analitza si la cobertura de la matèria d'un recurs d'informació s'adequa a la rellevància dintre del sistema d'informació. Referent a la cobertura els factors que influencien a la qualitat són la *comprensibilitat* del recurs dintre de l'àrea, el rang *de la cobertura* de les diferents matèries dins d'una mateixa àrea, també si un recurs cobreix la cobertura temàtica adequadament, si la cobertura retrospectiva de la matèria és arxivada o no, la relació del detall d'un recurs i si s'ajusta al context del sistema d'informació, i així com veure finalment si els enllaços que provenen d'un mateix recurs i la seva cobertura poden ser de valor afegit.

- *Precisió,* En termes generals es refereix a si la informació és editada correctament o no. També cal tenir present que hi ha factors a avaluar per saber si hi ha IQ:
 - Si la informació editada està referenciada,
 - Si la informació té unes bases de recerca,
 - Si la informació és suportada per un editor, etc.,
 - Si els materials són acurats, si no hi ha errors tipogràfics o gramaticals, etc.
 - Si són possibles les avaluacions dels recursos per altres autors
 - Si té un aspecte professional, etc.
- *Actualització.* Fa referència a la regularitat de la seva actualització, si s'actualitza freqüentment, si la informació actualitzada és apropiada o no, etc.

La darrera categoria o factor d'IQ que contempla Intute, és el **Format**, que sobretot fa referència a la seva accessibilitat, facilitat d'ús i el disseny. Podria correspondre's a la dimensió **Representacional** (Wang 1996 i Stvilia 2007).

Les dimensions que es representen en aquest context com a criteris d'IQ són:

- *Accessibilitat.* Com afecta a l'accessibilitat del recurs i si aquest és accessible constantment. També hi ha d'altres factors importants com serien el de la fiabilitat del servidor, si hi ha alternatives d'accés al document, si hi ha disponibilitat del document en diverses llengües, si hi ha algun tipus de restricció d'accés (geogràfic, subscripcions, tipus de

- software o hardware etc.), i finalment si el document té algun tipus de llicència o copyright.
- *Disseny i Etiquetatge*. En aquest apartat el nucli és sobretot la interfície i la seva usabilitat.
 - Si la navegació és assistida amb usos d'hipervincles i si hi ha un ús constant en la pantalla. La navegació entre les pàgines i els seus enllaços ha de ser fàcil i intuïtiva.
 - Si la informació es presenta adequadament (de forma professional), i si els recursos són descrits adequadament.
 - Si hi ha la possibilitat d'imprimir els recursos
 - Si hi ha ajuda per trobar la informació al web i si aquesta és efectiva: mapes web, índexs, sistemes de menús o facilitat en la recerca
 - Altres aspectes com l'ús d'imatges, si és adequat o se'n fa un ús decoratiu en excés, etc.
 - Facilitat d'ús. Aquest criteri també va lligat a l'accessibilitat i a la disponibilitat. Facilitat tant en l'accés com en la navegació del recurs i també en la localització de la informació que requereix l'usuari. Si l'usuari disposa de llistes de distribució, grups de suport, etc.

Com s'ha pogut veure en el cas de les SGs, la IQ va lligada a la prestació de serveis, i per això un servei és un valor afegit respecte del sistema.

El projecte *Desire*, contempla també una guia-manual dels diferents processos de control de qualitat que segueixen les *subject gateways*. En aquestes guies es fa sobretot èmfasi en l'esforç humà que hi ha al darrera d'aquests processos principalment quan fa referència a la selecció de recursos ja que no són processos automatitzats.

Els factors claus que engloben el model d'IQ de *Desire* es centren en els processos de la selecció de recursos d'Internet que afecten els usuaris, els propis recursos d'informació i els serveis. A partir d'aquí es construeixen cinc categories o criteris d'IQ:

- Criteri de **Competència** (considerant l'usuari). Inclou les dimensions de *Cobertura* i *Accés*. La cobertura ve marcada per l'audiència, per la tipologia d'usuaris, però, també contempla aspectes com la tipologia de recursos que es vulguin incloure, com la perdurabilitat del recurs, etc. El tipus d'accés que es vol incloure i les característiques quant a costos, a la tecnologia, si requereix registre o no, la seguretat, o contemplar altres característiques com restriccions geogràfiques o de llengua. Quin nivell de descripció tindran aquests recursos, si s'utilitzaran metadades o no, etc.
- Criteri de **Contingut**. Es contempen les dimensions ja conegudes en la literatura científica i com la *Validitat* segons la qual no hi ha d'haver pèrdues d'informació, també si hi ha bibliografia, i d'altra informació addicional. *Auditoria/Reputació del recurs*, Substantivitat que comporta si

hi ha informació de valor afegit com adreces de contactes o si la informació és a text complert. *Precisió* segons la qual no hi ha d'haver errors de dades ni de tipus ortogràfic o tipogràfic, i, també si la informació conté bibliografia. *Comprensió i Univocitat* de la informació, *Composició i Organització* de la informació, *Adequació* i *Actualització*.

- Criteri de **Forma** o avaluació del mitjà. On trobem els aspectes de *facilitat en la navegació*, *Suport a l'usuari* (si hi ha guies, FAQ's, etc.), ús d'*estàndards* (HTML, estàndards multimèdia, etc.), ús apropiat de tecnologia, aspectes estètics com per exemple si el sistema és amigable, i també aspectes d'integritat de la informació (sobretot en relació al temps de resposta del sistema i l'actualització dels recursos).

Criteri de **Procés** o avaluació del sistema. *Cobertura* de la col·lecció a partir de la qual cal detectar duplicats i analitzar la navegació per la col·lecció per detectar si hi ha àrees amb menys cobertura temàtica, *accessibilitat dels recursos d'Internet*, *accessibilitat dels recursos de la Biblioteca*.

Les dimensions de la IQ en les passarel·les són representades en conceptes diferents, però són coincidents amb la resta de criteris exposats. A diferència dels altres models d'IQ que hem exposat, les SGs, integren els *Serveis* com a part del model d'IQ, i fan incidència com molts altres autors en la gestió de la col·lecció i tot el que comporta, com són les dimensions *Intrínseca*, *Contextual* i *Representacional*.

Les SGs en si són models de qualitat i, tot i que en les "guies de qualitat" només tracten la IQ centrada en la "selecció de recursos", com a sistemes d'informació de qualitat, el que les diferencia de la resta de models d'IQ és la seva arquitectura de la informació. Desire, ens parla del gran esforç humà que hi ha al darrera d'aquests sistemes, i, de fet, les SGs conformen el seu propi sistema d'IQ ja que no només contribueixen a la qualitat en la selecció de recursos, sinó també en la riquesa de la descripció i estructura dels recursos amb l'ús d'alts estàndards de control de qualitat com són els vocabularis controlats i l'ús de metadades (preferentment el Dublin Core).

L'ús de vocabularis controlats (classificacions, tesaurs o encapçalaments de matèries) tradicionalment s'han utilitzat per a la indexació dels documents, donant IQ des de les dimensions de rellevància i precisió en la recuperació de la informació. Les SGs, a més a més de l'ús dels vocabularis controlats en la indexació també utilitzen els vocabularis en la navegació en dels seus sistemes d'informació. Sobretot, és més extensible l'ús de les classificacions com a sistemes de navegació ja que d'una banda el seu ús està més estandarditzat (CDU i DDC), i d'altra banda la seva estructura jeràrquica i numèrica és molt més senzilla alhora d'automatitzar els processos d'indexació o navegació (Golub, K., 2006): categorització de textos, *clustering* o agrupació de

documents, classificació de documents, implicacions en la navegació o ús de software específic com *ROADS*.

En el cas específic de les SGs, els llenguatges documentals són fonamentals perquè serveixen com a esquemes de mapeig entre diferents vocabularis controlats que poden integrar una SG⁵⁰, i per tant, han afavorit la creació d'una nova àrea d'experimentació i estudi sobre la interoperabilitat entre sistemes sobre la base de la indexació i la navegació creuada (Veure Capítol 3).

6.8. Anàlisi i comparativa de models d'IQ

L'anàlisi dels diferents models d'IQ seleccionats [Annex 6.2] ens porta a determinar que els projectes i la literatura especialitzada han identificat 70 dimensions que representen els criteris d'avaluació de la IQ [Annex 6.3]. Dimensions que cal tenir present si el que es vol és desenvolupar un model propi d'IQ aplicat a les SGs, tal com ens proposem en aquest treball.

Tanmateix, aquesta representació de dimensions/criteris de la IQ és molt extensa i no és globalment útil tenint en compte que a nivell de costos no seria assumible utilitzar tots els valors com a indicadors d'IQ. Així com resulta difícil seguir un únic model perquè cada autor utilitza conceptes diferents per atributs similars o idèntics.

Per tant, hem optat per seleccionar les dimensions més representatives dels models IQ, s'ha estudiat cada atribut i s'ha determinat un valor en funció de la freqüència que apareix en cadascun dels models d'IQ estudiats, tenint en compte que cada autor ja havia fet prèviament una selecció i, per això en els models d'IQ ja hi ha representades les dimensions més importants. Des del nostre punt de vista les dimensions (ordenades de major a menor importància) que ha de representar tot model d'IQ són:

- Precisió,
- Accessibilitat,
- Completesa,
- Actualització,
- Rellevància,
- Autoria/Reputació i
- Seguretat.

A més a més, es recullen d'altres dimensions importants extretes d'entre conceptes similars o idèntics que s'agrupen en les següents dimensions:

- Comprensió,
- Estructura,
- Representació i
- Integritat.

⁵⁰ Consulteu el capítol 3, dedicat a la Indexació i la Navegació creuades.

A continuació es presenten gràficament quines són les dimensions d'IQ més rellevants segons els models d'IQ existents i que hem analitzat. En el gràfic es mostra de esquerra a dreta, el grau descendent del valor d'importància dels criteris d'IQ extrets dels diferents estudis i de les *subject gateways* analitzades. Aquest valor ve donat per la suma dels indicadors coincidents o dimensions coincidents en cadascun dels models d'IQ, tenint sempre en compte que els atributs representants són també els més importants en els models d'IQ estudiats.

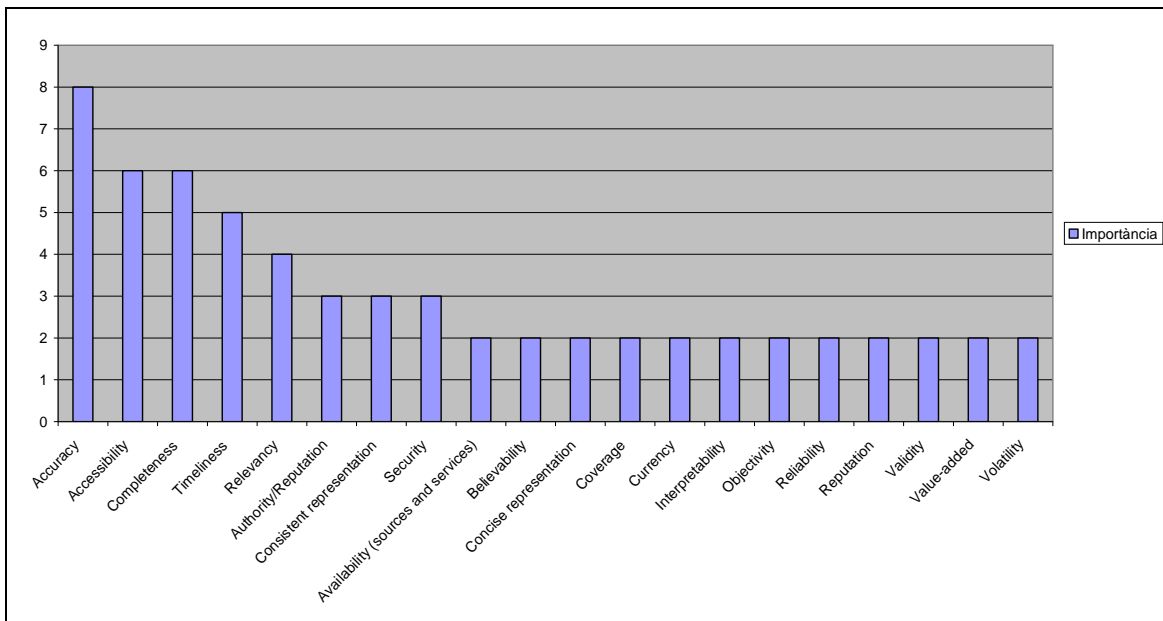


Figura 92. Atributs d'IQ rellevants (Font: Elaboració pròpia)

El resultat d'aquestes dimensions ens porten a representar una estructura clara de quins són els factors claus que es veuen afectats des de la perspectiva d'una SG. I l'estructura del model d'IQ en les passarel·les temàtiques queda agrupada de la següent forma: *l'Usuari, la Col·lecció, el Sistema i els Serveis*.

INTUTE	DESIRE	Model IQ Proposat [Propietat]
Contextual	Usuari	1.Usuari [intrínseca]
Contingut	Contingut	2. Col·lecció [Contextual]
Format	Forma/Mitjà	
	Procés	3. Sistema [Representacional]
	Col·lecció/Serveis	4. Serveis [Contextual]

Taula 27. Dimensions d'IQ bàsiques per a una SG (Font: Elaboració pròpia)

6.9. Model IQ proposat per a un sistema d'informació de qualitat distribuït

Després d'estudiar tots els models d'IQ existents i analitzar els valors dels indicadors que cada dimensió de la IQ té en cada model i també té en cada metodologia, i s'ha optat per contextualitzar la nostra proposta de Model IQ en funció de tres factors determinants pel sistema que es dissenya en aquest treball:

1. Entorn digital: tipus passarel·la temàtica
2. Context informacional: Ciències aquàtiques i/o ciències afins
3. Audiència: investigadors, docents, i professionals de la informació especialitzada.

De tots els models i mètodes d'IQ estudiats, en aquest estudi, han influït especialment els models d'IQ basats en entorns digitals com Intute, Desire i Stivilia (2007), perquè són els que responen als objectius d'aquest projecte.

El Model d'IQ que es proposa per al prototipus de sistema d'informació de qualitat especialitzat en ciències aquàtiques és un model híbrid amb una doble tipologia de característiques/atributs: fixes i variables. Aquesta híbrides en el model d'IQ es basa en la variabilitat del web, i en la importància i els factors externs que cada atribut pot tenir respecte de les quatre dimensions fonamentals triades: Usuari, Col·lecció, Sistema, i Serveis.

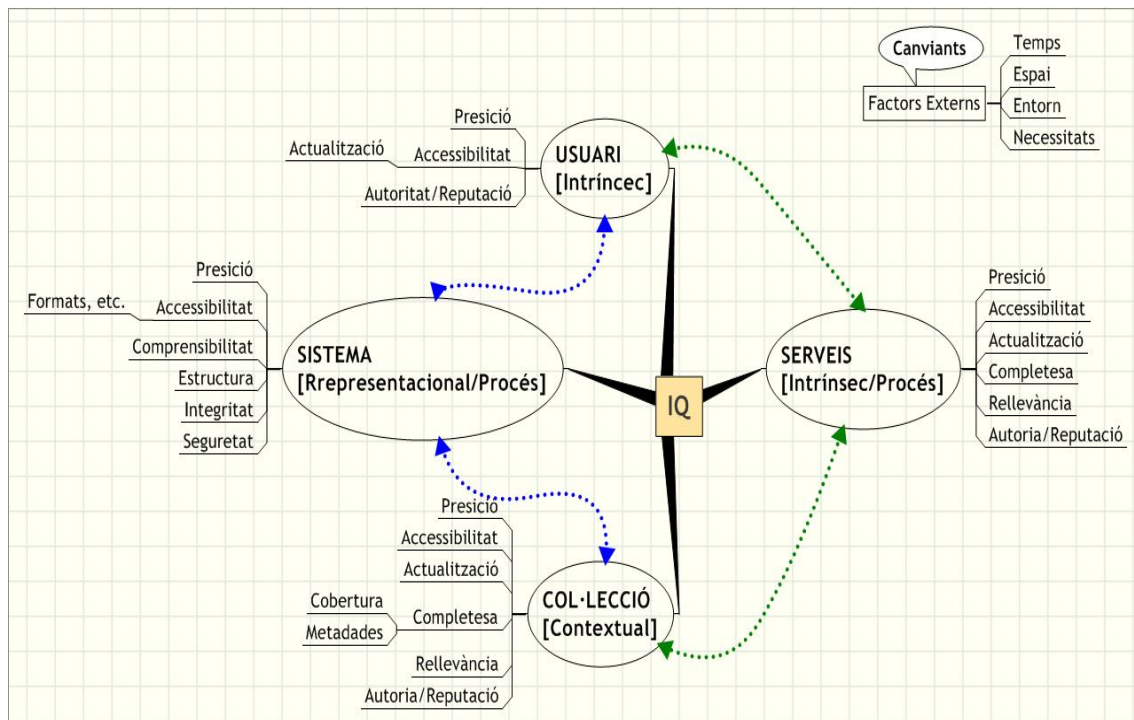


Figura 93. Model conceptual d'IQ proposat per a una SG (Font: Elaboració pròpia)

A continuació es descriuen tots els atributs del model d'IQ proposat (gràfic 2), la seva definició, i tipologia (si és un atribut variable o fix). Tots ells es classifiquen sota una dimensió d'IQ: usuari, col·lecció, sistema i serveis.

Dimensió usuari		
Atributs	Definició	Tipus
Precisió (Miller, 1996)	Una informació precisa o exacta reflexa que s'adequa a la realitat del usuari. La informació s'adequa a les necessitats del usuari que té en un moment determinat i poden variar.	Variable: Les necessitats de l'usuari varien i per tant, els nivells de precisió requerits també. Exemple: Usuari 1: "Algae + biodiesel"; Usuari 2: "microalgae + biodiesel"
Accessibilitat (Actualització) (Miller, 1996)	Poder accedir a un recurs quant sigui necessari (els canvis que succeeixin com per exemple de format o versió no s'han de notar. Dimensió lligada a l'oportunitat de les dades (dades obsoletes).	Fix: Tot i que els recursos i el seu accés, per l'usuari si un recurs no és accessible no serveix.
Autoria/Reputació (Wang, 1996, ja ho menciona com un atribut a tenir en compte però no apareix com a element clau fins l'aparició de models IQ de sistemes o documents)	Determina l'origen de les dades, d'on provenen les dades. Com a dimensió, és la mesura en què la informació està molt ben considerada quant a la seva font d'informació o contingut (Zhu 2007). També es pot considerar com a mesura de la posició d'una entitat o objecte d'informació en un àmbit o estructura cultural o d'activitat (Stvilia, 2007)	Variable: Com a indicador social és canviant i pot respondre a diferents criteris: expertesa dels usuaris, Indicators Científics (Factor d'Impacte) o altres valors determinats per les comunitats científiques.

Taula 28. Dimensió usuari

Dimensió Col·lecció: <i>Intute Evaluation Guidelines</i> i Stvilia (2007)		
Atributs	Definició	Tipus
Precisió	Si la col·lecció integrada en el sistema respon a les necessitats dels usuaris amb exactitud o a les preguntes que aquest realitza. Però hi ha altres factors claus com: si la informació és avaluada i revisada. Si no hi ha errors gramaticals o tipogràfics (Correctesa).	Variable: Des del punt de vista de la Recuperació de la Informació es un atribut que varia segons la perspectiva del usuari. Per tant, és difícil que sigui un atribut fix. Un exemple divergent de la perspectiva del usuari: "Microalgae Culture" i quan entrem al recurs és

		una pàgina comercial de compra de cosmètics
Accessibilitat	Facilitat d'accedir a una font. Si aquest està disponible constantment o no. Subscripcions per visualitzar informació. O altres aspectes relacionats amb la usabilitat (imatges sense llegenda), etc.	Variable: La col·lecció, s'ha d'actualitzar amb periodicitat, tot i que no s'han de trobar documents no accessibles en el sistema poden haver canvis. Per exemple: un accés obert canvia a accés sota subscripció.
Actualització	És la freqüència i regularitat d'actualització de la col·lecció. L'actualització és un factor clau per a la IQ ja que si els materials són obsolets la reputació del sistema serà negativa.	Variable: és canviant per definició. Per exemple, una plana web que canvia d'adreça URL (canvia l'adreça i per tant, s'hauran d'anar revisant els enllaços trencats, etc.). Implica l'adquisició de fons i per tant, la col·lecció pot incrementar a diari.
Completesa i (Cobertura Metadades)	És el factor clau que determinarà la col·lecció. Quant a la IQ s'inclouen factors diversos com: amplitud del recurs, exhaustivitat de la col·lecció i especificitat, cobertura retrospectiva, el grau de descripció del recurs (metadades), si hi ha avaluació dels recursos, etc.	Fix: per assegurar la IQ en aquesta dimensió és necessari marcar uns criteris fixos que cal assolir. Per exemple: quant a la descripció de la col·lecció usar estàndards com Dublin Core (15 elements bàsics). Definir la cobertura temàtica i geogràfica, etc. Això no vol dir que hi hagi recursos que no estiguin complets (manca editor, autor, data, etc.). O Redundància (duplicats).
Rellevància (Miller 1996)	La informació satisfà a les necessitats dels usuaris. És un dels components claus per a la Qualitat de la Informació.	Fix/Variable: Atribut mesurable segons la perspectiva de l'usuari. Una mateixa col·lecció pot ser rellevant o no per a unes necessitats d'informació diferents. Tot i que es podrien assignar criteris. Per exemple: rellevància segons l'autoria o reputació de la

		font o per precisió (es compleixen les necessitats dels usuaris), per exemple, la col·lecció correspon als descriptors definits pels usuaris.
Autoria/Reputació	<p>Mesurar la posició d'una entitat o objecte d'informació en un àmbit o estructura cultural o d'activitat.</p> <p>L'autoritat d'una font d'informació es basa principalment en una avaluació dels coneixements i l'experiència de l'autor (s), i si estan qualificats per subministrar la informació, així com la reputació de l'organització de la qual emana la informació.</p>	<p>Fix/Variable: és un indicador variable (factor social/cultural) però si es poden establir polítiques d'IQ per a la selecció de la col·lecció segons 'autories/reputació' que facin aquest atribut fix. Tot i que a vegades la reputació ve donada per la popularitat i això el fa un atribut variable.</p> <p>Exemple de criteris de selecció: "selecció de recursos científics provinents d'universitats, associacions professionals, organismes governamentals o organismes internacionals que siguin ben coneguts per les diferents comunitats científiques".</p>

Taula 29. Taula Dimensió Col·lecció

La perspectiva de la dimensió del Sistema sempre implica atributs variables ja que la seva avaluació s'enfoca cap a la recuperació de la informació i, aquesta sempre recau en la perspectiva del usuari.

Dimensió Sistema: Naumann (2000), Stvilia (2007)		
Atributs	Definició	Tipus
Precisió	Valor que es dona segons la correspondència de la correctesa de la font i el valor total que es doni al conjunt dels recursos recuperats. Qualitat de les dades (en oposició a informació de qualitat), taxa d'error, exactitud, integritat i precisió.	Variable: Cada equació de cerca pot donar resultats de precisió diferents. L'atribut varia segons la perspectiva del usuari. Exemple: "microalgae biodiesel" i recuperem també recursos que no volem com: " biofuel from algae "
Accessibilitat	Disponibilitat del temps en	Variable: es necessita

	que un recurs està al web. Cal tenir en compte els conceptes relacionats: recuperabilitat, fiabilitat, perdurabilitat i actualització de les fonts (formats i/o versions), etc.	una avaluació constant per detectar enllaços trencats (recursos que varien o desapareixen), així com les versions i formats dels documents, etc.
Comprensibilitat	Grau en què la informació pugui ser compresa per l'usuari. Per exemple: incloure documents o recursos amb llengua xinesa. Haurem de veure si la comunitat científica a la que ens dirigim coneix el xinès.	Variable: Atribut que varia segons l'usuari, els seus coneixements i la seva expertesa.
Estructura	Nivell pel qual l'estructura de la informació coincideix en la pròpia informació. Si la informació representada té coherència estructural, adequació, ocurrència i precisió.	Fix: l'estructura amb la que es representa la informació ha de correspondre amb el contingut. Per exemple: que una referència bibliogràfica correspongui amb la informació (autors, títol, font, etc.).
Integritat	Nivell pel qual l'estructura de la informació s'ajusta a altres fonts d'informació. Com per exemple, la homogeneïtat de la informació, coherència semàntica, coherència de valor, portabilitat i compatibilitat.	Fix: Tota la informació hauria de ser representada per una mateixa estructura com per exemple la utilització de metadades Dublin Core per a la descripció dels recursos. Així doncs, s'homogeneïtza el contingut del sistema, es dóna coherència semàntica i interoperabilitat entre els diferents recursos.
Seguretat	El nivell amb el que la informació de l'accés d'un usuari a la font d'informació consultada és privatitzada.	Fix: la privacitat i seguretat d'accés a la informació per part de l'usuari ha de ser un atribut fix. És un valor per al propi sistema.

Taula 30. Dimensió Sistema

Dimensió Serveis: els atributs definits en la dimensió dels serveis (Stvilia, 2007) són coincidents amb els de la col·lecció i per tant en la seva definició, ja que són atributs que han de correspondre a la política del sistema d'informació (per exemple una *subject gateway*) i a la comunitat que hi ha al darrere. Gonçalves

(2007), introdueix el concepte serveis com a mesura d'IQ, a través dels serveis es poden conèixer millor els usuaris (anàlisi de logs). A nivell de la definició dels atributs, Gonçalves se centra només en les característiques que afecten als objectes digitals, i en el cas de les subject gateway on Desire és qui més aprofundeix amb aquesta dimensió, no es defineixen atributs. Per això s'opta per definir els conceptes des del punt de vista d'Stilia on la seva dimensió contextual contempla tant als recursos com les activitats (serveis).

Dimensió Serveis: Stvia (2007)		
Atributs	Definició	Tipus
Precisió	El grau en què un objecte d'informació es representa correctament un altre context informacional, procés o fenomen en el context particular d'una activitat o cultura.	Variable: per exemple quant un usuari selecciona un recurs que li és de qualitat, el servei de cerca ofereixi buscar en "altres recursos o documents similars"
Accessibilitat	Velocitat i facilitat de localització i obtenció d'una informació relativa a un objecte o activitat particular. Cal afegir també el tipus d'accés a la informació, etc. (Intute).	Variable: L'accessibilitat anirà lligada al tipus de servei, de manera que variarà. Per exemple: accés amb diferents formats (html, RRSS, .RIS, etc.). Facilitat d'accés a través d'aplicacions necessàries per a la visualització (Java Virtual Machine o Adobe Acrobat Reader)
Actualització	És el context amb el qual la informació és nova o informativa en un context particular (I que no hi hagi redundància). I el temps en que la informació roman vàlida en el context d'una activitat determinada.	Variable: Dependrà del servei. Per exemple, quant a la temporalitat és refereix, un butlletí temàtic podrà tenir una periodicitat mensual i ser accessible el seu històric. I una subscripció a notícies via RSS l'actualització i enviament de la informació serà immediata a la seva producció.
Completesa	El context amb el que una informació, recurs o servei, coincideix amb la necessitat de l'usuari o	Variable: dependrà de la perspectiva de l'usuari Per exemple l'exportació de resultats: format cita

	activitat (referent a la precisió i completesa)	bibliogràfica (formats diferents que poden adequar-se o no a les necessitats del usuari), representació gràfica, <i>clustering</i> , etc.
Rellevància	Quant la informació obtinguda s'aplica al context.	Variable: Perspectiva de l'usuari. També cada servei pot respondre a diversos contextos: subscripció a novetats del portal (l'usuari pot només estar interessat en documents i no en notícies)
Reputació	El nivell de la reputació d'una informació o document, ve donada per una comunitat o cultura.	Variable: factor determinat per l'usuari. No tots els serveis poden tenir la mateixa reputació. Aquest atribut es podrà mesurar per l'anàlisi de logs (consulta a les bases de dades) o bé mitjançant puntuacions que els mateixos usuaris posin als serveis

Taula 31. Dimensió Serveis

6.10. Conclusions

L'estudi de la revisió bibliogràfica sobre l'estat de l'art de la Informació de Qualitat en sistemes d'informació distribuïts, conclou que un model o guia d'IQ d'un sistema d'informació ha d'incloure tant els aspectes teòrics com els pràctics, ha d'ésser sistemàtic i concís amb els tipus de criteris que posteriorment han d'avaluar la informació, però, també s'ha de ser realista en la quantitat de criteris que es poden avaluar, calculant els costos que comporta l'aplicació d'un model d'IQ.

Cal tenir present que si els criteris d'avaluació no són assumibles el model d'IQ no funcionarà. Primerament, s'ha de tenir en compte que són necessàries com a mínim dues fases en tot model d'IQ (Miller, 1996): assenyalar quins atributs són importants segons la tipologia d'usuaris, i determinar com aquests atributs afecten a les necessitats d'informació dels usuaris.

Posteriorment a l'aplicació del model, l'avaluació de la IQ ha de ser constant, ja que el model d'IQ que es necessita s'emmarca dintre d'un entorn digital i canviant com és Internet, on tant les necessitats d'informació com els atributs o dimensions d'IQ també aniran canviant i és necessari preveure la revisió dels models tan aviat com s'identifiquin els factors de canvi.

En el model s'han de contemplar o preveure alguns esquemes que analitzin i solucionin els problemes d'IQ que es detectin. I que permeti desenvolupar les bases per mesurar els criteris d'IQ, sempre preveient el canvi de criteris segons els canvis que es produeixin en l'entorn. Així doncs, cal oferir un mapa conceptual que pugui ser utilitzat per estructurar una varietat d'aproximacions, teories i fenòmens relacionals amb la IQ.

Altres aspectes a destacar són la importància de les dimensions i criteris d'avaluació ja que en tota la revisió dels models i mètodes s'observa que el grau de valor d'una dimensió canvia segons la perspectiva de l'autor, i que sovint, aquesta va lligada a l'aplicació o ús al qual va dirigit el model d'IQ. És a dir, si el model d'IQ que es dissenya s'enfoca a una organització, pren més rellevància l'usuari i les seves necessitats d'informació i en conseqüència els criteris d'IQ que afecten al usuari (*precisió, accessibilitat, cobertura, rellevància, autoria, etc.*). Si el model d'IQ va enfocat a un sistema d'informació digital els criteris d'IQ poden ser múltiples tant pel que fa a la seva *audiència*, l'*accessibilitat* del sistema (accés i representació de les dades), la *col·lecció* (cobertura, presentació, formats, accessibilitat, etc), la *recuperació de la informació* com pel propi *disseny* del sistema. (Veure Annex 6.2)

L'anàlisi i comparativa dels models d'IQ ha servit per veure quines dimensions són clarament representatives i necessàries per tot model d'IQ. Per exemple, l'*accessibilitat*, tot i que en entorns no digitals pot semblar un criteri no prioritari és, però, el criteri més important puix que sense aquesta dimensió la resta de d'avaluacions d'IQ no es poden dur a terme. Evidentment, si un usuari no té accés a la informació, la resta de criteris no seran avaluable perquè l'usuari haurà determinat que aquell document no té IQ. A més, l'*accessibilitat* depèn de molts d'altres factors com són l'actualització de les fonts o recursos d'informació i la freqüència amb que aquestes s'actualitzin. Això significa que és una dimensió que va sempre unida als objectes, documents o recursos d'informació que s'emmagatzemen en un sistema d'informació, i per tant, l'altre factor clau serà la *representació* d'aquests i el tipus de formats, versions i altres tipus de característiques tecnològiques que puguin afectar a la seva accessibilitat. De fet, aquest és un dels factors clau que tenen en compte els usuaris per avaluar la IQ d'un sistema o entorn d'informació. Després hi haurà altres factors associats com el cost, el tipus de disponibilitat (sota subscripció, *open access*, etc), l'autoria o reputació de la font i aquest factor a vegades depèn de l'expertesa de l'usuari per ser més o menys valorat.

Una altra dimensió a tenir present és la *completesa* de la col·lecció que, a més de ser una dimensió important per la majoria de models d'IQ, és on es té en compte l'anàlisi de les *metadades* (criteri i mesura d'avaluació d'IQ). Hi ha autors que uneixen la *completesa* a la dimensió de la *rellevància* (*Wang, Bovee i Miller*). La rellevància de fet, hauria de ser un factor clau en tot model d'IQ, perquè si la cobertura i rellevància de la col·lecció no és l'adequada per l'usuari

o per l'audiència a la que va dirigida el sistema o entorn d'informació, el model d'IQ es trenca ja que no compleix la premissa de tot servei i sistema d'informació i documentació de cobrir les necessitats dels seus usuaris que són el principal motiu de la seva creació.

D'altra banda, tot i que no s'afirma de manera explícita en cap estudi analitzat, les *Metadades* haurien de ser incloses en tot sistema d'informació i documentació digital i per tant, també han d'estar incloses dintre el model d'IQ. Les Metadades aporten un gran valor afegit en criteris d'avaluació d'IQ perquè ens donen informació molt valuosa com l'autoria, la matèria o la data del document entre d'altres. Serveixen per identificar i recuperar amb major precisió el contingut dels documents i els recursos digitals. A més, són fàcilment parametrizables i serveixen com a mètrica d'IQ per a la dimensió de "Completesa" dins d'una col·lecció. Però, el motiu principal és que representen un estàndard recolzat internacionalment, són flexibles, són una de les característiques principals de la interoperabilitat semàntica i per tant són necessàries per a l'intercanvi d'informació entre sistemes d'informació i documentació digitals.

CAPÍTOL 7. SISTEMES DE MAPEIG DE VOCABULARIS CONTROLATS EN CIÈNCIES AQUÀTIQUES

7.1. Introducció

En els anteriors capítols, els vocabularis controlats o sistemes d'organització del coneixement s'han definit com l'eina fonamental utilitzada per a la indexació i recuperació de la informació (RI) en qualsevol sistema d'informació (catàleg, base de dades, portals web, etc.), tant en els sistemes d'informació més tradicionals com més actuals o semàntics. En l'estudi de casos de diferents tipus de sistemes (vegeu capítol 2) s'observà la necessitat de l'ús de diversos vocabularis controlats per indexar el contingut de col·leccions (multidisciplinàries i interdisciplinàries), o bé, de col·leccions provinents de varis proveïdors (dominis similars però amb continguts i/o amb vocabularis diferents). És a dir, els vocabularis controlats poden funcionar com un mecanisme de connexió entre diferents models conceptuals d'organització del coneixement i, per fer-ho, la solució més estesa inclosa en aquest estudi és l'alineació de vocabularis (vegeu capítols 3 i 5).

L'alineació de vocabularis utilitza principalment amb quatre objectius o aplicacions:

- **Reindexar col·leccions.** Suport a la indexació de documents amb un Vocabulari "B" sobre la base de la indexació existent amb un "Vocabulari A", o viceversa).
- **Cerca de conceptes** (basats en els vocabularis que indexen col·leccions heterogènies). Suport a la recuperació de documents indexats amb un "Vocabulari A" amb consultes que utilitzen conceptes d'un "Vocabulari B", o viceversa.
- **Navegació** (mitjançant vocabularis). Suport a l'exploració dels espais conceptuals a través de vocabularis per accedir a elements de la col·lecció indexada amb els conceptes seleccionats.
- **Fusió de Vocabularis.** Donar suport a la construcció d'un nou vocabulari que abasta tant el "Vocabulari A" com el "Vocabulari B", o bé, la integració d'un vocabulari en un altre (com una extensió o satèl·lit de l'altre vocabulari).

Actualment hi ha molts casos exemplars d'alineació de vocabularis que cobreixen aquests objectius presentats. Cal destacar els següents exemples:

Enriquiment i descobriment d'informació: sistemes que han sol·licitat vocabularis d'origen o destinació que són	Exemple 1: Civil War Data 150 (EUA, 2011-). Projecte de col·laboració (biblioteques, arxius, museus, etc.) per compartir i connectar dades nord-americanes relacionades amb la guerra
--	--

part dels esforços d'alineació (DBPedia)	<p>civil a través de les institucions locals, estatals i federals. Utilitza <i>linked data</i> per crear connexions basades en diferents vocabularis com per exemple un de la guerra civil americana i una ontologia específica que han creat i estableixen correspondències (<i>crosswalk</i>) sobre regiments, batalles, camps de batalla, oficials, soldats i mariners amb altres vocabularis que els complementen com Freebase, DBPedia i LCSH, entre d'altres. Exemple d'Alineaments. En CivilWar 150 es pot consultar informació sobre un determinat lloc, regiment, batalla, o funcionari. Els resultats són dades estructurades vinculades amb els documents d'origen en les institucions que contribueixen per a la investigació (biblioteques, arxius, etc.)</p> <p>Exemple 2: MisMuseos (Madrid, 2013?). Col·lecció semàntica amb 15.000 obres d'art de set museus públics espanyols. L'aplicació, ja en funcionament, permet a l'usuari trobar amb precisió obres d'art i artistes en funció d'època, estils, escola, tècnica. Url: http://www.mismuseos.net/comunidad/museos (Software GNOSS)</p>
<p>Millora de vocabulari i la reutilització tant per estendre altres vocabularis de valor com per assentar les bases base per a la creació de vocabularis de valor nous. Per exemple: enriquiment de col·leccions, o bé, creació de nous sistemes d'informació: projecte ASEG (Reverté, C. i Sebastià-Salat, M., 2009b).</p>	<p>Exemple 3: HILT (High-Level Thesaurus). Servei de Cerca i Navegació creuada entre diferents vocabularis (generals LCSH, UNESCO, DDC, i AAT, o específics com MeSH). Diferents projectes com INTUTE, o EDINA.</p> <p>Exemple 4: LD Service. Library of Congress Vocabularies & Authorities: LD, mitjançant HTTP, URI i OWL, integra diferents vocabularis controlats i els utilitza per al descobriment i reutilització de recursos web de la LC.</p>
<p>Publicació, descobriment i manteniment de les eines o serveis d'alineació de vocabularis.</p>	

Taula 32. Exemples d'alineaments de vocabularis en sistemes d'informació

És evident que no estem davant d'una nova tecnologia, però sí d'una tecnologia amb gran potencial i en vies de desenvolupament, per tant, és necessari que s'estudiï detingudament i sota diferents punts de vista perquè aportï una major eficàcia als sistemes de recuperació de la informació de qualitat.

Un dels objectius d'aquest capítol és conèixer en profunditat algunes de les tecnologies que permeten mapejar ontologies i, a la vegada, realitzar un estudi

experimental sobre l'alineació i el mapeig automàtic de diferents vocabularis controlats en relació amb el camp de les ciències aquàtiques i marines, i estudiar, d'aquesta manera, els avantatges, inconvenients i la sostenibilitat d'aquest tipus de tecnologies.

En aquest estudi experimental de mapeig de vocabularis (període setembre 2009-juny 2011) es seleccionen els més estesos de les biblioteques especialitzades de ciències aquàtiques i marines, ASFA, AGROVOC, GEMET i NBII (presentats en capítol 5), tots ells són tesaurus, tipologia d'organització del coneixement considerada com a un tipus de vocabulari lleuger (*light*) i ontologia simple.

Abans de la realització dels experiments de mapeig de vocabularis, un estudi previ d'alineació manual dels quatre vocabularis (Annex 8.1 i 8.1.1) ja va descartar l'ús del vocabulari NBII per als experiments d'alineacions, perquè NBII utilitzava un format de representació de les dades incompatible amb la resta de vocabularis estudiats, és a dir, no tenia representació en un format estàndard i semàntic (OWL o SKOS).

7.2. Característiques d'alineació de vocabularis: vocabularis controlats versus ontologies

Com s'exposa en el capítol d'estudis de casos (capítol 2), i en el d'interoperabilitat semàntica (capítol 5), les ontologies són els nous vocabularis controlats emprats en els sistemes d'informació que milloren la qualitat de la navegació i recuperació de la informació (cerca i indexació).

Els vocabularis controlats tradicionals en comparació amb les ontologies tenen una cobertura i llenguatge limitats, codificació i semàntica simples (conceptes i relacions), problemes severs de manteniment per la seva estructura i la lectura per màquina és molt reduïda. Un domini o camp del coneixement es representa bàsicament per conceptes i relacions jeràrquiques i/o associatives.

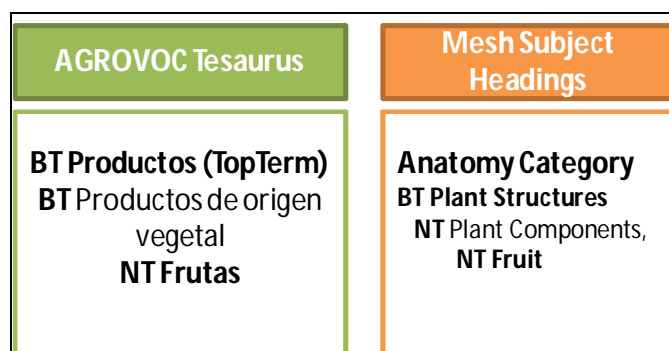


Figura 94. Exemple d'estructura de tesaurus (Font: Elaboració pròpia)

Les ontologies en canvi són llenguatges formals més comprensibles per màquina i tenen una major flexibilitat i manteniment, ja que reutilitzen el

coneixement d'un domini de forma automàtica. L'ontologia per definició es representada per un llenguatge més complex que permet a la màquina un treball més fàcil i llegible (informació formal). Les ontologies són representades per:

- Conceptes. Són idees bàsiques que es formalitzen. Els conceptes poden ser classes d'objectes, mètodes, plans, estratègies, processos de raonament, etc. I les classes poden ser representades per subclasses.
- Relacions. Representen la interacció i enllaç entre els conceptes d'un domini. Per exemple: subclasse de, part de, part-exhaustiva de, conectat-a, etc.
- Funcions. Són un tipus concret de relació on s'identifica un element mitjançant el càlcul d'una funció que considera diferents elements de l'ontologia. Per exemple, poden parèixer funcions com categoritzar-classe, assignar-data, etc.
- Instàncies. S'utilitzen per representar objectes determinats d'un concepte.
- Axiomes. Són teoremes que declaren sobre les relacions que han de complir els elements de la ontologia. Per exemple, "si A i B són de la classe C, llavors A no és de la subclasse B", o "per a tot A que compleixi la condició C1, A és B", etc.

Finalment, per representar una ontologia es necessiten definir tres elements bàsics:

1. Definir les classes de l'Ontologia
2. Definir la jerarquia de taxonomies de les classes (subclasse-superclasse)
3. Definir les propietats de les classes (atributs d'un objecte) i descriure els valors permesos per aquestes propietats.

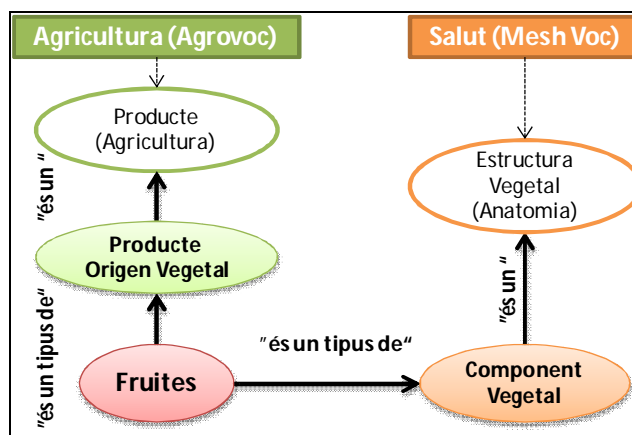


Figura 95. Exemple gràfic d'ontologia

En l'exemple anterior (figura 2), s'observen dos vocabularis amb diferents interpretacions conceptuals del mateix terme "Fruites", però, aquestes diferències no són un impediment perquè els tesaurs siguin alineats. Això és a causa dels vocabularis controlats tradicionals utilitzats amb biblioteconomia i documentació (tesaurus, llistes d'encapçalaments, etc.) els quals són convertibles fàcilment en ontologies. Aquest procés és possible gràcies a l'ús del llenguatge semàntic SKOS que s'aplica a aquests vocabularis (procés d'Skosificació). No obstant això, perquè aquest procés sigui possible és essencial que el vocabulari controlat sigui representat amb un llenguatge i format estàndard i interoperable com per exemple XML.

Descriptor "Top Concept": "Anatomy"	→ skos:TopConcept "Anatomy"
Terme Descriptor "Anatomy Category"	→ skos:prefLabel "Anatomy Category"
Terme No descriptor (USE)	→ skos:altLabel "Anatomia ES"
Concepte BT "Plant Structures"	→ skos:broader "Plant Structures"
Concepte NT "Plant Components"	→ skos:narrower "Fruit"
Concepte RT "X"	→ skos:related "X"

Figura 96. Exemple d'Skosificació (Font: Elaboració pròpia)

7.2.1. Funcionament: procés d'alineament

L'alineament d'ontologies, o procés de mapeig, passa per un procés automatitzat bàsic on s'avaluen un parell de vocabularis (Euzenat, et al. 2011):

1. Primerament, en el procés de mapeig (*matching*), el sistema de mapeig agafa dues ontologies i genera un alineament (A).
2. Posteriorment, hi ha un procés d'avaluació automatitzada (avaluació a través d'un *script*), en el qual el sistema, per mitjà d'un *script* d'avaluació, agafa l'alineament produït i el compara amb una alineació de referència que representa el resultat esperat del procés d'adaptació.
3. Finalment, hi ha un procés de comparació dels alineaments produïts: el sistema (avaluador automàtic) compara els dos alineaments i calcula mitjançant una mesura de qualitat de l'alineació (**M**) del component de mapeig que pot variar segons el sistema però, que normalment aquesta mesura fa referència a càlculs estadístics que mesuren la precisió i la recuperació dels alineaments obtinguts, finalment, compara els resultats amb de referència (**R**).

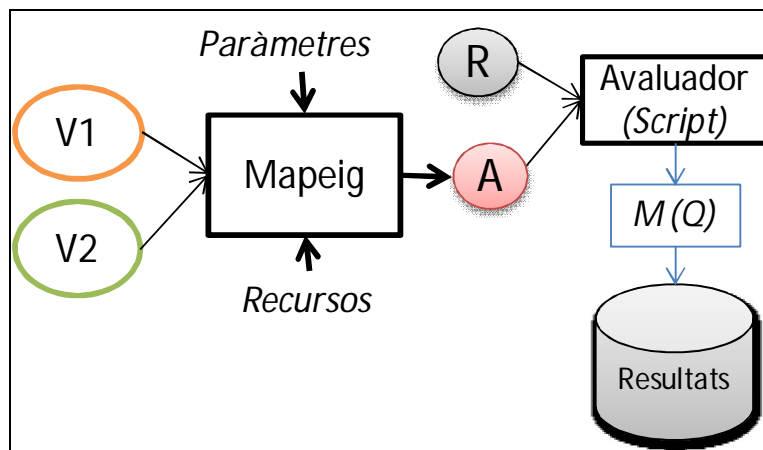


Figura 97. Procés de mapeig d'ontologies (Font: Elaboració pròpia)

Normalment existeixen unes tècniques esteses de mapeig i avaluació d'alineaments que consisteixen amb anàlisis estadístics i probabilístics. Algunes d'aquestes mesures qualitatives són:

- Càlcul de Precisió (*Precision*) = Número de documents o alineaments recuperats identificats com a rellevants dividit pel nombre total d'Ítems identificats (alineaments)
- Càlcul Recuperació (*Recall*) = Nombre d'Ítems (alineaments) rellevants identificats dividit pel nombre d'Ítems rellevants en la col·lecció (total d'alineaments possibles)
- F-Mesure = Suma o agregació de les mesures de Precisió i Recuperació

Aquestes mesures s'apliquen si el sistema de mapeig té accés a alineaments de referència per comparar i fer els anàlisis corresponents, però, si aquests alineaments de referència no existeixen, hi ha d'altres mesures alternatives a tenir en compte:

- Ús de recursos externs. Ús de sub-lexics, L'algoritme s'adapta al domini que s'està avaluant.
- Aplicació de nous paràmetres i mesures de qualitat (*OAEI initiative*)
- Tècniques de *datamining*. Es formulen preguntes per interrogar el sistema, com ara, quin sistema dóna la més alta o més baixa validació de correspondències, etc.

7.2.2. Problemàtiques del mapeig d'ontologies

Els problemes dels sistemes automàtics de mapeig es basen en els tipus de mètriques i càlculs elaborats a partir de les alineacions produïdes, referents a les avaluacions de les alineacions que en la majoria dels casos resulten ser imprecises. La major part de les mètriques se centren en l'avaluació

d'equivalències des del 2005 fins a l'actualitat, i mesuraven la presència o absència d'equivalències. Les dificultats s'enfocaven a trobar equivalències de classes en diferents ontologies i instàncies que representen un mateix individu.

A partir del 2009, ja comencen a aparèixer diferents tipus de mètriques enfocades en relacions d'equivalència "Exact Match" i "Close Match". En aquesta línia, també apareixen nous càlculs fonamentats en la distància entre termes, és a dir, en concordança amb el càlcul d'equivalències que mesuren la distància entre les relacions de les més properes a les més remotes. Entre el 2009 i 2011 fins l'actualitat, apareixen noves mesures alternatives basades en diferents tècniques:

- Tècniques de raonament:
 - Mètriques de càlcul sobre una població (n) amb major nombre de correspondències (etiquetatge manual). Es puntua sobre la precisió estimada
 - Tècniques de *Datamining* i patrons de correspondències (qüestions analítiques)
 - Raonament lògic. Mesura del nivell de incoherència que és causada per un alineament proposat.
- Aplicacions basades en rellevància (món bibliotecari). Complimenten el procés d'alineament a través de les anotacions que hi ha en els llibres/documents indexats en una col·lecció determinada:
 - Mètriques basades en el càlcul sobre una població (n), és a dir, es té en compte el major nombre de correspondències establertes. Es puntua sobre la precisió estimada. Per exemple:
 - Tècniques de *Datamining* i patrons de correspondències (qüestions analítiques)
 - Raonament lògic: mesura del nivell de incoherència que és causada per un alineament proposat.
 - Mesura d'eficiència. Executar diferents sistemes de mapeig en el mateix entorn d'avaluació controlada.

7.2.3. Factors determinants per a l'alineació de vocabularis

Existeixen diversos factors determinants per dur a terme els processos de mapeig entre vocabularis i seran requisits imprescindibles perquè les alineacions entre diferents vocabularis sigui possible:

- **Complexitat de les etiquetes.** Les etiquetes permeten descriure les classes, les relacions i les instàncies. Les etiquetes serveixen per trobar correspondències entre diferents entitats, a més, podem trobar-ne de diferents tipus i nivells de complexitat:
 - Etiquetes simples: per exemple etiquetes de tipus monolingüe
 - Etiquetes complexes: per exemple etiquetes multilingües, o etiquetes que estan lligades a vocabularis o recursos específics (WorNed, AGROVOC, aplicacions geoespaciales, etc.)

Cerca Tesaure Agrovoc (XML/RDF): Diatoms (EN)
USE: [32077](#) - Bacillariophyceae (EN)

(Exemple: Etiquetes multilingües) prefLabel	DIATOMEES (de) Diatomáceas (pt) Diatomeas (es) Diatomée (fr) Diatomee (it) диатомовые водоросли (ru) دیاتومها (fa)
(Exemple: Etiquetes lligades a Vocabularis)	Close Match: < http://aims.fao.org/aos/asfa/c_5261 > Exact Match: < http://lod.nal.usda.gov/nalt/14475 >

Figura 98. Exemples de diferents tipus d'etiquetes (Font: Elaboració pròpia a partir del tesaure AGROVOC)

- **Complexitat de les estructures.** Les estructures s'utilitzen per definir els elements emprats i, millorar i validar les hipòtesis inicials, és a dir, ajuden a trobar correspondències per mitjà de càlculs de similitud (estudi de l'estructura, el lèxic, etc.). D'una banda, un dels avantatges d'emprar les ontologies és el seu llenguatge de representació estàndard (RDF, OWL, SKOS), i, per tant, la interacció entre diferents vocabularis amb llenguatge ontològic és més fàcil. D'altra banda, el gran inconvenient en l'ús de diferents vocabularis és la utilització de varies estructures conceptuals. Per exemple, els tesaures tenen estructures jeràrquiques, i les ontologies, més expressives, empen diverses relacions entre classes, a més a més, aquestes poden ser limitades per distints axiomes (declaracions sobre les relacions): "Si A i B són de la Classe C, A no és subclasse de B", etc. L'estructura pot afectar tant a l'ús de diferents etiquetes com a la representació del contingut i les seves relacions. Per exemple, GEMET estableix les relacions dels conceptes mitjançant una categorització per grups i termes (s'estructura a través de 3 supergrups, 30 grups, 5 súper grups accessoris i 30 termes) a partir dels quals estableix les relacions jeràrquiques habituals (BT-NT) i associatives

(RT).

GEMET	AGROVOC
Groups list	<ul style="list-style-type: none">• BT resources (9001017) Natural resources (5091) • NT Agricultural resources (28001) • NT Biological resources (92341) ••• NT Animal resources (440) ••• NT Biodiversity (33949) •••• NT Agrobiodiversity (37977) •••• NT Genetic diversity as resource (33952) •••• NT Natural heritage (37967) ••• NT Plant resources (5979) ••• NT Wildlife (8389) •••• NT Wild plants (24126) •••• NT wild animals (24103) ••••• NT Game (3179) •••••• NT Game mammals (3182) •••••• NT game birds (3180) •••••• NT game fishes (3181) ••••• NT Wild boar (24131) ••• NT genetic resources (3218) •••• NT Gene pools (3212) •••• NT Germplasm (3249) •••• NT animal genetic resources (37418) •••• NT forest genetic resources (331594) •••• NT gene banks (11116) •••• NT plant genetic resources (37419) •• NT Energy resources (28006) ••• NT fuels (3136) •••• NT Biofuels (27465) ••••• NT Biodiesel (35013)
SOCIAL ASPECTS, ENVIRONMENTAL POLICY MEASURES	
ADMINISTRATION, MANAGEMENT, POLICY, POLITICS, INSTITUTIONS, PLANNING	
ECONOMICS, FINANCE	
ENVIRONMENTAL POLICY	
HEALTH, NUTRITION	
INFORMATION, EDUCATION, CULTURE, ENVIRONMENTAL AWARENESS	
LEGISLATION, NORMS, CONVENTIONS	
RESEARCH, SCIENCES	
RISKS, SAFETY	
SOCIETY	
HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE ENVIRONMENT	
AGRICULTURE, FORESTRY, ANIMAL HUSBANDRY, FISHERY	
CHEMISTRY, SUBSTANCES, PROCESSES	
EFFECTS, IMPACTS	
ENERGY	

Figura 99. Exemple de dificultats estructurals entre diferents vocabularis controlats (Font: tesaurus AGROVOC i GEMET)

Pel que fa a les ontologies, els tipus d'estructures són molt més complexes perquè presenten sempre una estructura de tripleta (subjecte –objecte - predicat) i per la dificultat de les relacions entre els termes a causa de la inclusió d'axiomes (teoremes que defineixen el tipus de relacions que han de tenir els elements d'una ontologia).

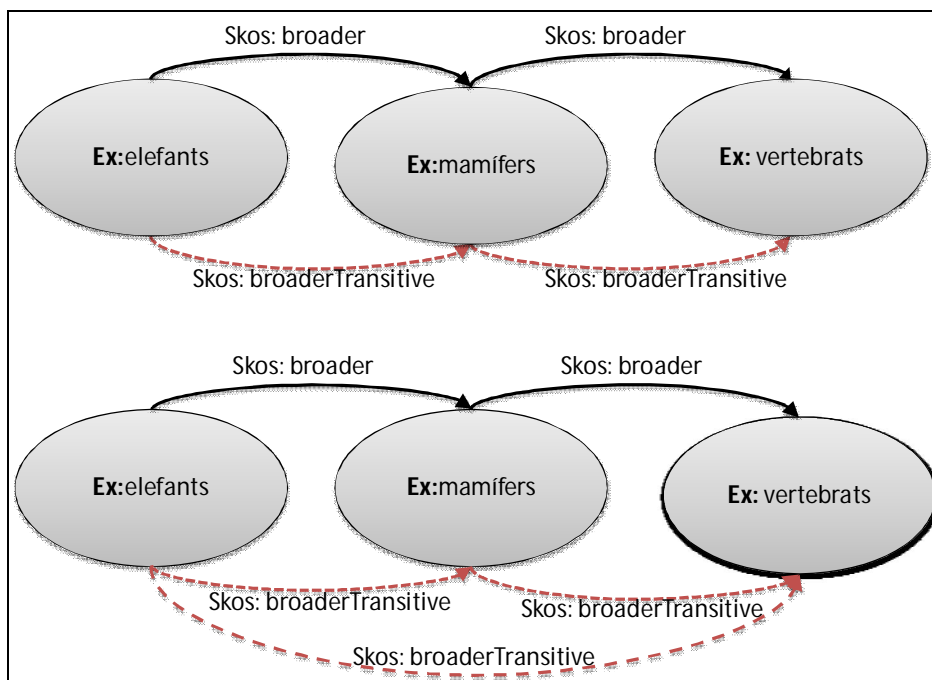


Figura 100. Exemple de relació Broader Term transitiva en una ontologia (Font: Elaboració pròpia amb suport guia SKOS-CoreGuide)

Aquest tipus de concordança s'ha de donar per representar totes les relacions jeràrquiques possibles entre les classes. Com s'exemplifica en la figura anterior, s'identifica i s'interrelaciona el concepte "gats" com un tipus específic d'animal, i alhora com un tipus determinat de mamífer, és a dir, tant "animal" com "mamal" són *broader terms* de "cats".

BT Vertebrats NT Mamífers
BT Mamífers NT Elefants
BT Vertebrats NT Elefants

Figura 101. Tipus de relacions transitives expressades de forma jeràrquica (Font: Elaboració pròpia)

- **Tipus de relacions semàntiques.** El tipus de relacions dependrà del sistema de mapeig, però les més habituals són:
 - Equivalència d'elements (classes i relacions). Per exemple, *skosmap:exactMatch* (mapeig exacte)
 - Subclasses. Establir relacions jeràrquiques inferiors a les classes, és a dir, definició de conceptes més específics en relació als seus conceptes genèrics: *skosmap:narrowMatch*, *skosmap:broadMatch* (relacions de tipus NT o BT)
 - Disjunció. Classes sense cap instància en comú, és a dir, classes que no tenen cap tipus de relació
- **Propietats formals dels alineaments.** La semàntica diferencia entre varis nivells d'interpretació de les relacions. En tot projecte de mapeig cal tenir definides prèviament els tipus de relacions que es volen establir en els alineaments, però, molts cops aquestes venen donades pels propis sistemes de mapeig. Per exemple:
 - Relació d'Equivalència o "equivalència lògica". Es tracta d'un alt nivell de similitud o intercanviabilitat entre 2 conceptes o classes d'objectes. Els tipus de relacions més habituals són: *skosmap:exactMatch* i/o *skosmap:closeMatch* (aquest darrer element no és tan comú)
 - Relació d'Equivalència. Si un alineament de 2 ontologies no inclou totes les classes o "x numero de classes" no és consistent. Aquest fet pot ser definit també en les relacions, per exemple: *skosmap:MinorMatch* o *skosmap:MajorMatch*; és a dir, podem

tenir un tipus de relació d'equivalència inexacta, "els termes A i B tenen una major o menor coincidència entre ells".

- **Cardinalitat i cobertura en l'avaluació (propietat significativa).** No hi ha restriccions en els alineaments i/o relacions entre els elements de diferents ontologies. Però, generalment la relació (mapeig) és de 1-a-1, perquè normalment el mapeig es produeix amb ontologies del mateix domini (cobertura), però, també podrien establir-se relacions amb ontologies de diferents dominis i cobertures i per tant, trobar relacions de tipus $n-a-m$ (de molts a molts).

Es conclou, que en tot sistema de mapeig s'ha de contemplar i analitzar prèviament els tipus d'ontologies que volen ser mapejades. És necessari estudiar la seva representació (anàlisi de l'estructura i les etiquetes que representen a les classes), els tipus de relacions que estableixen entre les classes (tipus, interpretació, consistència, etc.) i la cobertura.

7.2.4. Processos d'avaluació d'alineament

Actualment es contemplen 2 models generalitzats sobre el flux de dades dels processos d'avaluació d'alineaments.

El primer model, i més habitual, contempla dues ontologies (V1 i V2), i en alguns casos permet la inclusió de tests d'alineaments òptims proporcionats per l'usuari/institució d'interès (com ara, l'eliminació de n propietats de les ontologies, etc.). Aquests tests es produeixen per un "generador" que crea ontologies alternatives de forma automàtica, o bé, n'agafa una de referència. El primer pas és introduir el parell de vocabularis d'interès per a l'alineació a més de l'ontologia o ontologies de referència dintre d'un sistema de mapeig. Tot seguir, el sistema de mapeig realitza els processos d'alineaments corresponents amb les diferents eines i tècniques de mapeig per produir finalment uns alineaments (A). Aquests són valorats per un sistema o avaluador que executa diferents algorismes (per exemple, de similitud) que serveixen per avaluar els alineaments. En alguns casos l'avaluador pot contrastar els resultats amb alineaments de referència (R) que permeten una major efectivitat del procés, sempre i quant hi hagi disponible tests previs d'alineaments del parell de vocabularis d'interès. Finalment, els alineaments són sotmesos a mesures de qualitat (Q), d'anàlisi estadístics i probabilístics que permeten, aplicant mesures de recuperació de la informació i precisió, obtenir uns resultats finals de correspondències entre conceptes més precisos.

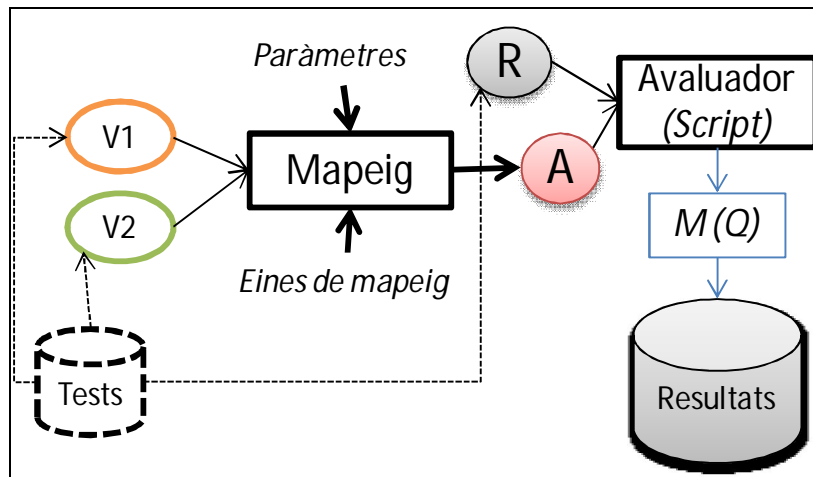


Figura 102. Procés de mapeig amb avaluació dels resultats amb tests o alineaments de referència (R) (Font: Elaboració pròpia)

Com s'ha apuntat en l'apartat 2.2, els sistemes de mapeig habituals no solen ser precisos, de fet, estan en constant canvi. Actualment, les tècniques d'avaluació no tenen en compte la qualitat i l'efectivitat de les intervencions dels usuaris amb els processos de mapeig i/o el sistema d'avaluació de mapeig en general (Euzenat, et al., 2011). Així doncs, una alternativa al model presentat anteriorment (figura 102) és un sistema de mapeig que inclou un avaluador-usuari com a part integrant del procés per valorar el sistema.

El sistema de mapeig realitza els processos descrits anteriorment però, quan aquest no té una ontologia de referència per avaluar-se l'usuari agafa el rol d'avaluador. L'usuari avalua i descriu els alineaments i després el sistema reutilitza la informació per a realitzar tests d'avaluació fets pel propi usuari. Per tant, hi ha una retroalimentació del sistema (generador de tests), que pot reproduir els tests per avaluar els alineaments entre classes, o bé, per valorar el mapeig de dues ontologies. Aquest procés s'anomena mapeig semiautomàtic, ja que no tots els processos de mapeig són automàtics o fets per una

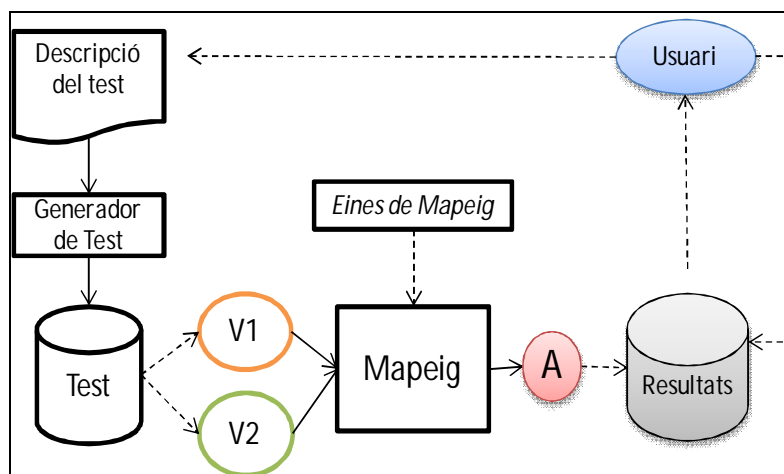


Figura 103. Procés de mapeig amb un usuari com avaluador (Font: Elaboració pròpia)

7.2.5. Avaluació de software de mapeig automàtic de vocabularis

El mapeig de les ontologies i la seva interoperabilitat dependrà del format amb què es desenvolupi. Principalment hi ha dos tipus de formats predominants, l'OWL i l'SKOS que provenen del format RDF. Aquests, no sempre compatibles entre ells, requereixen sistemes de gestió, visualització i mapeig determinats (vegeu capítol 4, Estàndards i Formats).

La bibliografia ens mostra que aquests llenguatges encara estan en fase de desenvolupament a l'igual que les eines utilitzades per a la seva gestió. Tot i així, en el camp de les ciències aquàtiques i afins predominen les ontologies en format SKOS, per aquest motiu s'estan avaluant eines de mapeig que treballin amb aquest format.

El funcionament de les eines de mapeig es basa en l'anàlisi de les relacions (coincidències exactes, coincidència a nivell dels termes genèrics i a nivell de termes específics). Aquests sistemes s'avaluen pel seu grau de precisió, tant a nivell d'equivalències exactes com per la resta de casos que es puguin produir (no equivalències, quasi-equivalències, etc.). L'organisme que s'encarrega de valorar aquests sistemes és l'OAEI (*Online Alignment Evaluation Initiative*), el qual identifica tres softwares destacats, avaluats al llarg del temps i que accepten el format SKOS: RIMOM, FALCON i COMA++.

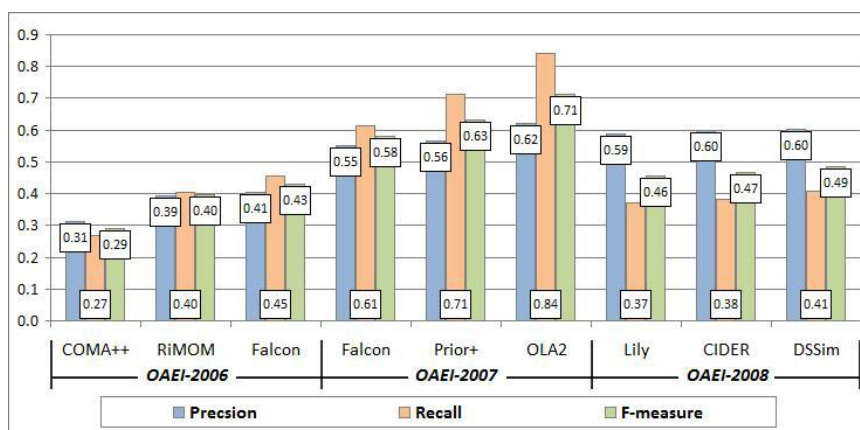


Figura 104. Resultats de l'evolució de la qualitat del mapeig durant 2006-2008 (OAEI, 2008)

Així doncs, tot i pensar en primer lloc en els sistemes de mapeig més representatius i avaluats per la OAEI, RiMOM, Falcon-OA, COMA++, per realitzar diversos experiments es va observar, després de diverses proves, que els sistemes no acceptaven "large ontologies" o ontologies de gran mida com son els tesaurus escollits com a objecte d'estudi. Finalment es va optar per l'ús d'un software en desenvolupament però operatiu, que permetés fer

proves amb “large ontologies”. Aquest software anomenat Amalgame (AMsterdam ALignment GenerAtion MEtatool) és un programari en desenvolupament de la Universitat d’Amsterdam (VU) que s’emmarca dintre del context dels projectes europeus PrestoPrime i EuropeanaConnect. L’ús del programa per aquesta tesi ha estat gràcies a un estada de col·laboració amb la Universitat d’Amsterdam (VU), la qual forma part del grup d’experts que formen l’[EuropeanConnect](#), que tenen com principal objectiu ajudar en l’accessibilitat dels diferents recursos de patrimoni cultural. Per conèixer tot el context d’Europeana i el paper rellevant que hi juga en el web semàntic cal veure el “White Paper” d’Europeana (verure “About us” d’Europeana). El grup de la UV, s’encarrega també d’enriquir les dades, especialment utilitzant els “vocabularis controlats” com a enllaços entre els diferents recursos. Estan treballant per construir un “semantic layer”, és a dir, una col·lecció de dades RDF on les col·leccions d’objectes proporcionades per les diferents biblioteques que es connecten amb diferents sistemes d’organització del coneixement (tesaurus, llistes d’autoritats, noms geogràfics, etc.). La finalitat és millorar l’accés al sistema d’informació. (Veure l’estudi de cas de cerca semàntica a [“ThoughLab”](#)). La descripció del treball realitzat per aquest grup es pot trobar a: <http://www.few.vu.nl/~aisaac/econnect/> (Annex 1 del projecte, descripció del projecte, p.55 i document D1.1.2 on es mostren resultats del “semantic layer”). Quan es van realitzar els experiments, Europeana encara no havia desenvolupat el model de metadades EDM, i només disposava de metadades

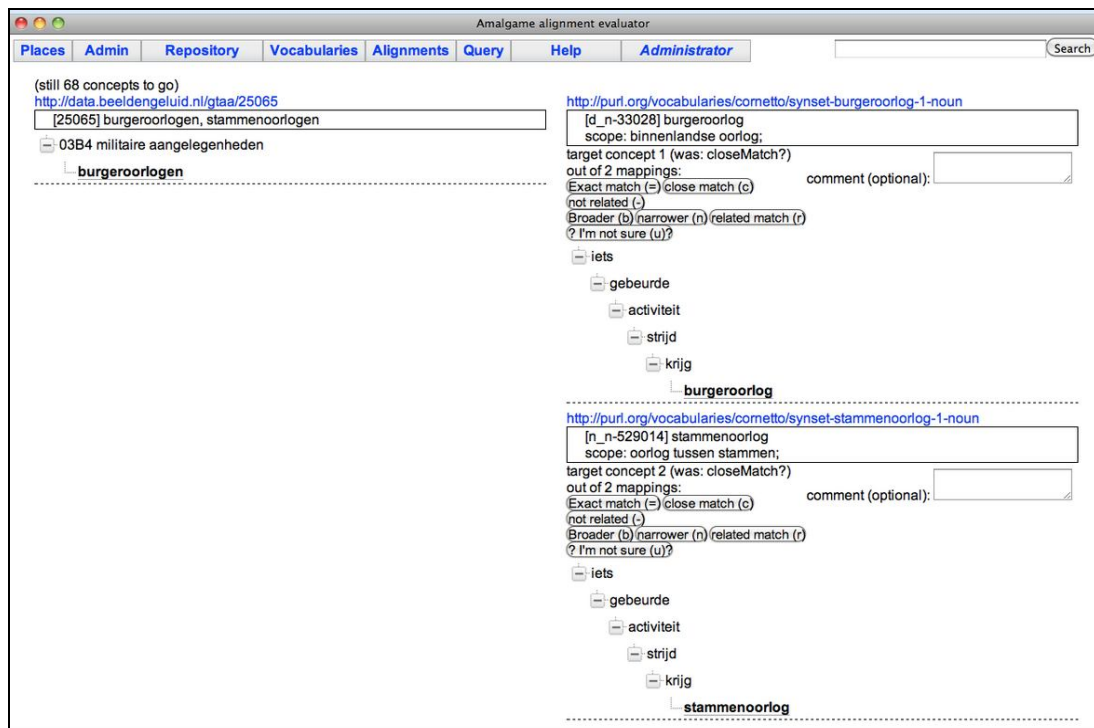


Figura 105. Interfície d’avaluació del sistema de mapeig Amalgame (Font: Elaboració pròpia)

Amalgame és una eina per trobar, avaluar i gestionar alineaments de vocabularis. Tracta de combinar les tècniques i mètodes de mapeig existents utilitzant una interfície de configuració amb un flux de treball bàsic. Utilitza les tecnologies del web semàntic SKOS i RDF, i el llenguatge de programació SWIProlog (Amagamae, W3C, 2013).

- Paral·lelament, també es va avaluar el sistema de mapeig de Neon Toolkit (European Commission Project, 2006-). El *plugin* d'alineament que incorpora el sistema pot calcular i gestionar automàticament les alineacions de l'ontologia. Aquest *plugin* d'alineament (versió 2.3) ofereix les següents funcionalitats: Troba alineacions entre ontologies introduïdes en el sistema o les que estiguin disponibles en el servidor
- Combinació d'ontologies
- Particions d'alineacions per aplicar-ho en les alineacions existents
- Representació i Recuperació dels alineaments en un format concret
- Emmagatzematge d'alineaments permanents en un servidor.

Quant a NeOn, es descarta com a sistema de mapeig per realitzar els experiments d'aquest estudi, perquè no llegeix correctament els següents vocabularis estudiats: AGROVOC, GEMET i ASFA. Potser, el problema resideix en els vocabularis controlats importats que són tesaurus convertits a ontologies, i en les seves estructures interpretades erròniament (vegeu Annex 7.1). A més, s'han importat altres vocabularis d'origen ontològic per provar el sistema de forma correcta, però els resultats no han estat els esperats per recomanar el sistema en l'alineament d'ontologies. Caldrà tenir aquest sistema com a referents pel seu continu canvi i desenvolupament.

Posteriorment, al 2012, encara en fase d'experimentació, es consulta un nou sistema HIVE (Greenberg, et al., 2011), que es presenta per primer cop en el *SKOS-2-HIVE workshop* a la Universitat Carlos III de Madrid el 19 de juny de 2012. HIVE (*Helping Interdisciplinary Vocabulary Engineering*) és un projecte del [Metadata Research Center \(MRC\)](#) de la [School of Information and Library Science, University of North Carolina at Chapel Hill \(UNC/CH\)](#) i el [National Evolutionary Synthesis Center \(NESCent\)](#) en Durham (Carolina del Norte), finançat per l'[Institute of Museum and Library Services \(IMLS\)](#) d'Estats Units. Aquest sistema té l'objectiu de crear un model per integrar múltiples sistemes d'organització del coneixement, però és també un sistema que suporta la generació automàtica de metadades, i, a més té l'enfocament de la integració dinàmica de la codificació de vocabularis mitjançant [SKOS \(Simple Knowledge Organization System o sistema d'organització del coneixement simple\)](#). Però, es descarta com a sistema d'alineament d'ontologies perquè en el moment de la seva avaluació, juny de 2012, no disposava d'un mòdul d'usuari com a tal per poder avaluar els alineaments.

7.3. Conclusions

Els processos automàtics d'alineaments, en general, assumeixen que un alineament només pot ser correcte o incorrecte, però hi ha molts matisos, com la quasi equivalència (*close match*). Cada cop més es troben sistemes que diferencien els tipus de relació de sinònims i de quasi-sinònims (*exact match* i *close match*).

La qualitat del procés de mapeig és també relativa, ja que molts cops dependrà de la utilitat de l'alineament generat. Cal tenir en compte els aspectes subjectius (processos de mapeig semiautomàtic) i el context per utilitzar el vocabulari alineat, entre d'altres aspectes. Així doncs, per tal d'avaluar la qualitat dels alineaments caldrà aplicar mesures addicionals com la rellevància que s'aplica en processos de mapeig automàtic i semiautomàtic.

Una de les tendències actuals, per millorar la qualitat dels alineaments passa per tenir en compte l'usuari i l'escenari/context d'ús. És a dir, donar més valor al processos de mapeig semiautomàtic que al sistemes completament automàtics. En aquests processos semiautomàtics hi ha un avaluador manual (usuari expert-especialitzat) dels alineaments generats pel sistema, i, per tant, es recomana utilitzar guies d'avaluació estàndards que especifiquin les casuístiques per mapejar els vocabularis, i incloguin casos d'ús per millorar els processos de mapeig.

Finalment, es conclou que encara calen altres mesures d'avaluació dels alineaments i dels processos de mapeig (com la inclusió de tests addicionals, etc), perquè en els sistemes de mapeig encara s'hi detecten errors. En general, és necessari definir un consens entre els sistemes de mapeig (com per exemple, COMA++, AgreementMaker, etc.) per tal de poder comparar esquemes de mapeig produïts per cada sistema (OAEI, 2011) i millorar així l'efectivitat dels processos automàtics de mapeig.

CAPÍTOL 8. PROTOTIPUS DE MAPEIG: ESTUDI EXPERIMENTAL D'ALINEACIÓ DE TESAURUS EN CIÈNCIES AQUÀTIQUES I AFINS

8.1. Introducció

En aquesta recerca⁵¹ s'estudia l'alineació de vocabularis des del punt de vista de la interoperabilitat entre els tesaurs multidisciplinaris i multilingües, ja que són els vocabularis més utilitzats en l'àmbit de les ciències aquàtiques. També es té en compte com els sistemes de mapeig i les tècniques d'alineaments d'ontologies poden ajudar i millorar els sistemes d'informació i documentació. Aquest estudi experimental està format per les següents parts; la introducció, on es defineix tant l'abast del projecte experimental com els objectius i les hipòtesis a assolir, la metodologia desenvolupada, l'anàlisi qualitativa dels vocabularis controlats objecte d'estudi per tal d'avaluar la seva interoperabilitat i consistència (anàlisi d'una mostra d'alineacions manuals), l'estudi experimental d'alineació semiautomàtica dels vocabularis i avaluació dels alineaments produïts, els resultats finals de l'estudi, la discussió i les conclusions finals.

L'objectiu d'aquest estudi és demostrar dues coses

1. Els sistemes de mapeig ajuden a la convergència de vocabularis diferents però, amb disciplines relacionades, així es dona la interoperabilitat semàntica entre diversos vocabularis controlats. Podem trobar fàcil i ràpidament la similitud entre dos tesaurs i/o la manca d'informació en una disciplina determinada.
2. Els sistemes de mapeig permeten el multilingüisme i/o la interoperabilitat de vocabularis amb diverses llengües.

De manera paral·lela es vol demostrar que l'ús del SKOS (format semàntic emprat per a l'elaboració d'ontologies), que és un model basat en el concepte, també és compatible amb els models de vocabularis tradicionals basats en el terme.

Per executar aquesta fase experimental es parteix de 3 vocabularis controlats (AGROVOC, GEMET i ASF A), relacionats amb les ciències aquàtiques. Tots tres tesaurs són multidisciplinaris i multilingües, però comparteixen disciplines comuns (medi ambient, agricultura i alimentació, biodiversitat i ciències aquàtiques i marines). Tot i que existeix informació sobre sistemes de mapeig i

51 Tots els gràfics i taules que no incorporen cites són d'elaboració pròpia. Aquestes han estat elaborades amb material de suport citat al llarg del text com són els tesaurs objecte d'estudi (AGROVOC, GEMET i ASF A) i el programari Amalgame utilitzat en els experiments de mapeig.

alineaments dels tesaurus AGROVOC i GEMET (*van Hague 2007 i 2010; Caracciolo 2012*), mai s'han realitzat estudis incloent el tesaurus ASFA i/o la branca de les ciències aquàtiques i marines. Tanmateix, la majoria de recerca publicada fa referència a l'eficàcia dels sistemes de mapeig automàtic d'ontologies, des del punt de vista de la recuperació de la informació, avaluant la rellevància, la precisió i l'exhaustivitat. No obstant això, hi ha altres estudis interessants que cal citar com Lauser, et al., (2008), inclòs també en la tesis de van Hague (2008), que avalua el mapeig humà vers l'automàtic i conclou que ambdues perspectives són complementàries. Quant a ASFA es troben escassos estudis sobre mapeig, alguns com el de Gangemi (2004) mostra l'alineació d'ASFA amb AGROVOC i un tercer vocabulari (OneFish) per a la integració d'aquests en un sistema d'informació sobre pesca. El treball de Caracciolo (2008) també estudia la superposició d'ASFA i AGROVOC, però en el cas d'AGROVOC només s'utilitzen les àrees temàtiques de "organisms," "vehicles" (*including vessels*), i "fishing gears" i, a més, l'estudi només es centra en els sistemes de mapeig automàtic. De fet, al 2012, la pròpia organització gestora del tesaurus ASFA publicà un treball sobre alineament de tesaurus (Caracciolo, 2012, pàgina 71) on cita explícitament que aquest vocabulari encara està en progrés de conversió a RDF/SKOS. Finalment, es conclou que no hi ha estudis sobre alineament dels tesaurus ASFA i GEMET.

Cal destacar també els alineaments per millorar un vocabulari existent, o bé, crear-ne un de nou, i l'ús dels sistemes de mapeig per fusionar vocabularis controlats per aconseguir millorar la recuperació de la informació d'un sistema d'informació multilingüe i multidisciplinari. D'una banda, com s'explicava en els primers capítols, les ciències aquàtiques és un camp interdisciplinari, i com s'ha demostrat en estudis previs (Reverté, C., 2006, 2008 i 2009a) té una cobertura temàtica insuficient per part dels vocabularis existents. D'altra banda, l'existència d'un vocabulari marc de la disciplina de les ciències aquàtiques (tesaurus ASFA) permet que el treball de mapeig (AGROVOC, GEMET i ASFA) millori el propi vocabulari de l'àrea d'estudi, amb l'objectiu d'actualitzar-se si es necessari, o bé, d'utilitzar l'estudi de mapeig com a base per a la creació d'un nou vocabulari dins d'un futur sistema d'informació de qualitat.

Cal recordar, que el plantejament d'aquest estudi experimental no és construir un nou vocabulari, que requereix molts elements, condicionants i un alt nivell de recursos i fonts d'informació, sinó que l'objectiu principal és avaluar la capacitat de convergència entre els tres vocabularis a través de la tecnologia semàntica (interoperabilitat). Perquè això sigui possible, cal avaluar l'extracció dels termes, la normalització d'aquests (evitar termes artificials en una o més llengües), la determinació de les àrees principals i la seva distribució, l'establiment de les relacions d'equivalència (quins termes són considerats preferits i quins no) i la seva estructura (relacions jeràrquiques i associatives). No obstant això, quan estem davant d'un projecte d'alineació de tesaurus multilingües, cal fer especial referència a l'anàlisi de les relacions entre els

tesaurus, és a dir, als termes preferits i no preferits existents entre una mateixa llengua i dos llengües diferents, on un terme preferit en un vocabulari A, pot ser un terme no preferit en el vocabulari B, i al contrari. De fet, aquesta és una bona metodologia per trobar les equivalències entre els termes de diversos vocabularis.

Quant a l'abast del projecte experimental, a priori, no es fan restriccions a nivell de cobertura geogràfica, cronològica ni tampoc a nivell temàtic, ja que els experiments es duran a terme amb tecnologies de mapeig de vocabularis experimentals per explotar la seva potencialitat (valor). A més, interessa modelar la tecnologia per tal d'identificar camps del coneixement afins a les ciències aquàtiques inclosos en els vocabularis controlats a objecte d'estudi. Finalment, com s'ha explicat en el capítol 7, el software emprat per al mapeig dels tesaurus és Amalgame.

8.2. Metodologia

En primer lloc, s'analitzen tres vocabularis controlats que són objecte d'estudi (ASFA, AGROVOC i GEMET) per avaluar principalment la seva interoperabilitat amb la tecnologia que ofereix el web semàntic. Per tal d'executar l'experiment, s'aconsegueixen els tres tesaurus amb format ontològic (SKOS). Els tesaurus amb format SKOS s'han obtingut mitjançant el web dels propis tesaurus (AGROVOC i GEMET), a excepció del tesaurus ASFA que és va demanar via correu electrònic al propi gestor del tesaurus, la FAO (*Food and Agriculture Organization of the United Nations*). El format SKOS només es va poder aconseguir la versió monolingüe, ja que encara no tenien la versió multilingüe amb format ontològic. A continuació, els vocabularis s'avaluen prèviament de forma manual (Annex 8.1) per conèixer les seves característiques i paràmetres (estructura, cobertura, etiquetes i relacions), que permeten saber el nivell d'interoperabilitat entre els vocabularis.

En segon lloc, es provaren diferents sistemes de mapeig automàtic reconeguts i d'accés obert, per realitzar diverses proves d'alineaments dels tres tesaurus i comprovar que la tecnologia del web semàntic permet la interoperabilitat entre vocabularis de diverses disciplines i llengües. La majoria de sistemes de mapeig estaven en fase de desenvolupament, així que no es podien fer proves amb vocabularis tan extensos com els exposats. En conseqüència, es va optar per cercar un altre sistema de mapeig que fos prou flexible i àmpliament testat per ser capaç de treballar amb els tesaurus a estudi. El sistema de mapeig emprat en l'experiment no era d'accés lliure però estava operatiu i integrat en un sistema d'informació europeu sobre patrimoni cultural (EUROPEANA), i, per tant, dins un entorn multilingüe. El sistema, suportat i gestionat per la

Universitat d'Amsterdam, es va testar gràcies a un acord de col·laboració entre la doctoranda i la *Vrije Universiteit Amsterdam*.

Un cop integrats els vocabularis en el sistema de mapeig escollit es va seguir el mateix model metodològic d'avaluació del sistema de la institució que va facilitar el programari. Aquesta avaluació tenia el funcionament estàndard dels sistemes semiautomàtics de mapeig d'ontologies. A priori, es van realitzar unes estadístiques bàsiques i generades automàticament pel sistema, que identificaren el nombre de paraules que contenia cada vocabulari, identificació del nombre i tipus de relacions (*broader*, *narrower*, *top concepts*, etc.). A posteriori, s'executà el mapeig de vocabularis, aquest només es donava per cada parell de vocabularis, així que es realitzaven fins a 6 processos de mapeig (A-B, A-C, B-A, C-A, B-C i C-B), i, tot seguit, el propi sistema generava mostres aleatòries de cada parell de tesaurus alineats (identificats com a estrats o *stratum*). Llavors, es van seleccionar diferents mostres individuals per cada parell de tesaurus per tal d'observar que les alineacions fossin correctes. Un cop es comprovà que els processos de mapeig havien estat òptims, per mitjà d'avaluacions manuals (visualització i lectura dels alineaments), es va decidir realitzar dos experiments més de forma separada. El primer experiment consistia en el mapeig dels tres vocabularis en la seva versió monolingüe (experiment 1), i el segon era un mapeig a nivell multilingüe dels dos vocabularis disponibles en aquesta versió (GEMET i AGROVOC). Cal recordar que era necessari separar els experiments perquè en el cas del tesaurus ASFA només es disposava de la versió monolingüe.

Després, es varen seleccionar varies mostres (*stratums*) que pertanyien als dos grups mencionats (mapeig monolingüe i mapeig multilingüe) amb l'objectiu d'avaluar el sistema de mapeig de forma adequada. Les avaluacions són les anàlisis dels resultats dels mapejos, és a dir, es realitza una anàlisi qualitativa dels alineaments d'un parell de termes, que el sistema identificava aleatòriament com a conceptes amb algun tipus de relació d'equivalència (equivalència exacta, quasi exacta, relació de tipus genèrica, específica, de tipus relacionat o no relacionat). Per tal de dur a terme aquestes avaluacions es va seguir una guia metodològica d'avaluació d'alineaments aplicada pel propi grup gestor del sistema d'alineament. Aquesta guia estableix les relacions d'equivalència habituals d'un sistema de mapeig, *broaderMatch*, *narrowerMatch*, *norelatedMatch* (entre d'altres com) ,*relatedMatch* i *closeMatch*, entre d'altres.

Finalment, l'avaluació dels alineaments o identificació dels tipus de relacions d'equivalència de termes (A versus B), era un procediment que el sistema emmagatzemava i comptabilitzava per acabar generant uns resultats d'alineaments. Aquests resultats ens proporcionaven, entre d'altres objectius, si les dades del sistema de mapeig eren correctes i precises. Els resultats van ser recollits amb taules de dades, per elaborar estadístiques, i així, comprovar els

promitjos existents per cada alineament i tipus de relació d'equivalència identificada.

8.3. Anàlisi de les característiques i paràmetres dels vocabularis controlats: alineacions manuals

En aquesta secció s'analitzen les estructures, etiquetes, cobertura, tipus de relacions existents entre els conceptes del cada vocabulari i possibles restriccions entre d'altres elements que puguin ser significatius.

D'una banda, es parteix d'estudis previs realitzats (Reverté, C., 2006) on després d'una fase de proves d'alineaments manuals s'observava un cert nivell d'interoperabilitat entre els vocabularis estudiats. D'altra banda aquestes proves es reproduïen i s'amplien en aquest estudi (Annex 8.1), els resultats del qual verifiquen que entre els vocabularis hi ha un cert grau de consistència, però també un alt grau d'inconsistència sobretot a nivell cultural i de cobertura (exemplificat també en el capítol 7). En aquest estudi, els alineaments manuals es centren concretament en l'anàlisi de descriptors afins a les ciències aquàtiques (aigua, aqüicultura, biodiversitat, biologia, canvi climàtic, ciències aquàtiques, ecosistemes, geologia, gestió, hidrologia, hidrobiologia, medi ambient, recursos naturals, zones humides, etc.). També s'estudia un quart tesaurus sobre biodiversitat, l'NBII (Annex 8.1), que s'exclou de l'anàlisi comparativa perquè no es pot incloure en els experiments de mapeig automàtic per no disposar de versió ontològica.

A continuació, es resumeixen els resultats que mostren la diferència estructural (a nivell de mida) i de cobertura (nivell d'exhaustivitat, segons el nombre de descriptors que engloba cada concepte analitzat, i identificació de diferències de tipus cultural o semàntic) entre els tesaurus:

Mostra de Conceptes (mapeig manual)⁵²	ASFA (A)	AGROVOC (B)	GEMET (C)⁵³
Aquaculture	24	20	12
Aquatic Sciences	21	37	18
Ecology	27	32	11
Ecosystems	9	15	8
Geology	16	17	11
Management	10	50	16
Environment	8	25	0

⁵² Nota: quan apareix el nombre 0, significa 0 termes, és a dir que el concepte no existeix.

⁵³ Nota: es comptabilitzen també com a descriptors, els termes classificats dintre de "Groups" i "Themes"

Oceanography	19	10	7
Natural Resources	18	27	11
GIS	2	6	4
Remote Sensing	18	16	10
Biodiversity	4	14	4
Climatic Change	16	7	5
Aquatic Ecosystem ⁵⁴	0	0	8
Limnology	14	3	3
Hydrology	9	7	4
Hydrobiology	7	5	4
Hydrodynamics	12	4	0
Hydrogeology	1	0	0
Hydrogeomorphology	1	0	0
Aquatic environment	17	12	8
Body Water/Water Mass	25	3	8
Wetlands	8	9	7
Materials ⁵⁵	20	15	3
Modelling	3	0	4
Monitoring	6	16	7
Total N° Descriptors	315	350	173

Taula 33. Nombre de descriptors per cada concepte objecte d'estudi del mapeig manual

Tenint en compte que s'han escollit conceptes de l'àmbit del thesaurus ASFA, a excepció de termes més genèrics com és el cas de conceptes de tipus faceta (materials, modelització i monitoreig), s'observa que hi ha encara un 27% dels descriptors analitzats d'altres thesaurus que ofereixen major cobertura (*ecology, management, environment, natural resources, biodiversity* i *wetlands*) vers els seus corresponents termes equivalents a ASFA. Corroborant la hipòtesis que

⁵⁴ Nota: el terme "Aquatic Ecosystem" no existeix en ASFA i AGROVOC, si es troba "Aquatic Environment" o "Aquatic Sciences" (com a terme superior)

⁵⁵ Nota: el thesaurus GEMET no té un sol terme per designar materials, contempla 2 "top concepts": "ciències dels materials" i "tecnologies dels materials"

l'ús dels sistemes de mapeig poden servir per a la fusió de tesaurus amb l'objectiu d'estendre un vocabulari determinat.

Un cop es realitza el mapeig manual (Annex 8.1) s'analitzen els descriptors a nivell només d'*exact match*. Aquestes anàlisis mostren diferències específiques en relació amb la cobertura dels tesaurus analitzats.

Mostra de conceptes (mapeig manual)	A-B (alineacions en la jerarquia)	A-C (alineacions en la jerarquia)	B-C (alineacions en la jerarquia)
Aquaculture	10	7	7
Aquatic Sciences	3	0	0
Ecology	9	3	3
Ecosystems	1	0	0
Geology	9	2	2
Management	3	2	3
Environment	1	0	0
Oceanography	8	2	2
Natural Resources	9	5	4
GIS (Geographic information systems)	2	1	1
Remote Sensing	3	1	2
Biodiversity	2	1	0
Climatic Change	1	1	1
Aquatic Ecosystem	0	0	0
Limnology	1	0	0
Hydrology	1	0	0
Hydrobiology	2	1	1
Hydrodynamics	1	0	0
Hydrogeology	0	0	0
Hydrogeomorphology	0	0	0
Aquatic environment	5	2	2

Body Water/Water Mass	1	2	1
Wetlands	2	0	0
Materials ⁵⁶	5	1	0
Modelling	0	2	0
Monitoring	2	1	0
Alineacions totals	80	34	29

Taula 34. Alineacions resultants del procés de mapeig manual

8.3.1. Estructura i relacions

L'opinió més comuna és que totes les diferents versions lingüístiques d'un tesaurus multilingüe han de ser idèntiques i simètriques, cada descriptor ha de tenir un i només un equivalent en totes les llengües i estar relacionat de la mateixa manera amb d'altres descriptors en la llengua corresponent (relació asimètrica de diccionari a diccionari). Però en els casos estudiats (AGROVOC, GEMET i ASFA) ens trobem amb estructures que no són ni idèntiques ni simètriques. El nombre de descriptors en cada vocabulari no és necessàriament el mateix, ni tampoc la forma en què els descriptors estan relacionats entre ells. Les principals raons que expliquen aquestes diferències entre els tesaurus són de tipus culturals i semàntic.

Respecte a les diferències d'estructura i relacions, cal a dir, com s'observa en la figura 99, que GEMET és el tesaurus amb major diferència d'estructura en comparació amb la resta. GEMET estableix les relacions dels conceptes partint d'una categorització per grups i termes (s'estructura a través de 3 supergrups, 30 grups, 5 súper grups accessoris i 30 temes) abans d'ordenar les relacions jeràrquiques habituals (BT-NT) i associatives (RT), de fet, aquestes relacions només es troben en els "temes". En canvi, AGROVOC disposa des del primer concepte, les relacions jeràrquiques i associatives habituals (BT-NT-RT).

Altres problemes sorgeixen per l'heterogeneïtat de les estructures jeràrquiques, poden existir diverses relacions entre dos o més conceptes i classes de conceptes. O sigui, són relacions polijeràrquiques (Veure taula 35) on el concepte "Surface Water", segons el tesaurus ASFA i GEMET, és un terme específic de "Water", en canvi, per al tesaurus AGROVOC és un terme específic del concepte "Water resources" i "Land cover". Així com "Water" és un terme genèric (BT) en els tesaurus ASFA i GEMET, en AGROVOC és un terme relacionat (RT).

⁵⁶ Nota: En el cas d'ASFA hi ha un descriptor relacionat que és "Materials technology" que és l'únic concepte equivalent que correspon a un dels "Top Concepts" de GEMET, "Materials technology".

ASFA	GEMET	AGROVOC
<p>SURFACE WATER</p> <p>(No hi ha definició)</p> <p>Broader Term (BT) Water</p> <p>Related Term (RT) Bottom water Epilimnion Evaporation Shallow water Surface layers Surface water masses</p>	<p>SURFACE WATER</p> <p>Definition: All waters on the surface of the Earth found in streams, rivers, ponds, lakes, marshes or wetlands, and as ice and snow. (Source: LANDY/BJGEO)</p> <p>Broader Term (BT) water (geographic)</p> <p>Narrow Term (NT) brackish water drainage water freshwater inland water salt water seepage water thermal water</p> <p>Scope note: <i>scope note is not available</i></p> <p>Groups: HYDROSPHERE (freshwater, marine water, waters)</p> <p>Themes: water</p> <p>Other relations: Has exact match AGROVOC: Surface water EuroVoc: surface water</p>	<p>SURFACE WATER (creada:1988-03-11) (No hi ha definició)</p> <p>Broader Term (BT) 8325 - Water resources (EN) 37897 - Land cover (EN)</p> <p>Narrow Term (NT) 8326 - Water springs (EN) 6697 - Runoff water (EN) 6617 - Rivers (EN) 6105 - Ponds (EN) 50112 - gulfs (EN) 4160 - Lakes (EN) 24124 - Wadi (EN) 14931 - Torrents (EN) 8330 - Waterfalls (EN)</p> <p>Related Term (RT) 3727 - Hydrography (EN) 25305 - Water harvesting (EN) 8309 - water (EN)</p>

Taula 35. Exemple d'heterogeneïtat entre estructures i relacions entre vocabularis controlats (tesaurus)

8.3.2. Etiquetes (lèxic)

AGROVOC i ASFA, elaborats per la mateixa organització, divergeixen de GEMET a nivell d'etiquetes. Per exemple, GEMET usa singulars i ASFA i AGROVOC plurals. A continuació hi ha un exemple que mostra les grans diferències entre els tres tesaurus a nivell de singulars i plurals, a més, s'hi pot observar també les variacions existents en quant a cobertura temàtica (GEMET en general inclou menys descriptors, sobretot específics com són els *Narrower terms* o NT):

ASFA	AGROVOC	GEMET
Natural Resources	Natural Resources	Natural Resource
(Note: Restricted to resources within or beneath the aquatic environment)		Definition: A feature or component of the natural environment that is of value in serving human needs, e.g. soil, water, plantlife,

		wildlife, etc.
BT Resources	↔ BT 9001017 - resources (EN)	broader terms
NT Common property resources	NT 92341 - Biological resources (EN)	resource
NT Energy resources	NT 35106 - Feed resources (EN)	narrower terms
NT Food resources	NT 28012 - Marine resources (EN)	biological resource
NT Living resources	NT 28006 - Energy resources (EN)	climate resource
NT Marine resources	NT 28001 - Agricultural resources (EN)	mineral resource
NT Mineral resources	NT 24056 - Nonrenewable resources (EN)	sea resource
NT Nonrenewable resources	NT 7187 - Soil resources (EN)	soil resource
NT Renewable resources	NT 3026 - Food resources (EN)	water resource
NT Unconventional resources	NT 3050 - Forest resources (EN)	related terms
NT Water resources	NT 3218 - genetic resources (EN)	natural material
rt Multiple use of resources	NT 2942 - Fishery resources (EN)	nature protection
rt Protected resources	NT 6502 - Renewable resources (EN)	Groups: RESOURCES (utilisation of resources)
rt Rare resources	RT 36649 - Biological production (EN)	Themes: resources
rt Raw materials	RT 2476 - Economic resources	

Taula 36. Exemple sobre diferències entre etiquetes (ús diferent de singulars i plurals) i nivell de cobertura temàtica entre termes equivalents.

També existeixen diferències entre termes preferits i no preferits dins d'un mateix idioma i entre dos idiomes diferents, o sigui, un terme preferit d'A pot ser un terme no-preferit de B i al contrari. Els sistemes de mapeig poden utilitzar aquest raonament també com a un bon mètode per trobar les equivalències entre termes de diferents tesaurus.

AGROVOC	ASFA
----------------	-------------

Preferred Term (ES) ⁵⁷	Preferred Term (EN)	Preferred Term (EN)
Teledetección	“Remote Sensing”	Geosensing (<i>Use for remote sensing of earth surface from space. Before 1986 search</i>)
Non-Preferred Term (ES)	Non-Preferred Term (EN)	Non-Preferred Term (EN)
	Teledetection	Remote Sensing

Taula 37. Exemple de termes preferits i no preferits en diversos idiomes en un mateix tesaurus (AGROVOC) i en diversos tesaurus

Preferred Term	“Biological Cycle” (GEMET)	“Life Cycle” (AGROVOC)
Non-Preferred Term		UF: “Biological Cycle”

Taula 38. Exemple de termes preferits i no preferits dintre un mateix idioma, i entre diferents vocabularis

8.3.3. Semàntica

Alguns resultats previs conclouen que la primera gran diferència entre els tres vocabularis és la seva cobertura sobretot temàtica, ja que cadascun inclou punts de vista diferents. Per exemple, GEMET conté un punt de vista geogràfic que els altres no integren, i AGROVOC comprèn un vocabulari taxonòmic no existent en la resta de tesaurus. També s’identifiquen diferències culturals i semàntiques entre els termes, que es tradueixen en paraules amb un lèxic idèntic però amb significat diferent (casos d’homonímia).

Respecte a la cobertura, GEMET representa el tesaurus, generalment, amb menor cobertura amb els descriptors escollits i relacionats en l’àmbit de les ciències aquàtiques (vegeu taula 33). Igualment, els descriptors de GEMET solen tenir una visió molt més genèrica que els descrits per ASFA i AGROVOC, com per exemple els conceptes “hydrodynamics” o “wetlands”.

Una altra diferència significativa entre els tres tesaurus és el punt de vista de cadascun. Els resultats mostren que un mateix descriptor pot tenir punts de vista diferents, tal com es pot veure en la jerarquia del terme “Aquaculture” (Annex 8.1), on la jerarquia del terme a AGROVOC exposa conceptes referents a tècniques d’aqüicultura (exemple: *aquaponics*) i a tipus d’espècies amb les que es realitza l’aqüicultura (exemple: *Shellfish culture, fish culture, etc.*). ASFA proporciona un punt de vista més general i, a més, dona un punt de vista del concepte d’exemple relacionat amb el medi ambient (exemple: aqüicultura d’aigua dolça, salada, etc.). Finalment, GEMET, que ofereix una jerarquia més simple, s’acosta més al punt de vista d’AGROVOC. Un segon

⁵⁷ ES= Espanyol; EN=English o Anglès

exemple de diferències a nivell semàntic s'observa en el terme "surface water" (taula 35). ASFA i GEMET ofereixen un abast del terme a nivell geogràfic, i AGROVOC a nivell temàtic (fa referència) es refereix als recursos naturals com a matèria física de la terra. Un altre exemple en relació a la semàntica és el concepte "water masses", que ASFA segueix com a punt de vista dels ecosistemes aquàtics, ja que defineix els diferents tipus d'aigües incloses en el concepte de "water masses" (aigües costeres, llacunes, oceans, etc.), per contra, AGROVOC i GEMET el contempla com un punt de vista hidrològic i com un procés biològic, físic o químic.

8.3.4. Avaluació dels resultats: alineaments manuals

Resumint, es troben diversos nivells d'equivalència basat en la semàntica, la cultura i l'estructura. Tanmateix, com citen les directrius per a l'elaboració de tesaurus multilingües (IFLA Reports n°115, 2009 i ISO Standard 5964: 1985), els aspectes semàntics i culturals es refereixen al significat dels termes i la forma en què s'usen en una llengua o cultura determinada, i pel que fa a l'aspecte estructural, es refereix a les relacions jeràrquiques i associatives existents entre els termes.

A partir de les avaluacions manuals, els resultats corroboren els diferents graus d'equivalències existents en els tesaurus i els problemes associats en els seus alineaments. Tal com marquen les directrius trobem:

1. Equivalència exacta. Es refereix a un sinònim exacte. Per exemple: "climatic change" (AGROVOC) i "climatic change" (GEMET)
2. Equivalència inexacta. És un quasi sinònim i s'utilitza com a terme d'indexació, però amb una diferència en l'especificitat. Un exemple és l'ús del singular i plural, com el cas del terme "climatic change" utilitzat en GEMET i AGROVOC, mentre que ASFA emprà el terme en plural "climatic changes".
3. Equivalència parcial. És similar al quasi-sinònim, però un dels termes estrictament vistos denota un concepte més ampli o més estret. Aquests dos termes es poden arribar a considerar sinònims, però amb una diferència de punt de vista. Per exemple: el concepte "tree" en GEMET representa una gran planta llenyosa perenne amb un tronc diferent que dóna lloc a branques o fulles a certa distància de terra, en canvi, en AGROVOC es troba el terme amb plural "trees" que representa un conjunt de plantes llenyoses, etc.
4. Equivalència d'un terme individual a un múltiple. Aquesta és la situació més complexa, ja que es refereix principalment als termes compostos. ISO 2788 defineix tres casos principals i solucions:
 - a. El terme només existeix en un idioma, però es pot trobar en una altra llengua amb un o més termes específics que en combinació cobreixen el concepte més ampli.

English	Spanish
Culture	Cría + Cultivo

Taula 39. Exemple d'equivalència simple a múltiple a través del multilingüisme

- b. Similar a la situació a), però es refereix a termes compostos, no s'estudien en les mostres analitzades però es refereix a:
- i. *Semantic factoring* (ocorre en termes simples i compostos). Un terme es pot expressar en forma complexa i en forma d'elements simples, o bé, mitjançant la definició en què es troba la combinació de termes per representar-ne un altre. Per exemple: "biocontrol" (UF=ASFA) *versus* "biological control" (USE=ASFA i AGROVOC). ASFA defineix el terme com "l'ús d'organismes o virus per controlar els paràsits, males herbes aquàtiques o altres plagues", per la definició és conclou que un altre terme equivalent pot ser "biological pest control" emprat per GEMET.
 - ii. *Syntactical factoring*. Termes compostos acceptats com a indexats per separat, com ara: "Fisheries Management" (GEMET: "gestió de la indústria pesquera que inclou captura, processament i venda de peix"). Ambdós termes són indexats, també per separat, pel mateix vocabulari.
 - iii. *Factoring rules*. També es podrien trobar altres situacions, depenent de la construcció de cada tesaurus i les seves regles de recomanació utilitzades (vegeu directrius Aitchison), però en els casos estudiats no es troben altres casuístiques específiques definides pels propis vocabularis estudiats.
- c. Quan hi ha dos termes genèrics en un idioma i només un en un altre, la solució depèn dels criteris de la biblioteca o de la comunitat que procedeixi al mapeig dels vocabularis.

AGROVOC	
Aquatic Environment (EN)	"Ambiente acuático" (ES)
BT 331336 – Natural environment (EN)	BT 331336 - this term does not exist in ES
BT 2593 - environment (EN)	BT 2593 – "Medio ambiente" (ES)

Taula 40. Exemple de diferències d'equivalències entre termes genèrics en tesaurus multilingües

5. Termes no equivalents. Es poden veure diferències a nivell de cobertura quan un terme equivalent no està expressant entre un parell de vocabularis en les proves manuals realitzades. Però, no sempre es troben termes equivalents entre els tesaurus analitzats. En general, la majoria dels casos de "no equivalència" es troben entre GEMET i la resta

de vocabularis, fet evident, tenint en compte que és un vocabulari més petit i d'una d'especialització diferent. Veure Annex 8.1, conceptes com Aquatic Sciences, Ecosystem, Hydrodynamics, Hydrogeology, Hydrogeomorphology o Ocenaography.

8.4. Execució mapeig automàtic/semi-automàtic de vocabularis

L'execució del projecte o prototipus d'integració de vocabularis consta de tres fases: pre-anàlisi (pre-mapeig), alineacions i/o anàlisi de les alineacions (post-mapeig), i resultats finals (avaluacions de les alineacions resultants).

8.4.1. Fase de pre-mapeig

Aquesta fase consta de l'anàlisi estadístic bàsic dels tres tesaurus estudiats, ASFA, ARGOVOC i GEMET, a nivell de quantificació en l'ús de l'estructura (tipus de relacions i nivells) i de termes descriptius. Cal assenyalar que, a priori, ASFA és l'únic tesaurus del qual només es disposa de la versió SKOS en anglès, i, per tant, només es pot comparar amb la resta de tesaurus a nivell monolingüe, mentre que ARGOVOC i GEMET són avaluats a nivell multilingüe.

Tesaurus	SKOS Concepts	Top Concepts	Pref Labels (all lang)	Pref Labels (en)	altLabels (all lang)	altLabels (en)	Broader Relations	Narrower Relations	Skos Related	Skos Related pair
ASFA	6.267	1.737	6.267	6.267	3.746	3.746	4.530	4.785	22.290	0
GEMET	5.244	4	109.612	5.244	0	0	5.332	5.332	2.086	1.043
ARGOVOC	28.953	0*	329.932	146.791	146.791	10.119	29.007	28.999	28.357	14173

En aquesta primera anàlisi s'observa que ARGOVOC i GEMET tenen una estructura jeràrquica comparable (vegeu relacions *Broader* and *Narrower* de la taula 42). A nivell de profunditat però, GEMET és el vocabulari amb una major profunditat jeràrquica, com s'ha vist en exemples anteriors, atès que contempla més diversitat de nivells jeràrquics que la resta de vocabularis. En canvi, ASFA és el que té l'estructura més plana, ja que disposa de menys nivells de relacions.

Altres nivells diferenciadors en quant a l'estructura dels tesaurus són els "top concepts":

Thesaurus	# SKOS concepts	# TopConcepts
ASFA	6267	1737
GEMET	5244	4
ARGOVOC	28953	0

Taula 42. Anàlisi de Top Concepts

A nivell de "Top Concepts", el tesaurus amb major quantitat de descriptors és l'ASFA amb 1737 "top concepts", GEMET amb 4 i l'ARGOVOC no en té cap. Aquestes dades representen la diferència d'estructura entre els tres tesaurus

analitzats. GEMET, com s'ha explicat anteriorment, té una estructura amb més nivells jeràrquics, a més, agrupa tots els conceptes amb 4 grans àrees temàtiques interpretades pel sistema de mapeig com a “top concepts”, esdevenint així una classificació superior. Pel que fa a AGROVOC no apareix “top concepts”, perquè aquests són modelats pel sistema com a conceptes d'esquemes o “skos:ConceptSchemes”.

Un altre element diferenciador entre els tres tesaurus a nivell d'estructura és el tipus d'etiquetes que representen. ASFA, com s'ha apuntat, no té etiquetes de llengua com la resta de tesaurus, no obstant això, incorpora etiquetes alternatives (*altlabels*). En canvi, GEMET no n'inclou d'alternatives tot i tenir-ne. Pel que fa a AGROVOC, a diferència de la resta de tesaurus analitzats, integra etiquetes de llengua específiques tant a nivell de *preflabels* com d'*altlabels*.

Thesaurus	#prefLabels (all lang)	#prefLabels (en)	# alt labels (all Lang)	#altlabels (en)
ASFA	6267	6267 (0)	3746	3746 (0)
GEMET	109612	5244	0	0
AGROVOC	329932	28722	146791	10119

Taula 43. Diferències estructurals a nivell d'etiquetes alternatives i preferides

A nivell de relacions entre conceptes (RT), AGROVOC i GEMET són completament asimètrics i ASFA no ho és (més de 22 mil relacions però 0 relacions parells). A més ASFA té moltes més relacions de tipus associatiu que jeràrquic, això indica que no té molt nivell d'especificitat en les relacions.

Thesaurus	# skos relateds	# skos related pair
ASFA	22290	0
GEMET	2086	1043
AGROVOC	28357	14173

Taula 44. Diferències estructurals a nivell de relacions associatives

A nivell de relacions entre conceptes (RT), AGROVOC i GEMET són completament asimètrics, en canvi ASFA no ho és (més de 22 mil relacions però 0 relacions parells). Tot i això, ASFA disposa d'un grau més elevat de relacions de tipus associatiu que jeràrquic indicant el seu baix nivell d'especificitat en les relacions.

Thesaurus	# broader relations	# narrower relations	average broader depth	max broader depth
-----------	---------------------	----------------------	-----------------------	-------------------

ASFA	4530	4785	1.72	2
GEMET	5332	5332	5.43	12
AGROVOC	29007	28999	4.34	10

Taula 45. Nivell de profunditat jeràrquica dels tesaurus

A nivell semàntic, a priori, sense l'anàlisi qualitativa dels resultats, el sistema de mapeig automàtic indica les diferències quant a profunditat en les jerarquies o en els tipus de relacions semàntiques. Pel que fa a GEMET i AGROVOC els resultats tenen una jerarquia comparable, tot i que GEMET té la jerarquia més profunda a nivell de tipus de relacions (*broader - narrower* relations), mentre que ASFA té la més plana.

Thesaurus	# skos relateds	# skos related pair
ASFA	22290	0
GEMET	2086	1043
AGROVOC	28357	14173

Taula 46. Propietats objecte dels tesaurus

Cal afegir també que tots 3 tesaurus tenen notes d'abast (*scope notes*), però a nivell semàntic no són tan rics perquè no tots disposen de definicions. GEMET és l'únic amb un major nombre de definicions de termes (19.790), ASFA no en té, i AGROVOC inclou 30 definicions respecte als milers que incorpora GEMET

Thesaurus	#skos:definitions	#skos:definitions (en)	#skos:scope Notes	#skos:scopeNotes (en)
ASFA	0	0	1285	1285
GEMET	19790	4909	197200	9860
AGROVOC	30	15	13365	1183

Taula 47. Propietats descriptives dels tesaurus

8.4.2. Fase de mapeig

En la fase de mapeig automàtic diverses proves d'alineament posen a punt el sistema. Els tipus de mapeig que aquest suporta són *ExactMatch*, *CloseMatch*, *Broader*, *Narrower*, *Related*, *Unrelated* i *Unsure*. Un cop preparat el sistema, en aquesta fase de mapeig es realitzen dos experiments:

1. Mapeig monolingüe. Alineaments només per “etiquetes en anglès” (EN) dels tres tesaurus, perquè ASFA-SKOS només tenia la versió anglesa.
2. Mapeig multilingüe: alineaments per “totes les etiquetes” d’idiomes.

Els alineaments, en ambdós experiments (vegeu taula 48 i 49) (els alineaments), es donen per cada parell de tesaurus (T1-T2, T1-T3, T2-T3, T2-T1, T3-T1, T3-T2), i les alineacions es produeixen tant per a les etiquetes *preflabels* (principals/preferides) com les *altlabels* (alternatives).

La fase de mapeig està molt unida a la fase de post-mapeig, per això s’amplia en la secció següent on s’explica l’experiment en detall.

8.4.3. Fase post-mapeig: anàlisi d’alineaments

En aquest procés de post-mapeig s’avalua el nivell de precisió del sistema, és a dir, es valoren les alineacions correctes obtingudes pel sistema de forma automàtica, i es corregeixen manualment en cas que no ho siguin. Els experiments, el mapeig monolingüe dels tres tesaurus i el mapeig multilingüe d’AGROVOC i GEMET, es van realitzar en paral·lel. La metodologia a seguir en aquest procés ha estat la següent

1. Comptar amb un equip d’avaluadors (humans). En aquest cas, es va comptar amb 2 avaluadors humans, un coneixedor dels vocabularis i de les disciplines que formen part dels vocabularis, i un avaluador extern no expert en la matèria, però membre de l’equip de disseny del sistema de mapeig.
2. Escollir una mostra a avaluar. El sistema escolleix les mostres (alineacions). Per cada parell de tesaurus s’avaluen aproximadament una mostra de 200 termes avaluats i organitzats amb grups de 10. Com els tres tesaurus no compleixen les mateixes característiques a nivell de llenguatge, s’efectuen 2 tipus d’estudis. D’una banda, es realitzen tests per tots tres tesaurus a nivell monolingüe, mapejant sols les etiquetes que corresponen a l’anglès i, de l’altra banda, es realitza un estudi a nivell multilingüe pels dos únics tesaurus que tenen 10 llengües en comú, AGROVOC i GEMET.
3. Fixar una guia d’avaluació d’alineaments. És important tenir una guia d’avaluació per tal d’evitar la subjectivitat en el procés i establir uns criteris mínims d’estàndards d’avaluació, fet imprescindible si en les validacions hi ha més d’un avaluador humà. Per exemple, la ISO 6964 (tesaurus multilingües) assenyala fins a 5 nivells d’equivalències per avaluar “equivalències entre tesaurus”. En aquest estudi experimental però, s’agafen els criteris d’avaluació establerts per *vrije Universiteit Amsterdam*, ja que són els que s’apliquen també en el sistema de

mapeig. No es pot mostrar la guia emprada per temes de copyright, però, si clarificar que els criteris d'avaluació fan referència als següents tipus de mapeig: *exactMatch*, *closeMatch*, *broaderMatch*, *narrowerMatch*, *relatedMatch*, *unrelatedMatch* i l'opció *Unsure*.

Ambdós experiments realitzats es donen per a cada parell de tesaurus, realitzant alineacions de T1> T2 i viceversa: T2> T1. En totes les alineacions també s'han emprat etiquetes *prelabels* i *atlabels*, considerant que no són sensibles a majúscules i minúscules.

8.4.2.1. Estudi de mapeig de tesaurus a nivell monolingüe

En el primer experiment de mapeig monolingüe, només es realitzen alineaments amb etiquetes en llengua anglesa (EN).

Els resultats mostren una gran diferència entre els tres tesaurus tant per la seva estructura (com ja s'observava en la taula 42) com per l'ús dels conceptes. No obstant això, els de major percentatge d'afinitat són ASFA i AGROVOC, ambdós pertanyents al mateix organisme (FAO), fet que corrobora la seva semblança definida amb els experiments d'alineació manual. Entre ASFA i AGROVOC hi ha una superposició del 28%, mentre que GEMET i AGROVOC tenen un 22% de conceptes equivalents. Els tesaurus amb menor afinitat són ASFA i GEMET, els quals tenen un 12% de superposició.

Source	Target	Source mapped	Target mapped	# maps	Source %	Target %
GEMET	ASFA	649	607	649	0,12	0,10
ASFA	GEMET	576	581	583	0,09	0,11
GEMET	AGROVOC	1168	1087	1176	0,22	0,04
AGROVOC	GEMET	1004	1082	1093	0,03	0,21
ASFA	AGROVOC	1783	1625	1790	0,28	0,06
AGROVOC	ASFA	1822	1881	2013	0,06	0,30

Taula 48. Mapeig de tesaurus monolingüe

Una dada curiosa és que els resultats del mapeig dels parells de tesaurus (per exemple, T1-T2) comparant-los amb els seus viceversa (per exemple, T2-T1) no obtenen els mateixos resultats. Tot i no saber el perquè d'aquestes divergències es preveu un índex d'error del sistema de mapeig que explicaria aquestes diferències. En el cas dels alineaments entre GEMET-ASFA hi ha un error d'un 11%, entre GEMET-AGROVOC es contempla un error del 14%, i entre ASFA-AGROVOC hi ha només un del 2%. No obstant això, com el sistema de mapeig és semiautomàtic aquests errors són corregits a posteriori en una anàlisi manual dels alineaments.

8.4.2.2. Estudi de mapeig de tesaurus multilingüe

En l'estudi de mapeig multilingüe es realitzen alineaments per a totes les etiquetes d'idioma coincidents entre AGROVOC i GEMET, ja que tenen fins a 10 etiquetes d'idiomes comuns. En l'experiment s'inclou la versió anglesa del tesaurus ASFA, per tant, els estudis no són comparables i, com era d'esperar, els resultats mostren que el nivell de superposició divergeix. Tanmateix, en aquest anàlisi manual dels alineaments realitzat a posteriori, s'inclou només els tesaurus disponibles amb versió multilingüe (AGROVOC i GEMET), ja que és l'única forma possible d'avaluar el sistema d'una forma més precisa i objectiva.

	Target	Source mapped	Target mapped	# maps	Source %	Target %
GEMET	ASFA	694	646	697	0,13	0,10
ASFA	GEMET	646	694	697	0,10	0,13
GEMET	AGROVOC	2431	2420	14223	0,46	0,08
AGROVOC	GEMET	2420	2432	14502	0,08	0,46
ASFA	AGROVOC	1909	1850	2059	0,30	0,06
AGROVOC	ASFA	1850	1909	2059	0,06	0,30

Taula 49. Mapeig de tesaurus multilingüe (10 llengües)

Com ja s'ha exposat, en el cas de mapeig de tesaurus multilingüe es produeixen alineaments per mitjà de l'eina Amalgame entre GEMET i AGROVOC (amb 10 idiomes en comú), i es té en compte també el mapeig d'etiquetes alternatives (*altlbels*) i preferides (*prelabeleds*) per part del tesaurus AGROVOC.

Per a cada idioma, es mapegen les etiquetes de llengua dels conceptes del tesaurus d'origen (T1) i el tesaurus de destinació (T2).

En aquest procés, els mapejos realitzats per les 10 llengües comuns identificades pel parell de tesaurus d'estudi, es combinen. El mapeig resultant d'aquesta fusió consisteix en 2.323 mapejos provinents dels conceptes dels tesaurus d'origen.

En els primers resultats del mapeig automàtic de tesaurus (taula 48) s'observa clarament diferències significatives entre l'experiment de mapeig monolingüe vers el segon experiment de mapeig de tesaurus multilingüe (taula 49). Comparant resultats, s'observa l'augment considerable dels percentatges quant a la identificació automàtica de conceptes equivalents. Veure el percentatge de superposició des del tesaurus d'origen (*source*):

% Superposició Experiment 1		% Superposició Experiment 2	
GEMET-ASFA	12%	GEMET-ASFA	13%
GEMET-AGROVOC	22%	GEMET-AGROVOC	46%
ASFA-AGROVOC	28%	ASFA-AGROVOC	30%

Taula 50. Nivell d'equivalència o mapeig entre els tesaurus a partir de dos experiments

Recordar que l'experiment 1 fa referència al mapeig de tesaurus en versió monolingüe en anglès i l'experiment 2 fa referència al mapeig dels tres tesaurus en versió multilingüe (més de 10 llengües, a excepció d'ASFA que només inclou la versió monolingüe en anglès).

D'una banda, clarificar que no es poden considerar i comparar els dos experiments on s'inclou ASFA, ja que no hi és accessible la seva versió multilingüe en format ontològic, de manera que els experiments no es produeixen en les mateixes condicions.

D'altra banda, el segon experiment que inclou els tesaurus en les seves versions multilingües, presenta un increment en el nombre d'alineaments, per tant, és un primer indicatiu que el multilingüisme facilita una major recuperació de termes equivalents, o bé, pot contemplar algun tipus de relació d'equivalència (*exact match*, *close match*, *broader* or *narrower*). Aquesta hipòtesi es corrobora també en la fase d'avaluació manual dels alineaments.

Per posar un exemple, s'agafen uns descriptors concrets i s'observen grans diferències entre els experiments de mapeig multilingüe i monolingüe:

Mapeig Multilingüe	Mapeig Monolingüe
Plant production (AGROVOC). Url: http://www.fao.org/aims/aos/asc%23c_F (Mapeig AGROVOC-GEMET)	
Aquest concepte (skos:ConceptScheme) té un concepte superior ("Top Concept") amb 450 relacions (o "skos:Concepts"), amb un o més mapejos a 105 conceptes de totes les llengües, i 30 conceptes que tenen dos o més coincidències (<i>match</i>) En referència al concepte de "plant production", aproximadament un quart d'AGROVOC s'alinea (és alineat) amb els conceptes de GEMET.	42 conceptes mapejats només amb llengua anglesa.
NATURAL ENVIRONMENT, ANTHROPIC ENVIRONMENT (GEMET). Url:	

http://www.eionet.eu.int/gemet/supergroup/5499 (Mapeig GEMET-AGROVOC)	
Aquest concepte en té 9 d'específics (<i>Narrower terms</i>), i, a més, és l'ancestre de 1145 conceptes. D'aquests, 641 són mapejats amb conceptes d'AGROVOC, i 124 termes tenen com a mínim una coincidència.	204 mapejos només amb llengua anglesa.
SOCIAL ASPECTS, ENVIRONMENTAL POLICY MEASURES (GEMET). Url: http://www.eionet.eu.int/gemet/supergroup/2894 (Mapeig GEMET-AGROVOC)	
Aquest concepte té 2.338 conceptes per sota, 869 són mapejats amb conceptes de GEMET, i més de 103 tenen més de 2 coincidències o alineaments.	452 mapejos només amb llengua anglesa.

Taula 51. Resultats exemple de mapeig multilingüe versus mapeig monolingüe

A continuació, i per tal d'extreure conclusions més objectives, es realitza l'anàlisi qualitativa dels resultats de forma individual estudiant cada alineament produït a nivell manual pel sistema de mapeig automàtic. Aquest procés de validació es realitza dins el mateix sistema, ja que Amalgame com a sistema semiautomàtic permet l'avaluació dels resultats de forma manual. Un cop s'obtenen els resultats o alineaments, s'han de comprovar les equivalències entre termes produïdes pel sistema de forma automàtica i corregir-les en cas que no siguin correctes.

Aquestes anàlisis qualitatives del capítol permeten aconseguir els dos següents objectius considerats cabdals: conèixer i veure l'eficàcia dels sistemes de mapeig de vocabularis, i comprova com diverses disciplines s'agrupen i es complementen (poden ser agrupades i complementàries) per tal d'aconseguir distints objectius (vegeu taula 33).

En aquesta etapa, s'han estudiat també diversos mapejos, escollits aleatòriament pel sistema. Aquesta mostra aleatòria es divideix en estrats (*stratums*) o en grups de conceptes que s'analitzen individualment per un o diversos avaluadors humans. Un estrat "X" es defineix com un conjunt de mapejos pel qual hi ha una evidència exacta d'equivalència amb "X" llengües. Per exemple, l'estrat 2 d'una mostra aleatòria, proporcionada pel sistema, conté un mapeig entre conceptes per mitjà dels quals s'han mapejat etiquetes d'idiomes d'alemany i castellà, i aquestes, a la vegada, són coincidents amb els mapejos produïts per etiquetes d'idiomes francès i italià.

8.4.3. Avaluació dels resultats: procés de mapeig semi-automàtic

En aquesta fase d'anàlisi qualitativa s'ha arribat al consens que per poder extreure conclusions de forma més objectiva només es realitzaran anàlisis comparatives del segon experiment sobre el mapeig multilingüe (GEMET-AGROVOC). No obstant això, s'han realitzat anàlisis qualitatives sobre el primer experiment però no s'han inclòs en aquest estudi pels motius ja citats.

El sistema proporciona de forma aleatòria 10 grups (stratums) de resultats d'alineaments per analitzar. A continuació es mostren els resultats dels mapejos produïts:

Stratum	GEMET	AGROVOC	Matches
1	725	698	849
2	341	329	351
3	304	298	307
4	295	289	298
5	240	235	240
6	196	194	196
7	185	181	185
8	193	188	193
9	135	133	135
10	8	8	8

Taula 52. Mostra de resultats dels alineaments del mapeig multilingüe

Aquests resultats són analitzats per dos avaluadors humans, avaluador 1 (Carmen Reverté, autora de la tesis) i avaluador 2 (*Victor de Boer*, col·laborador d'aquest experiment i investigador de la Universitat d'Amsterdam, institució desenvolupadora del software de mapeig Amalgame).

En cadascun dels estrats es seleccionen aleatòriament mostres a analitzar. Per l'estrat 1 es seleccionen 40 alineacions, pels estrats del 2 al 9 es van extreure 20 alineacions i per l'estrat 10 es va extreure 8 alineacions. Veure Annex 8.2.

Després del procés d'avaluació dels alineaments, els resultats mostren un acord dels avaluadors d'un 68% respecte a l'establiment d'equivalències entre els termes (que els 2 avaluadors han tingut un acord del 68% respecte a l'establiment d'equivalències entre els termes). Veure els resultats en la taula següent:

Alineaments			
Stratum	Agreed	Not agreed	%
1	26	20	56,52174
2	7	18	28
3	15	5	75
4	16	3	84,21053
5	18	2	90
9	21	0	100

68,21192

Taula 53. Percentatge total d'acord entre els dos avaluadors

En les avaluacions qualitatives fetes (dutes a termes) pels avaluadors (Annex 8.3 i 8.4), els resultats conclouen que on hi ha un major consens en l'avaluació dels alineaments és en la identificació de termes amb equivalència exacta o "exact match", concretament, un 81% de coincidència:

Alineaments "Exact Match"			
Stratum	Agreed	Not agreed	%
1	4	4	50
2	6	12	33,33333
3	14	1	93,33333
4	16	0	100
5	17	1	94,44444
9	20	0	100
			81,05263

Taula 54. Percentatge d'acord d'"exact match" entre els dos avaluadors

Per tal d'apreciar les diferències de cada avaluador alhora de valorar els alineaments cal tenir en compte diferents aspectes.

D'una banda, un dels inconvenients principals a tenir en compte és que sols l'avaluador 1 té experiència en la temàtica dels vocabularis mapejats, per tant, valora un total de 10 estrats, en canvi, l'avaluador 2 no és expert en la matèria i valora 6 estrats (1-5 i el 9). Tot i això, en la comparativa de les avaluacions (taula 48 i 49) es té en compte el mateix nombre d'estrats avaluats. Cal puntualitzar que els avaluadors pertanyen a diferents països, i es podrien veure diferències culturals, però aquesta possible desigualtat es minimitza perquè l'avaluador 2 (no-expert), alhora de prendre decisions i avaluar un alineament s'ha de cenyir als recursos que ofereixen els propis vocabularis (per exemple: notes d'abast, definicions, etc.). Ara bé, si fos un expert en la matèria, probablement podria emetre altres tipus de judicis i fer interpretacions completament diferents. És a dir, la definició de les equivalències entre els termes es recolzarà molts cops en factors externs, com és el cas de l'establiment de les relacions de tipus *close match versus exact match*, sobretot, depenent dels avaluadors (persones, institucions o col·leccions) es podran incloure diferents punts de vista sobre la indexació adequada amb els vocabularis alineats.

D'altra banda, també cal tenir en compte l'ús dels alineaments, per exemple, no és el mateix voler alinear vocabularis diferents per integrar varies col·leccions en un mateix sistema i trobar una "equivalència exacta" entre els termes, que desenvolupar un tercer vocabulari per indexar una col·lecció determinada. Aquest darrer cas es dona per la manca d'un vocabulari especialitzat o per

l'escàs abast de l'àrea temàtica d'una col·lecció, llavors interessa identificar tant els tipus de relacions exactes com les de quasi-equivalència (*close match*).

A nivell global i amb l'objectiu de l'avaluació del propi sistema de mapeig es compara també les anàlisis dels avaluadors respecte el propi sistema (Annex 8.2). Els resultats dels alineaments (obtinguts per) de l'avaluació manual conclouen que el sistema de mapeig té una eficiència alta però millorable, ja que només s'han obtingut entre un 70% i 77% de relacions d'equivalència exacta. L'avaluador 1 identifica que un 30% de termes no tenen cap tipus de relació amb el termes alineats, i l'avaluador 2 identifica que un 23% dels termes no són equivalents. Veure resultats:

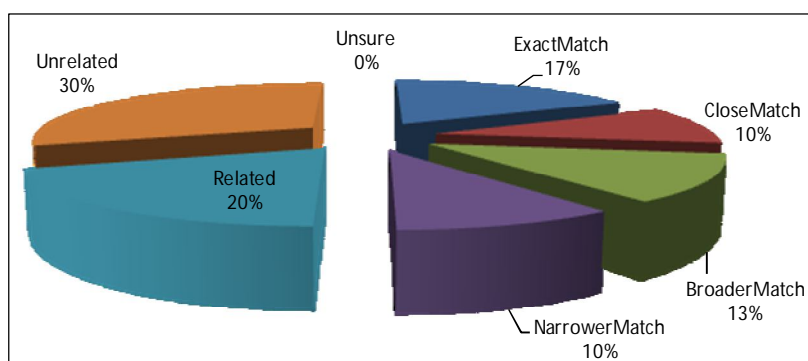


Figura 106. Resultats de l'avaluació manual per Avaluador 1 (10 estrats)

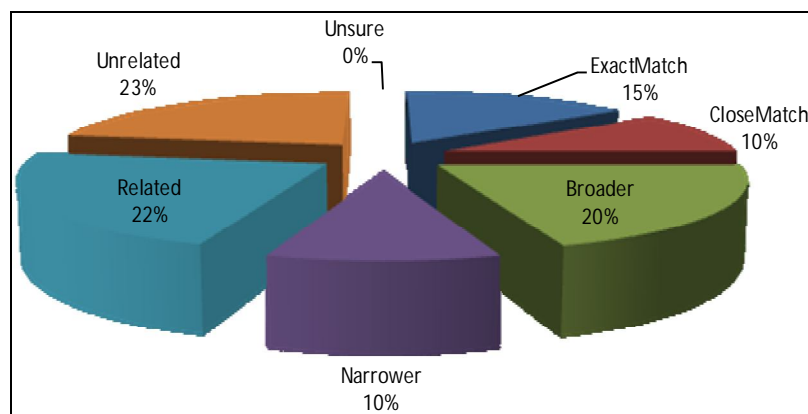


Figura 107. Resultats de l'avaluació manual per Avaluador 2 (6 estrats)

Amb l'anàlisi qualitativa (Annex 8.3), els resultats (obtinguts) han donat veu a moltes de les hipòtesis prèvies sobre la caracterització dels tres tesaurus i els nivells d'equivalència (cobertura, especificitat, jerarquia, etc.) que s'exposen a continuació:

- 1. Diferències estructurals entre els tesaurus.** Les més comunes es troben entre singulars *versus* plurals, o nivells de relacions. Veure exemples:

- AMINE (GEMET) *versus* AMINES (AGROVOC)
- GEMET: Mycorrhiza (Singular) *versus* AGROVOC: Mycorrhizae (Plural)
- GEMET: Organism (Singular) *versus* AGROVOC: Organisms (Plurals)

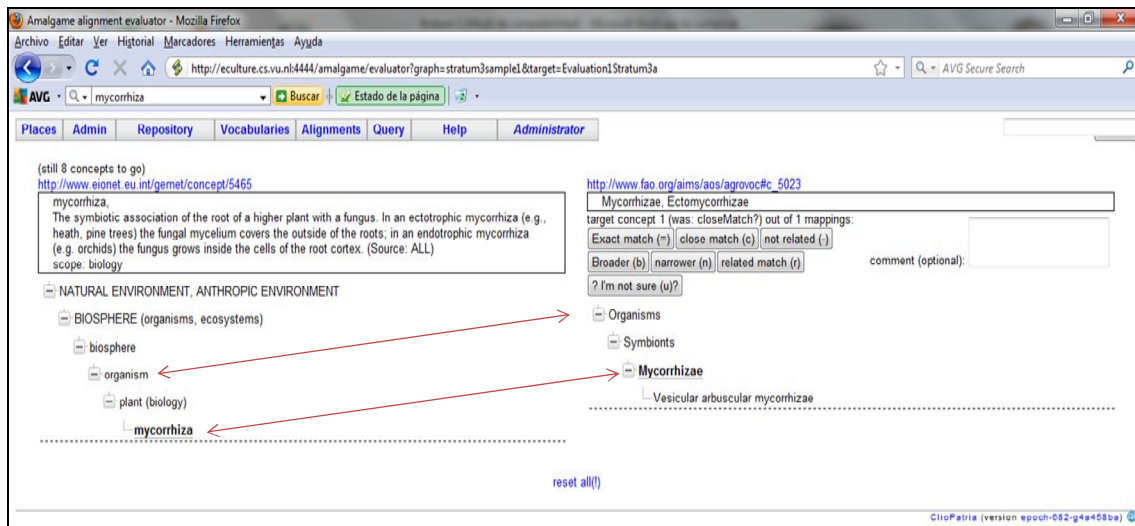


Figura 108. Exemple de diferència estructural entre tesaurus

2. **Diferències culturals:** nivell de complexitat semàntica i de cobertura entre termes diferents, etc.). Exemple:

- *Wood* (GEMET) *versus* *Forests* (AGROVOC): *wood*, segons la pròpia definició de GEMET, seria un NT de *Forests* (AGROVOC). Però per AGROVOC, *wood*, representa un tipus de material (fusta), per tant, estem veient que hi ha diferents nivells de cobertura temàtica. Veure exemple

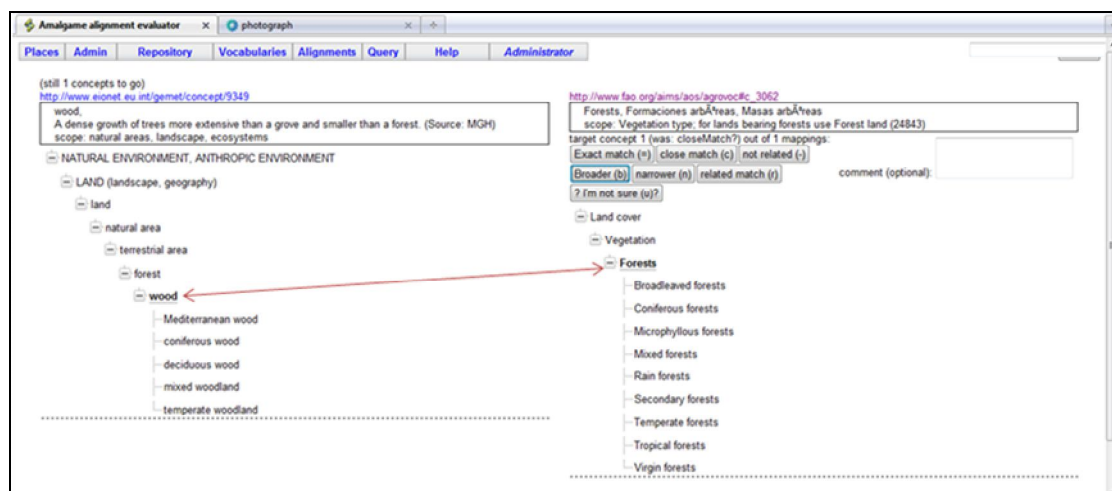


Figura 109. Exemple 1 de diferència de cobertura entre tesaurus a nivell cultural

- *Organic substance versus Organic matter.* A part que “matèria orgànica” és un concepte més genèric que “substància orgànica” (vegeu Annex 8.3), els nivells jeràrquics d’ambdós conceptes també mostren la diferència de cobertura temàtica d’ambdós tesaurus, on GEMET en aquest cas ofereix un vocabulari més ric. Veure exemple:

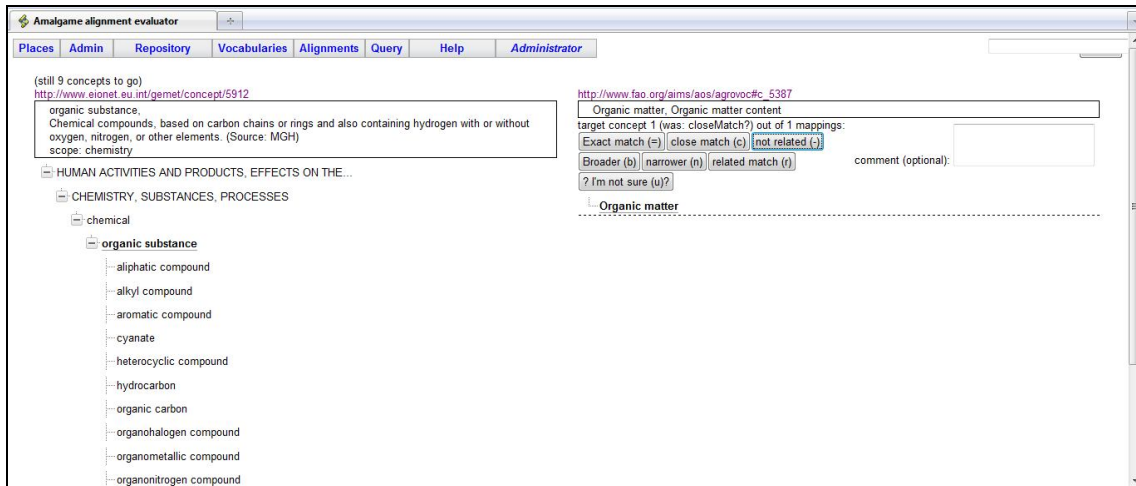


Figura 110. Exemple 2 de diferència de cobertura entre tesaurus a nivell cultural

3. Diferències de cobertura entre tesaurus: com s’exposa en l’exemple anterior (figura 110) existeixen estructures jeràrquiques amb diferents nivells (hi ha vocabularis amb major profunditat de relacions jeràrquiques que d’altres i/o major especificitat)

- *Climatology (GEMET) versus Climatology (AGROVOC):* GEMET mostra una major especificitat sobre el concepte, AGROVOC no contempla termes específics:

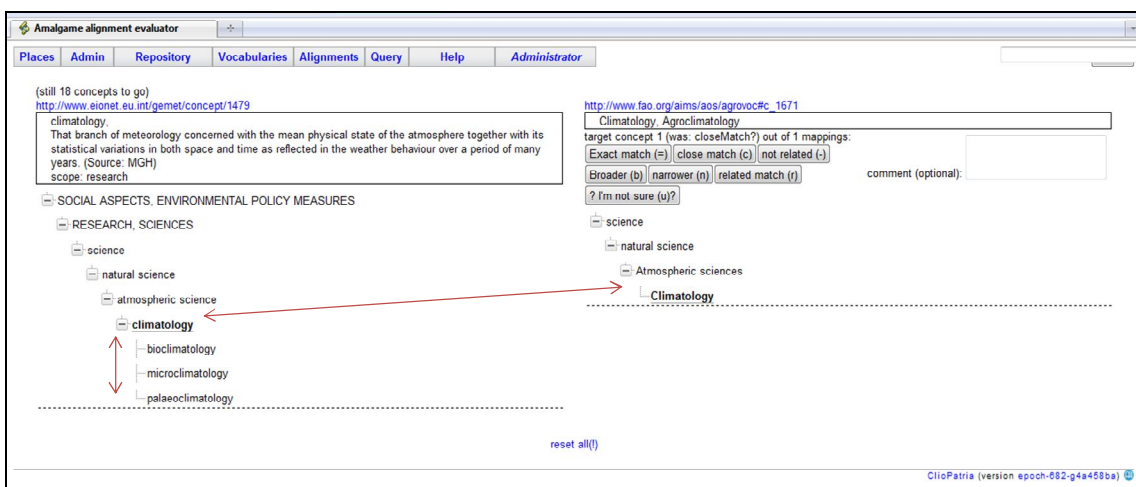


Figura 111. Exemple de diferència de cobertura a nivell jeràrquic entre tesaurus

4. Problemàtiques d'irrellevància. Encara que no amb xifres molt elevades també es troben casos d'alineacions identificades pel sistema com a equivalents, però que en realitat no ho són. El tipus de relació que s'estableix és de "No-equivalència":

- *Justice* (GEMET) versus *Justicia* (AGROVOC)

D'una banda, aquest alineament és fruit de l'equivalència entre els termes per l'alineament d'etiquetes multilingües, perquè el terme "Justice" (dret) en anglès és idèntic al terme en castellà "Justicia" que AGROVOC l'identifica com a nom d'una planta. D'altra banda, aquest és un exemple on cal millorar el sistema de mapeig. Per exemple, es podria aplicar un algoritme addicional que diferencies tipografia i semàntica, per diferenciar dos conceptes que s'escriuen gairebé igual però amb (que tenen) significats diferents, o bé, especificar que entre la resta de vocabularis, sols AGROVOC inclogui un glossari taxonòmic.

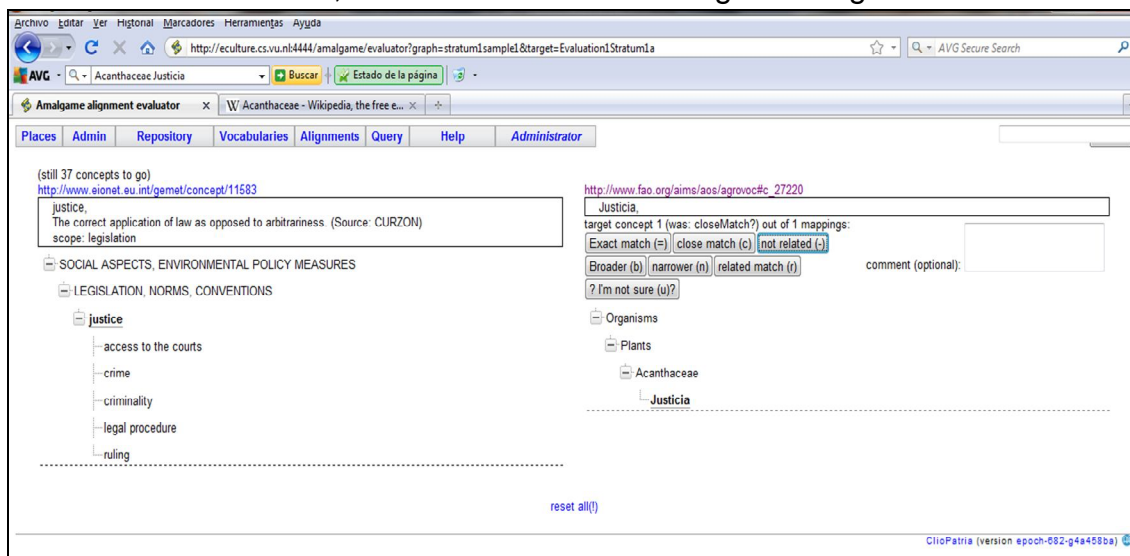


Figura 112. Exemple de termes no equivalents

Resumint, exposades totes les dificultats associades amb la precisió d'un sistema de mapeig, convé ressaltar que els resultats obtinguts de l'experiment de mapeig multilingüe, gràcies a l'ús de múltiples fonts d'evidència, com són els vocabularis multilingües, han permès millorar la precisió dels sistemes de mapeig.

Tal i com es demostra en els resultats presentats (taula 50), l'alineament d'un concepte amb els seus termes homòlegs en altres idiomes permet la identificació d'un major nombre de termes equivalents. Per aquest motiu, en el primer experiment de mapeig monolingüe no s'aconsegueixen els mateixos alineaments que en l'experiment de mapeig multilingüe. En aquest darrer experiment, amb l'ús de 10 etiquetes d'idiomes diferents, s'han recuperat fins a un 24% més d'alineaments o tipus de relacions semàntiques. A més, un cop feta l'avaluació qualitativa d'aquests alineaments, es corrobora que el multilingüisme augmenta significativament la precisió dels sistemes de mapeig

perquè en la majoria dels casos permet identificar els termes amb una major precisió (vegeu múltiples exemples en Annex 8.3).

Els resultats de la part experimental d'aquest capítol conclouen que cal millorar els mètodes d'alineaments emprats en els sistemes de mapeig des de diferents punts de vista, per tal d'incrementar l'eficiència i precisió dels sistemes de mapeig. Primerament, és necessari incloure avaluadors humans en el flux de treball del sistema de mapeig, i intentar que la seva valoració sigui, amb l'ajut de les guies d'avaluació comunes, el més objectiva possible. Es demostra que l'ús d'avaluadors humans també ajuda a millorar la qualitat dels alineaments, i per tant, a perfeccionar els sistemes de mapeig i a augmentar la seva eficàcia. A posteriori, és important estudiar els vocabularis que es volen integrar o alinear per tal de, si s'escau, incorporar nous algoritmes, i millorar els mètodes d'alineament (estructures, diferenciació de plurals i singulars, si contenen definicions i/o notes d'abast, etc.). Finalment, també s'ha demostrat la necessitat de comptar amb vocabularis multilingües, o bé, amb aplicació d'algoritmes de correspondència en les etiquetes d'idiomes incloses en els vocabularis, perquè milloren la qualitat dels alineaments produïts automàticament, tal i com s'ha mostrat en els resultats exposats.

8.4.4. Discussió

Els resultats mostren que els sistemes de mapeig ajuden a la convergència entre tesaurus monolingües i multilingües els quals comparteixen disciplines afins tot i les diferències existents en quant a estructura i contingut. Aquesta interoperabilitat es dona gràcies a la conversió dels tesaurus en format SKOS.

D'una banda, en l'alineació manual dels tres tesaurus analitzats (taula 33 i 34), els resultats mostren diferències estructurals, culturals o de cobertura (ús de termes diferents, nivells d'especificitat i exhaustivitat diferents) i de forma (ús d'etiquetes diferents per a la definició dels termes). Aquestes dades també es corroboren en la fase de pre-mapeig automàtic. En els resultats de la taula 42 fins la 47 s'identifiquen aquestes diferències:

1. Respecte a l'estructura dels tesaurus, tots tres són molt distints tant pel que fa a l'extensió com per la seva pròpia estructura jeràrquica. Quant a la mida, AGROVOC contempla 28.953 termes, per contraposició a ASFA i GEMET amb 6.267 i 5.254 termes respectivament. A nivell d'estructura també es troben divergències, ASFA té un major nombre de descriptors "top concepts" (1.737), en canvi, GEMET en té 4 i AGROVOC no en té cap, perquè passa directament a la descripció de les etiquetes preferides (*preferred labels*). ASFA disposa d'una estructura més plana i només inclou 4 nivells jeràrquics que pegen d'aquests "top concepts". GEMET amb una estructura més complexa, que classifica per grups i temes abans del desplegament dels conceptes preferits, el sistema de mapeig

identifica 4 grups temàtics com a “top concepts”. De tota manera, aquestes diferències no causen errors al sistema de mapeig per a identificar relacions d'equivalència entre els termes.

2. A nivell lèxic o d'etiquetes. Malgrat la distància existent entre els tesaurus, també a nivell de les etiquetes de representació dels conceptes, el sistema de mapeig automàtic pot identificar igualment els descriptors equivalents. Recordar que ASFA i AGROVOC (ambdós de la FAO) empen diferent terminologia que GEMET per a la descripció de termes (exemple: BT *versus* “Broader term”). El sistema de mapeig automàtic (també) distingeix etiquetes simples i compostes que hi ha en els tesaurus (exemple: etiquetes monolingües *versus* multilingües). A més, el sistema respon correctament a la manca d'etiquetes alternatives del tesaurus GEMET (USE FOR, UF, etc.).
3. A nivell semàntic, quan s'alineen conceptes equivalents també s'observen les diferències de cada tesaurus, diverses cobertures i punts de vista que engloben dos conceptes aparentment similars, i, en aquest punt es fa evident l'eficàcia del sistema automàtic del mapeig per trobar assignacions lèxiques simples i fàcils. Ara bé, en casos d'alineaments ambigus, homonímies i d'altres on cal aportar coneixement humà és on el mapeig manual (avaluacions humanes) és necessari, i en conseqüència, ajuda a millorar els alineaments, el mapeig automàtic i semiautomàtic.
4. Respecte al multilingüisme, els resultats dels tesaurus multilingües milloren l'eficàcia dels esforços dels alineaments, ja que el mapeig amb les etiquetes multilingües dona com a resultat un alineament de major qualitat.

El mapeig d'aquests tres tipus de tesaurus convertits en ontologies, amb punts de vista dispars faciliten la indexació i la recuperació de la informació de col·leccions heterogènies d'una institució (biblioteques i/o centres de documentació en l'àmbit de les ciències aquàtiques). En l'actualitat, la majoria de les biblioteques especialitzades en aquesta àrea afirmen que solen emprar més d'un vocabulari en la indexació de les seves col·leccions, per tant, les ontologies poden millorar els sistemes d'informació actuals aportant una major flexibilitat i comprensibilitat per a una recuperació de la informació més efectiva.

Els resultats sobre el mapeig i integració de diferents vocabularis multilingües de disciplines afins (poden facilitar) comporten una futura col·laboració entre biblioteques europees i internacionals en l'àmbit de les ciències aquàtiques i marines. Aquesta col·laboració es pot traduir amb la construcció d'un sistema o portal d'informació especialitzat que englobi les col·leccions de cadascun dels

membres, de manera que aquestes siguin accessibles i recuperades amb les diferents ontologies.

Aquests resultats poden ajudar a la millora o extensió d'un vocabulari especialitzat en ciències aquàtiques com és ASFA.

La principal limitació del present estudi, ha estat no poder incloure els tests de mapeig multilingüe amb el tesaurus ASFA del que només s'ha obtingut la versió monolingüe en anglès en format SKOS. Tenint en compte el mapeig automàtic per accedir a totes les parts d'un vocabulari de forma automàtica, i el mapeig multilingüe per recuperar termes alineats amb més eficàcia que el monolingüe, no s'ha pogut explorar en profunditat la hipòtesis que ASFA, com a tesaurus, no cobreix àmpliament alguns àmbits actualment representats de d'un punt de vista molt genèric (com ara, l'agricultura, medi ambient, ecologia o canvi climàtic).

Les dades obtingudes del mapeig dels vocabularis multilingües, que faciliten els alineaments entre els conceptes, obren una línia de recerca per millorar els sistemes de mapeig amb el multilingüisme. Convé subratllar que seria interessant estudiar si a part de l'exhaustivitat pot augmentar també la precisió i la rellevància dels sistemes de mapeig automàtic o semiautomàtics gràcies a l'ús de vocabularis multilingües.

8.5. Conclusions

Els resultats corroboren que les tècniques d'alineació són processos d'alt valor afegit per a l'enriquiment de vocabularis ja existents, és a dir, poden ser útils per a l'actualització d'un vocabulari i/o per crear un tercer vocabulari, ja sigui perquè l'existent hagi pogut quedar obsolet, o perquè no existeixi cap tesaurus interessant del domini específic.

A més, tant en els tests realitzats com en d'altres estudis existents (Jean-Mary, et al., 2009), s'ha demostrat que utilitzar un tesaurus de domini específic per a l'alineació d'ontologies especialitzades és un avantatge, ja que millora la precisió del sistema de mapeig i la recuperació de la informació en qualsevol entorn (bases de dades, *Word Wide web*, etc.).

La intervenció humana demostra que és un gran avantatge en l'avaluació manual, el mapeig semiautomàtic (avaluació i verificació d'equivalències dels alineaments produïts) i també en el mapeig automàtic, ja que permet la rectificació del propi sistema de mapeig per millorar la qualitat dels alineaments, i en alguns casos, introduir o corregir d'algoritmes (per exemple, en l'avaluació humana dels alineaments es detecta la identificació de plurals i singulars, i l'ús de vocabulari vulgar *versus* vocabulari científic, etc).

Cal ressaltar també, com l'ús de vocabularis multilingües (i/o etiquetes ontològiques d'idiomes) millora la qualitat dels alineaments, perquè ajuden a trobar un major nombre d'equivalències entre els termes emprats en diferents vocabularis. A més, fins a la data dels experiments no existia literatura científica al respecte, i encara en l'actualitat n'hi ha poca (Sphor, Hollink i Cimiano, 2011; Moonet 2010-2012). Aquest fet mostra una línia de recerca poc desenvolupada en l'avaluació dels sistemes de mapeig que cal investigar per tal de donar suport al futur desenvolupament de sistemes de mapeig i, paral·lelament, recolzar la millora de la interoperabilitat semàntica en general. En conseqüència, es considera continuar investigant aquesta línia de recerca per tal de donar suport al desenvolupament futur de sistemes de mapeig i paral·lelament ajudar en la millora de la interoperabilitat semàntica en general.

Finalment, es conclou que l'ús d'ontologies, independentment de la seva diversitat d'esquemes o KOS (diferències semàntiques, lèxiques i d'estructura), permet integrar múltiples models de dades de diferents punts de vista i també en sistemes d'informació existents (tal i com es mostra en la part experimental, veure taula 33). En canvi, en les bases de dades tradicionals, tot i que s'integren diversos KOS, no hi ha aquesta flexibilitat per oferir un model de dades al mateix temps (*Gangemi, 2004*). En conseqüència, aquests tipus de models de KOS i la seva tecnologia d'aplicació (web semàntic) és una línia de recerca que s'ha convertit en tendència els darrers anys, però on s'ha comprovat que encara hi ha molta recerca futura a realitzar.

9. CONCLUSIONS GENERALS

Com a colofó a la tesis doctoral, en aquest capítol de conclusions es confirma la consecució de les hipòtesis de partida i es finalitza detallant les principals aportacions d'aquest treball al camp d'estudi híbrid entre l'organització del coneixement i la recuperació de la informació en els sistemes d'informació, així com també es perfilen les noves línies d'estudi que s'obren a través de les investigacions realitzades.

Cada conclusió desenvolupada, respon a un punt de partida que s'ha plantejat com a pregunta i s'ha treballat i respost al llarg de cada capítol.

1. Quins són els antecedents, característiques i tendències dels sistemes d'informació de qualitat de tipus distribuït i en un entorn científic?

En el primer capítol (Evolució dels sistemes d'informació de qualitat), a través de l'estudi de l'estat de l'art sobre els sistemes d'informació de qualitat i de tipus distribuït, es conclou que el futur dels sistemes d'informació de qualitat passa per donar accés a la informació sense cap tipus de barreres socials o lingüístiques. Els sistemes s'adapten als usuaris i permeten la seva participació de forma activa en la creació de coneixement dins el propi sistema d'informació, i en la cerca i recuperació de la informació.

La tendència actual és anar cap al disseny de sistemes d'informació semàntics, ja que donen la possibilitat de cercar informació mitjançant diferents vocabularis controlats (ontologies o altres sistemes d'organització del coneixement convertits a ontologies) i permeten la indexació, cerca i navegació creuada entre diferents col·leccions, i alhora, milloren la qualitat de la recuperació de la informació. El gran avantatge d'aquests sistemes semàntics és la flexibilitat, estandardització i interoperabilitat que ofereixen als sistemes d'informació. Les dades contingudes en aquests tipus de sistemes es converteixen en una xarxa de metadades que permeten integrar i interpretar les dades internes i externes del web, amb de llenguatges de descripció de la informació estàndards (RDF, OWL, SKOS, etc.) i formant part del moviment *Linked Data*.

Aquesta tecnologia ha estat un gran descobriment per als sistemes d'informació de biblioteques i centres de documentació i viceversa. D'una banda, les biblioteques són llocs on es concentra informació rellevant (avaluada, filtrada, etc.), descrita amb formats estàndards (MARC, Dublin

Core, etc.) i adaptables a la web semàntica, per tal que el coneixement quedi obert a altres aplicacions web. D'altra banda, la connectivitat, la portabilitat i la integració de les dades fa possible la retroalimentació entre diferents sistemes d'informació facilitant així, que els sistemes d'informació (de biblioteques, centres de documentació, etc.) puguin enriquir-se d'altres dades web gràcies a l'intercanvi d'informació que permet el *Linked Data*.

El gran avantatge del *Linked Data* és la interrelació de diferents tipus de dades a través de diversos protocols i identificadors web. La comunitat web (W3C) ha estès la noció del protocol de transferència de l'hipertext (URL's) i el localitzador de recursos uniforme (URL's) no només com a identificadors de documents, sinó també com a identificadors de conceptes, llocs (físics o electrònics) i objectes. La presència d'aquests identificadors és un dels punts claus per a la integració de la informació heterogènia dins els sistemes.

A nivell de sistemes d'informació de tipus científic, la potencialitat d'aquests tipus de sistemes és molt gran, ja que permet la interrelació de diferents tipus de dades (atmosfèriques, bibliogràfiques, taxonòmiques o d'espècies, etc.) independentment de la seva localització (diverses bases de dades o organitzacions) que tinguin algun interès comú (temàtica, localització, autoria, etc.). Conseqüentment, l'usuari final pot accedir a la informació desitjada amb un únic punt d'accés.

A nivell tècnic, però, aquesta tecnologia presenta també algunes problemàtiques encara no resoltes. Cal aconseguir la coherència entre la varietat de tipus de dades enllaçades (llocs, persones, llengües, categories, etc.), així com també l'heterogeneïtat dels recursos i la inconsistència de les dades (procedència de les dades de diferents tecnologies i formats, per exemple RDF i XML o MARCXML i OWL o SKOS, etc.).

La tendència general per tal de resoldre aquests problemes tècnics és anar cap a la convergència de normes àmplies que s'adaptin a l'intercanvi dels diferents tipus de metadades existents en els sistemes d'informació i de la web en general. És a dir, anar cap a la universalització i/o ús de formats estàndards emprats en la web semàntica. De fet, els sistemes d'informació de qualitat actuals no es limiten a un públic especialitzat, sinó que hi ha un gran nombre d'iniciatives que se centren en usuaris i col·leccions de tipus universal (catàlegs *online* d'accés públic, sistemes d'informació patrimonial o biblioteques/arxius nacionals digitals, etc.)

2. Quines són les característiques dels sistemes d'informació distribuïts de qualitat?

L'anàlisi de casos ha permès concloure que aquest tipus de sistemes són plataformes úniques per a l'intercanvi d'informació dins un àmbit cooperatiu tant a nivell inter-institucional (cooperació entre diferents institucions i/o regions, amb àmbits de recerca multidisciplinaris però amb un interès comú de difusió del coneixement científic: LIBRIS, VASCODA, Intute, etc.) com a nivell disciplinari (diferents institucions de diferents països o no que tenen un interès temàtic en comú: Europeana, AustLII).

L'estudi comparatiu dels dos models de sistemes distribuïts d'informació de qualitat, tipus passarel·la temàtica i tipus sistema semàntic, comparteixen una arquitectura de la informació similar, i ambdós utilitzen els vocabularis controlats en la indexació, cerca i navegació creuades, com a eines d'interoperabilitat entre diferents col·leccions i/o sistemes d'informació.

La principal diferència entre els dos tipus de models de sistemes és que els sistemes semàntics se centren principalment en la contextualització del contingut a través de les dades entrelaçades, mentre que la resta de sistemes ofereixen poca relació amb el contingut que mostren (text, web, dades, vídeo, so, etc.) i amb el context del propi contingut (informació sobre objecte, persones, llocs, temes relacionats amb la informació recuperada). No obstant això, els sistemes tipus passarel·la temàtica són sistemes d'alta qualitat perquè segueixen estrictes controls de producció. Per exemple, alhora de seleccionar els recursos web que formaran part de la seva col·lecció apliquen protocols d'anàlisi de contingut. Per contra, els sistemes d'informació semàntics, ara per ara, no inclouen aquests tipus de control, tot i que la procedència de les dades (biblioteques, museus, centres de documentació, etc.) ja té implícita un control de qualitat (per exemple: Europeana).

Adicionalment, els sistemes semàntics, gràcies a la seva estandardització i interoperabilitat, faciliten la gestió i desenvolupament de software, la col·laboració entre diferents institucions i la interrelació entre diferents tipus d'informació i documentació (*Linked Data*). Per tant, són eficients amb grans maneigs de dades distribuïdes i la seva visualització. En contraposició, la gestió de la informació en la resta de sistemes és molt més feixuga i costosa, i la prova està en que molts dels sistemes tipus passarel·la temàtica han anat desapareixent amb el temps i les innovacions s'orienten cap a models de sistemes semàntics.

3. Quines són les característiques essencials dels sistemes d'informació de qualitat, distribuïts i especialitzats en ciències aquàtiques?

3.1. La indexació i navegació creuada és imprescindible per aconseguir sistemes d'informació de qualitat?

La indexació i navegació creuada és el resultat del procés de mapeig entre vocabularis controlats emprats en un sistema d'informació distribuït de tipus passarel·la temàtica. Convé recalcar que els sistemes d'informació distribuïts semàntics també utilitzen el mapeig de vocabularis en la cerca i recuperació de la informació, i és per això que aquests tipus de sistemes apunten a ser els predecessors de les passarel·les temàtiques.

El mapeig entre vocabularis es considera com una gran eina ecològica que permet la consistència i la interoperabilitat entre diferents sistemes d'informació distribuïts, i per tant, millora la recuperació de la informació de manera que el resultat final és un sistema distribuït d'informació de qualitat. Els usos ecològics que proporciona el mapeig de vocabularis i que la recuperació de la informació sigui de qualitat són:

- **La conversió entre diferents sistemes** (OPACS, bases de dades, serveis d'Internet, etc.) per incorporar registres d'una estructura local a una passarel·la temàtica (sistema distribuït), on també es produeix una conversió entre els esquemes de classificació (indexació creuada).
- **Suport per a la traducció de categories i termes entre diferents llengües.** Es pot representar la cobertura de termes en diferents llengües i cercar l'equivalència entre termes (*dimensió lingüística, Capurro*).
- **Provisió de la Navegació Creuada** entre diferents serveis els quals poden tenir els seus propis sistemes de classificació, així com també la provisió de la Cerca Creuada entre diferents sistemes.
- **Permet la interoperabilitat** entre diferents sistemes (esquemes de classificació, regles de catalogació, metadades, etc.).

No obstant això, els resultats dels diferents casos d'estudi analitzats mostren que la metodologia del mapeig té alguns punts de conflicte, com és la discussió que hi ha entre la utilització d'esquemes de classificació i/o la utilització de matèries (tesaurus o llistes d'encapçalaments de matèries) en la indexació. D'una banda, l'ús de classificacions estàndards com DDC o CDU, juntament amb l'ús de metadades Dublin Core, són considerats unes de les eines més eficaces en el mapeig de la informació continguda en els sistemes d'informació, per la seva fàcil aplicació. D'altra banda, aquesta hipòtesi és una línia de recerca que cal estudiar i avaluar en profunditat per tal de veure com afecten aquests processos en la recuperació de la informació científica (especialitzada),

ja que els sistemes de classificacions inclouen un punt de vista més generalista i multidisciplinari. Caldria analitzar també les estructures de cerca i navegació per poder concloure si els vocabularis controlats són realment els més eficaços o no, perquè hi ha exemples on no s'utilitzen esquemes de classificació com a eines d'indexació i navegació creuades, com per exemple SOSIG (ús de thesaurus) o el projecte CARMEN (llistes d'encapçalaments de matèries).

3.2. Són els estàndards i formats l'element clau de la interoperabilitat dels sistemes d'informació especialitzats?

Els estàndards en sistemes d'informació digitals (distribuïts o no) serveixen principalment per proporcionar interoperabilitat entre la informació continguda en el sistema d'informació i els serveis d'informació que siguin integrats en una plataforma digital.

Es poden trobar formats i estàndards en tots els nivells d'interoperabilitat semàntica (estructures de dades, dades categòriques i dades factuais). Els estàndards generalment també es poden ajuntar en quatre grups: estàndards d'identificació i localització (exemples de formats: DOI, URI, etc.), estàndards de llenguatge de marcat (HTML, XML, RDF, MARC, etc.), estàndards de protocols d'intercanvi d'informació (Z39.50, SRU, OAI-PMH, etc.) i estàndards de metadades (Dublin Core, MODS, METS, etc.). Principalment s'estudien els estàndards del grup 2 i 4 perquè són els que fan referència al nivell de descripció i recuperació de la informació en els sistemes digitals, i la resta d'estàndards responen a la tecnologia pròpiament (protocols d'intercanvi d'informació, identificadors, etc.).

El context de les metadades és molt heterogeni i poden existir tants esquemes de metadades com àmbits d'aplicació. En l'àmbit de les ciències aquàtiques i marines, els estàndards i formats d'estructures de dades més utilitzats són els formats més habituals coneguts en l'àmbit de les biblioteques especialitzades, els formats MARC i Dublin Core, no obstant això, s'identifiquen fins a uns 62 estàndards d'esquemes de metadades relacionats en l'àmbit de les ciències aquàtiques. Ambdós formats internacionals identificats com els referents (MARC i Dublin Core), normalment, no es troben en el mateix sistema d'informació perquè són creats amb objectius diversos i cadascun inclouen col·leccions de diferent tipus (catàlegs, repositoris digitals, etc.). Els estàndards coneguts en el camp bibliotecari sempre inclouen la premissa de la interoperabilitat, però, no estan preparats per als nous usos demandats en l'escenari actual del web semàntic, tanmateix, les seves característiques com a estàndards els permeten ser adaptables als nous sistemes d'informació.

Actualment s'està produint un canvi en els entorns digitals, on es desenvolupen sistemes d'informació més interoperables i flexibles que permeten, al mateix temps, l'ús de diverses col·leccions i diferents formats de descripció i accés a la informació. L'escenari global que envolta la informació digital, presenta a un usuari-creador d'informació, és a dir, l'usuari no només té accés i consum de la informació de forma universal a través d'Internet, sinó que també és el gestor i creador de la informació. Per tant, el creixement de la informació és constant, desmesurat i conseqüentment augmenta la informació desestructurada i heterogènia.

La tendència actual i els nous sistemes d'informació s'encaminen cap a la construcció d'una web més estructurada (web semàntica o web 3.0), amb nous estàndards i formats semàntics (RDF, OWL, SPARQL, SKOS, etc.) que permeten una recuperació de la informació de qualitat, aportant un major significat a la informació continguda en la xarxa i essent més comprensible per a les màquines. En el camp de les biblioteques també canvia el concepte de col·lecció com a un conjunt d'objectes, no necessàriament físics, ja que una col·lecció potser un conjunt de metadades que apunta a recursos d'informació distribuïts de diverses col·leccions seleccionades en la xarxa per la rellevància que ha definit la pròpia biblioteca digital (per exemple, similitud temàtica o cooperació bibliotecària), o a través de la consulta/interrogació d'un usuari a un sistema d'informació online.

Es conclou, que per tal d'assegurar la interoperabilitat dels sistemes d'informació digital, és primordial utilitzar metadades elaborades conforme a estàndards d'ús comú en el seu àmbit d'aplicació, que siguin flexibles i interoperables. Conseqüentment, els recursos descrits mitjançant diferents esquemes poden ser recuperats independentment de si pertanyen a diferents sistemes d'informació i/o responen a informació distribuïda en el World Wide Web.

3.3. La Interoperabilitat semàntica pot fer front a la indexació, cerca, navegació, i accés a la informació heterogènia del camp de les ciències aquàtiques?

Els sistemes d'informació distribuïts i de qualitat, per tal de garantir la interoperabilitat semàntica, han de tenir en compte la cooperació i coordinació entre les diferents institucions implicades (proveïdors de dades), per augmentar la interoperabilitat global de les dades disponibles (dades internes o dades externes accessibles al web).

El repte més important per fer front a la interoperabilitat semàntica, des de la biblioteconomia i la documentació, passa per la normalització, la estandardització i l'ús dels sistemes d'organització del coneixement (KOS).

La normalització o estandardització va des de la definició dels termes, utilització de normes, vinculació de dades i metadades fins a seguir senzilles instruccions que faran que les dades del sistema siguin accessibles per a molts més usuaris, proporcionant informació de qualitat, etc.

Pel que fa als sistemes d'organització del coneixement (KOS), cal garantir la interoperabilitat entre ells, amb l'ajut dels usuaris i a través de l'adaptació de polítiques i normatives referents al manteniment dels vocabularis controlats, i al contingut dels sistemes d'informació corresponents.

En la majoria dels casos estudiats, aquestes premisses encara no s'han aconseguit fins ara, i per aquesta raó, és necessari seguir treballant amb els desenvolupaments del web semàntic com és el cas del mapeig de vocabularis controlats, assenyalat com a un factor clau d'interoperabilitat semàntica per als desenvolupaments presents i futurs de sistemes d'informació semàntics. Cal destacar l'important funció del mapeig de les col·leccions web, per tal d'enriquir la col·lecció, i facilitar l'accés i l'intercanvi d'informació entre diferents sistemes i usuaris).

La tendència actual i objectiu general per a fer front a la interoperabilitat semàntica, rau en poder millorar els vocabularis controlats (conversió de vocabularis controlats a ontologies o llenguatges més comprensibles per màquina) i també, les eines relacionades amb l'alineació de vocabularis per tal d'aconseguir un alt grau d'interoperabilitat semàntica.

Finalment, es conclou que la interoperabilitat semàntica és un element clau per a la sostenibilitat del sistema d'informació, ja que permet la reutilització de la informació, intercanvi i enriquiment de les col·leccions amb documents o informació relacionada.

3.4. Quin és el model d'IQ (Informació de qualitat) pertinent per a un sistema d'informació de qualitat, distribuït i especialitzat?

Aquest estudi, mostra que un model o guia d'IQ d'un sistema d'informació ha d'incloure tant els aspectes teòrics com els pràctics, ha d'ésser sistemàtic i concís amb els tipus de criteris que posteriorment han d'avaluar la informació, però també s'ha de ser realista en la quantitat de criteris que es poden avaluar, calculant els costos que comporta l'aplicació d'un model d'IQ.

Primerament, s'ha de tenir en compte que són necessàries com a mínim dues fases en tot model d'IQ⁵⁸: 1) assenyalar quins atributs són importants segons la tipologia d'usuaris, i 2) determinar com aquests atributs afecten a les necessitats d'informació dels usuaris.

Posteriorment a l'aplicació del model, l'avaluació de la IQ ha de ser constant i ha de prevenir els canvis. Un model d'IQ d'un sistema d'informació distribuir s'emmarca dintre d'un entorn digital i canviant com és Internet, per tant, les necessitats d'informació i els atributs o dimensions d'IQ també seran variables. En el model s'han de contemplar o preveure alguns esquemes que analitzin i solucionin els problemes d'IQ que es detectin (per exemple, prevenir canvis que es produeixen en l'entorn).

L'anàlisi i comparativa dels models d'IQ conclou que les dimensions més representatives i necessàries per tot model d'IQ són:

- L'**accessibilitat**, si un usuari no té accés a la informació, la resta de criteris no seran avaluables perquè l'usuari haurà determinat que aquell document no té IQ. Altres factors dependents són l'actualització de la informació i la freqüència amb que aquestes s'actualitzin.
- La **representació de la informació** i el tipus de formats, versions i altres tipus de característiques tecnològiques que puguin afectar a l'accessibilitat de la informació.
- **Disponibilitat**, si hi ha costos associats a l'obtenció de la informació (sota subscripció, *Open Access*, etc), l'autoria o reputació de la font, etc.
- La **completesa de la col·lecció**, és on es té en compte l'anàlisi de les *metadades* (criteri i mesura d'avaluació d'IQ). Hi ha autors que uneixen la *completesa* a la dimensió de la *rellevància* (*Wang, Bovee i Miller*). La rellevància de fet, hauria de ser un factor clau en tot model d'IQ, perquè si la cobertura i rellevància de la col·lecció no és l'adequada per a l'usuari o per a l'audiència a la que va dirigida el sistema o entorn d'informació, el model d'IQ es trenca, ja que no es cobreixen les necessitats dels seus usuaris que són el principal motiu de la seva creació.

⁵⁸ Miller, H. 1996. The Multiple Dimensions of Information Quality. *Information Systems Management*, vol. 13, no. 2, pp.79-82.

D'altra banda, cal fer notar que tot i no afirmar-se s'afirma en cap dels estudis analitzats, si que en tot sistema d'informació de qualitat s'han d'incloure les metadades dintre el model d'IQ. Les metadades, aporten un gran valor afegit en criteris d'avaluació d'IQ, donen informació valuosa (autoria, matèria o data del document, etc.), serveixen per identificar i recuperar amb major precisió la informació, són fàcilment parametrizables (serveixen de mètrica d'IQ per a la dimensió de "Completesa"), normalment responen a estàndards internacionals (exemple, Dublin Core), i són una de les característiques principals de la interoperabilitat semàntica, per tant, faciliten l'intercanvi d'informació entre sistemes d'informació digitals.

4. Són els sistemes de mapeig de vocabularis controlats mètodes eficaços d'interoperabilitat semàntica en sistemes d'informació de qualitat en ciències aquàtiques?

El mapeig de vocabularis controlats funciona com un mecanisme de connexió entre diferents models conceptuals d'organització del coneixement. El context actual mostra que hi ha molts sistemes d'informació que apliquen sistemes de mapeig per a la alineació de vocabularis amb usos diversos com indexació, cerca i navegació creuades (sistemes distribuïts), o per a la fusió de vocabularis per a la integració de diferents KOS, que corroboren la hipòtesis que els sistemes de mapeig són eficaços per assegurar la interoperabilitat semàntica. No obstant això, existeixen diferents tipus de sistemes de mapeig, i la majoria estan en desenvolupament o en canvi constant, al igual que els sistemes semàntics que els suporten.

Els resultats d'aquest estudi confirmen que és necessari seguir amb el desenvolupament o millora dels sistemes de mapeig. Són necessaris sistemes de mapeig que diferenciïn diferents tipus de relacions semàntiques, així com també que tinguin en compte el context dels vocabularis o el propi usuari final (processos d'avaluacions o mapeig humà, guies d'avaluacions, etc.), i alhora incloguin mesures o mètodes d'alineaments addicionals.

Els estudis realitzats conclouen que encara calen d'altres mesures d'avaluació dels alineaments i dels processos de mapeig que hi ha al darrera (com la inclusió de tests addicionals, etc), perquè en els sistemes de mapeig encara hi ha errors. I en general, cal definir un consens entre els sistemes de mapeig existents, per tal de poder comparar esquemes de mapeig produïts i avaluar-los de forma adequada per millorar l'efectivitat dels processos automàtics de mapeig.

Per tal d'avaluar l'eficàcia dels mètodes de mapeig en l'àmbit de les ciències aquàtiques, es du a terme un experiment d'alineament de tres tesaurus de l'àmbit de les ciències aquàtiques i afins, i es conclou que:

Els resultats mostren que els sistemes de mapeig ajuden a la convergència entre diferents tesaurus (monolingües i multilingües), amb diferents estructures, lèxic, i contingut (cobertura) i semàntica.

Els mètodes d'alineament serveixen per identificar aquestes divergències: els resultats mostren diferències estructurals, culturals o de cobertura (ús de termes diferents, nivells d'especificitat i exhaustivitat diferents) i de forma (ús d'etiquetes diferents per a la definició dels termes).

Els resultats confirmen també que les tècniques d'alineació són processos d'alt valor afegit per a l'enriquiment de vocabularis ja existents, és a dir, poden ser útils per a l'actualització d'un vocabulari i/o per crear un 3r vocabulari, ja sigui perquè l'existent s'ha pogut quedar obsolet, o bé, perquè no existeixi cap tesaurus del domini específic que interessi.

Es demostra també que la intervenció humana és un gran avantatge per als sistemes de mapeig automàtics, ja que permeten la rectificació del propi sistema de mapeig perquè millori la qualitat dels alineaments.

Cal ressaltar també, com l'ús de vocabularis multilingües (i/o etiquetes ontològiques d'idiomes), milloren la qualitat dels alineaments, perquè ajuden a trobar un major nombre d'equivalències entre els termes emprats en diferents vocabularis.

Finalment es conclou que l'ús d'ontologies, independentment de la seva diversitat d'esquemes o KOS (diferències semàntiques, lèxiques i d'estructura), permeten integrar múltiples models de dades amb diferents punts de vista.

5. Recerca futura

Aquest estudi i anàlisi sobre la qualitat en la indexació i recuperació de la informació en ciències aquàtiques i marines, constitueix la base per a la creació d'un sistema d'informació distribuït tipus passarel·la temàtica. Les dades aportades en aquest treball demostren la necessitat del desenvolupament de sistemes d'informació de tipus col·laboratiu i cooperatiu que aportin informació d'alt valor afegit i facin front a l'heterogeneïtat i multidisciplinarietat del propi camp temàtic. Les biblioteques i centres de documentació especialitzats en aquest àmbit, tant a nivell europeu com internacional, han posat de manifest l'interès i els esforços en temes de treball col·laboratiu sobre gestió de la informació científica i intercanvi de coneixement, però encara són escassos els

desenvolupaments en l'àmbit de les ciències aquàtiques i marines, i molts més escassos els treballs relacionats amb el *Linked Data* i l'*Open Linked Data*. Cal remarcar que la majoria d'aplicacions relacionades amb ciències afins (agricultura, etc.) són projectes liderats de forma individual per organitzacions governamentals. Per tant, s'espera que aquest treball juntament amb la voluntat de les xarxes existents (EURASLIC i IAMSLIC) pugui aportar el coneixement per a noves investigacions així com el desenvolupament d'un sistema d'informació distribuït en ciències aquàtiques i marines que afronti els nous reptes del web semàntic.

10. BIBLIOGRAFIA

Nota: El darrer accés a bibliografia online ha estat entre desembre 2013 i gener 2014. En els casos que el recurs ja no està accessible, es posa la data de darrera accés en el propi recurs.

3.6. Interoperability. In: Desire Handbook. 3.6 Interoperability. Url: <http://www.desire.org/handbook/3-6.html>

Agenjo-Bullón, X.; Hernández-Carrascal, F. (2010). La construcción de esquemas semánticos para bibliotecas virtuales". *I Jornada profesional 'Lenguajes y gestión de información'*, organizada por Sedic y el Instituto Cervantes, 17 junio 2010. Url: <http://goo.gl/3gMo9>

AGRIS Application Profile for the International Information System on Agricultural Sciences and Technology Guidelines on Best Practices for Information Object Description. FAO, 2005. Url: <http://www.fao.org/docrep/008/ae909e/ae909e00.htm>

AGROVOC Tesauro: http://www.fao.org/aims/ag_intro.htm

AGROVOC architecture: <http://www.fao.org/aims/agrovoccs.jsp>

Amalgame. W3C, 23 January 2013. Url: <http://www.w3.org/2001/sw/wiki/Amalgame>

Aitchison, J.A.; Gilchrist, A.; Bawden, D. (2000). Thesaurus construction and use: A practical manual (4th ed.). Chicago, IL., Fitzroy Dearborn.

Aitchinson, J.; Gilchrist, A.; Bawden, D. (1997). Thesaurus construction and use: a practical manual. 3th ed. London : *Aslib*, 1997. 212 p.; 30 cm. ISBN: 0-85142-390-6

Shiri, A.A.; Revie, C. (2000). Thesauri on the web: current developments and trends. *Online Information Review*, 24(4), p.273-280

Alemu, G., Stevens, B.; Ross, P. (2012). Towards a conceptual framework for user-driven semantic metadata interoperability in digital libraries: A Social Constructivist Approach. *New Library World*, 2012. 113(1-2): p.38-54.

Alexander, J.E., Tate, M.A. (1999). web wisdom : how to evaluate and create information quality on the web. Mahwah (N.J.) : Lawrence Erlbaum Associates, 1999. XV, 156 p.: il. ISBN: 0805831231

Algergawy, A.; Massmann, S.; Rahm, E. (2011). A clustering-based approach for large-scale ontology matching. *Advances in Databases and Information Systems*, Springer, 2011.

Alonso-Arevalo, J., et al. (2010). Medical Information Library & Knowledge: MILK o el descubrimiento de la información científica en Ciencias de la Salud por medio del etiquetado social: Connotea and CiteUlike. *EAHIL (Lisboa, Portugal, 14th-18th June, 2010)*. Url: <http://repositorio.ispa.pt/bitstream/10400.12/232/1/EAHIL%20-%20MILK42spa.pdf> .

Allen, R. (2000). With a *wysh* and a prayer: An experiment in cooperative development of legal knowledge bases. *The Journal of Information, Law and Technology (JILT)*, 2000, (2). Url: <http://elj.warwick.ac.uk/jilt/00-2/allen.html>

Amalgame. W3C, 23 January 2013. Url: <http://www.w3.org/2001/sw/wiki/Amalgame>

AMALGAMAE ClioPatria toolkit. Video-Demo: <http://vimeo.com/23420503> ;
Autors: vrije Universiteit Amsterdam (Michiel Hildebrand, et al.); Col-laboradors: Europeana Connect (Antoine Isaac). Pàgina web d'Amalgame. Url: <http://semanticweb.cs.vu.nl/amalgame/>

Amann, B.; Fundulaki, I.; and Scholl, M. (2000). Integrating ontologies and thesauri for RDF schema creation and metadata querying. *International Journal on Digital Libraries*, 2000, 3(3), p. 221-236.

ANSI/NISO Z39.19-2005. Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. Bethesda, Maryland: NISO, 2005, 188 p. ISBN 1-880124-65-3. NISO. Url: <http://www.niso.org/standards/index.html>

Antoniou, G.; Van Harmelen, F. (2010). Manual de web semàntica. Núria Casellas; Manual Atencia (traducció). Granada: Comares, S.L., 2010. XXVI, 301 p. ISBN: 978-84-9836-780-5.

Angjeli, A., et al., (2009). Semantic web and Vocabulary Interoperability: An Experiment with Illumination Collections. *International Cataloguing and Bibliographic Control*, 2009, 38, p. 25-29.

Araújo, R.; Pinto, H.S. (2007). Towards semantics-based ontology similarity. *Ontology Matching*, 2007, 304, p. 37.

ARC (Australian Research Council). Url: <http://www.arc.gov.au/>

Ardö, A., et al. (2000). Nordic Interconnected Subject-based Information Gateways (NISBIG), 2000. URL: <http://www.lub.lu.se/nisbig/slutrapport.html>

Aristar, A., et al. (2009). Language engineering for the semantic web: a digital library for endangered languages. *Information Research*, 2004, 9(3).

ASFA Tesauros: <http://www4.fao.org/asfa/asfa.htm>

ATUA: <http://www.atua.org.au/atua.htm>

AustLII. (1997a). Greenleaf, G.; Mowbray, A.; King, G.; Cant, S; Chung, P. 'New Legal Services via the web - AustLII's Research on Legal Inferencing', Chapter 5 of Greenleaf, et al. *New Directions in Law via the Internet: The AustLII Papers*" [1997] Col. 2. Url: <http://www.AustLII.edu.au/au/other/col/1997/2>

AustLII. (1997b). Mowbray A.; Greenleaf G.; King, G.; Cant, S. Wysh Developers Manual, 1997, Col. 1. Url: <http://www.AustLII.edu.au/au/other/col/1997/1>

Austlit. Url: <http://austlit.edu.au/>

AustLII Guide to Legal Research on the web. Url <http://www.AustLII.edu.au/AustLII/guide/current/20030315-Contents.html>

AustLII Twitter. Url: <https://twitter.com/AustLII>

Australian Libraries Gateway. Url: <http://www.nla.gov.au/libraries/about.html>

Australian Subject Gateway Forum. Definitions for web-Based Services. Url: <http://www.nla.gov.au/initiatives/sq/servicetypes.html>

Avel: <http://avel.library.uq.edu.au/about.html>

Babu, P.B., et al. (2012). Rationale of institutional repository categories and IR development challenges in India. *Library Review*, 2012, 61(6), p. 394-417.

Baker, T. (2012). Libraries, languages of description, and *Linked Data*: a Dublin Core perspective. *Library Hi Tech*, 2012. 30(1). p. 116-133.

Barite, M.G. (2000) The notion of 'category' : its implications in subject analysis and in the construction and evaluation of indexing languages. *Knowledge Organization*, 2000, 27(1/2), p. 4-10.

Barckow, A. (2009). Vascoda – Co-Ordination of International Activities of German Subject Portals. Proceedings of the Sino-German. *Symposium on Development of Library and Information Services 2009 (Hunming, China, 2009)*

Baruzzo, A., et al. (2009). Toward Semantic Digital Libraries: Exploiting web 2.0 and Semantic Services in Cultural Heritage. *Journal of Digital Information*, 2009. 10(6), p. NP

Bauer, F., Kaltenböck, M. (2011). Linked Open Data: The Essentials - A quick start guide for decision makers REEEP Viena, Austria. Url: <http://www.semantic-web.at/LOD-TheEssentials.pdf>

Bawden, D.; Robinson, L. (2002). Internet subject gateways revisited. *International Journal of Information Management*, 2002, 22, p.157-162

Becker, H.J., Neuroth, H. (2002). Crosssearchen und crossbrowsen von 'Quality-controlled *Subject Gateways*' im EU-Projekt Renardus. Cross-searching and cross-browsing with 'quality-controlled subject gateways' in the EU project Renardus. *Zeitschrift für Bibliothekswesen und Bibliographie*, 2002, 49, p.133-146.

Berners-Lee, T. (2009). LOD = Linking Open Data; Berners-Lee. *TED 2009 conference, "The Great Unveiling"* (Long Beach, CA., USA, 4 February, 2009).
Url: <http://www.w3.org/2009/Talks/0204-ted-tbl/#%281%29>

Berners-Lee, T. (1998). Semantic web Road map. Url:
<http://www.w3.org/DesignIssues/Semantic.html>

Berners-Lee, T., et al. (2005). Uniform Resource Identifier (URI): Generic Syntax. *IETF (January, 2005)*. Url: <http://tools.ietf.org/html/rfc3986>

Berners-Lee, T. Hendler, J., Lassila, O. (2001). The Semantic web. *Scientific American*, 2001, 284(5), p.34-43.

Beyond book indexing: How to Get Started in web Indexing. Information Today Phoenix, ed. D. Brenner and M. Rowland. American Society of Indexers; Medford, N.J.; 2000, 150 p. ISBN: 1-57387-081-1

Biagetti, M.T., (2009). Digital Libraries and the Semantic web. *Bollettino AIB*, 2009, 49(4), p. 577-582

Bikakis N., et al. (2012). The XML and Semantic web Worlds: Technologies, Interoperability and Integration. A survey of the State of the Art. *Semantic Hyper/Multi-media Adaptation: Schemes and Applications*. Springer, 2012. Url:
<http://www.dblab.ntua.gr/~bikakis/XMLSemanticWebW3CTimeline.svg>

Blair, D.C. (1990). Language and Representation in Information Retrieval. Amsterdam: Elsevier Science, Ltd., 1990, 350 p. ISBN: 9780444884374

Boast, R.; Bravo, M.; Srinivasan, R. (2007). Return to babel: Emergent diversity, digital resources, and local knowledge. *Information Society*, 2007. 23(5), p. 395-403.

Bonura, Larry S. (1994). The Art of indexing. New York, Wiley, 1994, 233 p.

Borst, T., et al., (2010a). How Do Libraries Find Their Way in to the Semantic web?. *A.B.I. Technik*, 2010, 30, p. 31-34.

Borst, T., et al., (2010b). On the Path to the Semantic web: Application Examples and Solution Scenarios in Libraries. *A Presentation by hbz and ZBW. BuB Forum Bibliothek und Information*.

Bosch, M. (2006). Ontologies, different reasoning strategies, different logics, different kinds of knowledge representation: Working together. *Knowledge Organization*, 2006, 33(3), p. 153-159.

Both, P. F. (2001). *Indexing: The Manual of Good Practice*. Munchen, Saur, 2001.

Bovee, M.; Srivastava, R.P.; Mak, B. (2003). A conceptual framework and belief-function approach to assessing overall information quality. *International Journal of Intelligent Systems*, 2003, 18(1), p. 51-74.

Brooks, T.A. (2002). The Semantic web, universalist ambition and some lessons from librarianship. *Information Research*, 2002, 7(4).

Brown, T.C. and Daniel, T.C. (1990). Scaling of ratings: concepts and methods. *Res. Pap. RM-293*. Fort Collins, CO, U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, 1990, 24 p. Url: http://www.fs.fed.us/rm/pubs_rm/rm_rp293.pdf

BS 8723-2:2005: Structured vocabularies for information retrieval. Guide. Thesauri. Published : November 2005 [Replaced By : [BS ISO 25964-1:2011](#)]

Buckingham Shum, S.; Motta, E.; Domingue, J. (2000). ScholOnto: an ontology-based digital library server for research documents and discourse. *International Journal on Digital Libraries*, 2000. 3(3), p. 237-248.

Building a Movement: Europeana Annual Report and Accounts 2012. Europeana Foundation. Url: <http://pro.europeana.eu:9580/documents/858566/99cea5ba-46ce-4037-a913-a9c81080079b> [Darrer accès: 01/12/2013]

Business Information Systems Workshops. Springer Berlin Heidelberg, pp. 244-255. *BIS 2012 International Workshops and Future Internet Symposium (Vilnius, Lithuania, May 21-23, 2012)*

Burke, M., (2009). The Semantic web and the Digital Library. *Aslib Proceedings: New Information Perspectives*, 61, p. 316-322.

Bygstad, B., et al., (2009). Organisational Challenges of the Semantic web in Digital Libraries: A Norwegian Case Study. *Online Information Review*, 33, p. 973-985.

Campbell, D. (2000). Australian subject gateways: political and strategic issues. *Online Information Review*, V.24(1), Pp.73-74.

Campbell, D. (1999). An overview of subject gateway activities in Australia. *Ariadne*, (21). Url: <http://www.ariadne.ac.uk/issue21/subject-gateways/>

Campbell, D.G., Fast, K.V. (2004). Academic libraries and the Semantic web: what the future may hold for research-supporting library catalogues. *Journal of Academic Librarianship*, 2004, 30, p. 382-390.

Candan, K.S., Liu, H., Suvarna, R., (2001). Resource description framework: metadata and its applications. *SIGKDD Explorations. Newsletter*, 2001, 3, p. 6-19.

Capurro, R. (1990). Towards an Information Ecology. *NORDINFO International seminar "Information and Quality"*, Royal School of Librarianship, Copenhagen, 23-25 August 1989. Proceedings: I. Wormell ed.: Information Quality. Definitions and Dimensions. London, Taylor Graham 1990, p. 122-139. Url: <http://www.capurro.de/nordinf.htm>

Caracciolo, C., et al., (2012). Thesaurus maintenance, alignment and publication as *Linked Data: the AGROVOC use case*. *International Journal of Metadata, Semantics and Ontologies*, 2012, 7, p. 65-75.

Caracciolo, C., et al., (2008). Results of the ontology alignment evaluation initiative 2008. *Proc. 3rd ISWC workshop on ontology matching (OM)*, p. 73-119.

CARMEN (Content Analysis, Retrieval and MetaData: Effective Networking). Url: <http://www.bibliothek.uni-regensburg.de/projects/carmen12/index.html.en>; Accesible també a: <http://www.mathematik.uni-osnabrueck.de/projects/carmen/>

Catàleg Col·lectiu de biblioteques de Museus Municipals de Barcelona. Url: <http://catalegbibliotequesicub.bcn.cat/CercaSimple.aspx?idioma=es>

CC:DA (ALCTS/CCS/Committee on Cataloging: Description and Access). (2000). *Task Force on Metadata: Final report*, June 16, 2000. Url: <http://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html>

CIP-ICT Policy Support Programme 2009. EC. Information Society and Media, 2009. Url: http://ec.europa.eu/information_society/activities/ict_psp/documents/leaflet_call_3_cip_ict_psp_v.0309.pdf

CIP-ICT PSP work programme 2012. *CIP-ICT Policy Support Programme*. EC. Information Society and Media, 2012. Url: http://ec.europa.eu/information_society/activities/ict_psp/documents/cip_ict_psp_wp2012_adopted_01022012.pdf

Cleveland, D. B. (2001). Introduction to indexing and abstracting. 3rd edition. Englewood, Libraries Unlimited, 2001.

Cliff, P. (2001). Building Resource Finder. *Ariadne*, 2001, (30). Url: <http://www.ariadne.ac.uk/issue30/rdn-oai/>

Codina, LI. (2006). Metodología de análisis y evaluación de recursos digitales en línea (v. 2006). Barcelona: UPF. Área de Biblioteconomía y Documentación. Dep. De Periodismo y de Comunicación Audiovisual, 2006, 56 p. (documento reprografiado). URL: <http://www.lluiscodina.com/metodos.htm>

Concordia, C.; Gradmann, S.; Siebinga, S. (2010). Not Just Another Portal, Not Just Another Digital Library: A Portrait of Europeana as an Application Program Interface. *IFLA Journal*, 2010. 36(1), p. 61-69.

Guia per la integració de registres RDA en el CCUC. *Concrecions a les RDA*. CCUC, abril 2013. Url: http://pautes.cbuc.cat/doku.php?id=guia_per_la_integracio_de_registres_rda_en_el_ccuc#guia_per_la_integracio_de_registres_rda_en_el_ccuc

Cornella, A.1996. "Como darse de baja y evitar la infoxicación en Internet". *Extra!-Net*, Mensaje 187, 16/12/1996. Url: http://www.infonomia.com/pdf/1996_12_16_extranet.187.infoxicacion.pdf

Correspondence between ISO25964 and SKOS/SKOS-XL Models. Niso, 23/06/2012. Aaccessible a: http://www.niso.org/apps/group_public/download.php/9627/Correspondence%20ISO25964-SKOSXL-MADS-2012-10-21.pdf

Coyle, K. (2010). Cataloging Horizons: Navigating the Bibliographic Space with Linked Data. *American Libraries*, 2010, 41(3), p. 26-26.

CrossROADS, <http://www.ukoln.ac.uk/metadata/roads/crossroads/>
Crosswalks tools in Marine Metadata Interoperability. MMI, 2007?. Url: <http://marinemetadata.org/tools>

Chen, Z., et al. (2005). Knowledge Service and Digital Library: A Roadmap for the Future. *Digital Libraries: International Collaboration and Cross-Fertilization*. 2005, Springer Berlin Heidelberg. p. 104-114

Choi, Y. (2010). Traditional versus emerging knowledge organization systems: Consistency of subject indexing of the web by indexers and taggers. *Proceedings of the American Society for Information Science and Technology*, 2010, 47, p. 1-2.

Chowdhury, G.G. Introduction to Modern Information Retrieval. 2 ed. London : Facet, 2004. 474 p. ISBN: 1-85604-480-7.

D5.3.1: Semantic Interoperability in Digital Library Systems. DELOS Project no.507618. A Network of Excellence on Digital Libraries. 2004.

Danskin, A. (2013a). Deployment of RDA (Resource Description and Access) Cataloging and its Expression as Linked Data. *NISO/DCMI Joint Webinar Series*, 2013. Url: <http://www.niso.org/news/events/2013/dcmi/rda/>

Danskin, A., (2013b). Linked and open data: RDA and bibliographic control. *JLIS.it*, [S.I.], jan. 2013, 4(1), p. 147. ISSN: 2038-1026. Url: <http://leo.cilea.it/index.php/jlis/article/view/5463>. Doi:10.4403/jlis.it-5463.

Dappert, A.; Enders, M. (2008). Using METS, PREMIS and MODS for Archiving eJournals. *D-Lib Magazine*, 2008, 14(9/10). Url: <http://www.dlib.org/dlib/september08/dappert/09dappert.html>

Darek M.H.; Kajtazi, M.; Mirijamdotter, A. (2011). A Review of Information Logistics Research. Abramowicz, W. (ed.). *Lecture Notes in Business Information Processing*, 2011, 97, p. 244-255.

Davenport, T.; Prusak, L. (1997). Information ecology: mastering the information and knowledge environment. New York : Oxford University Press, 1997. ISBN: 9789706134486

Day, M. (2004). Cross-browsing subject gateways with the Decimal Classification in the Renardus Service. Bath: University; UKOLN, 2004. Url: <http://www.ukoln.ac.uk/metadata/presentations/jiscterm-2004/demo.html>

Day, M., et al., (2005). Searching and browsing multiple subject gateways in the Renardus service, Recent developments and applications in social research methodology. *Proceedings of the RC33 Sixth International Conference on Social Science Methodology*. Amsterdam: Budrich Verlag,

DCMI Education Community. Url: <http://dublincore.org/groups/education/>

DCMI Libraries Community. Url: <http://dublincore.org/groups/libraries/>
DCMI Metadata Terms, 14 January 2008. Url: <http://dublincore.org/documents/dcmi-terms/>

DCMI Science And Metadata Community. Url: http://wiki.dublincore.org/index.php/Main_Page

Definitions for web-Based Services. Australian Subject Gateway Forum, 4 July 2005. Url: <http://www.nla.gov.au/initiatives/sg/servicetypes.html> [Darrer accés:02/07/2008; pàgina web actualment no disponible, arxivada a: <http://pandora.nla.gov.au/pan/60664/20060915-0000/www.nla.gov.au/initiatives/sg/index.html>]

DBPedia. Url: <http://wiki.dbpedia.org/Downloads37>

Debra, H. (2006). Retrospective on the RDN. [en linea]. *Ariadne*, 2006, (47). Url: <http://www.ariadne.ac.uk/issue47/hiom>

Dempsey, L. (2000). The subject gateway: experiences and issues based on the emergence of the Resource Discovery Network. *Online Information Review*, 24(1), p. 8-23.

Dempsey, L. (1996). ROADS to Desire : Some UK and Other European Metadata and Resource Discovery Projects. *D-Lib Magazine*, July/August 1996.

Dempsey, L. et al. (c2004-05). Metadata switch: thinking about some metadata management and knowledge organization issues in the changing research and learning landscape. *Forthcoming in LITA guide to e-scholarship* (working title), ed. Debra Shapiro, 25p. Url: <http://www.oclc.org/research/publications/archive/2004/dempsey-mslitaguide.pdf>

DESIRE Information Gateways Handbook.
URL:<http://www.desire.org/handbook/welcome.html>

Dietrich, D. (2010). Metadata Management in a Data Staging Repository. *Journal of Library Metadata*, 2010, 10(2-3), p. 79-98.

Dillon, D. (2007). Google, libraries, and knowledge management: from the Navajo to the National Security Agency. *Journal of Library Administration*, 2007, 46(1), p. 27-40.

Di, H. (2003). A Review of Cross Searching Technique for Heterogeneous Databases. *Library and Information Service*, 2003, (6), p. 94-97.

Digital Preservation Metadata Standards. *ISQ: Information Standards Quarterly*. Special Issue: Digital Preservation, 2010, 22(2), 13 p. Url: http://www.loc.gov/standards/premis/FE_Dappert_Enders_MetadataStds_isqv2_2no2.pdf

Do, H.-H., E. Rahm, et al. (2002). Chapter 53 - COMA - A system for flexible combination of schema matching approaches. *VLDB '02: Proceedings of the 28th International Conference on Very Large Databases (Hong Kong, China, August 20-23, 2002)*. San Francisco, Morgan Kaufmann, 2002, p. 610-621.

Doerr, M., et al., (2010). The europeana data model (edm). *World Library and Information Congress: 76th IFLA general conference and assembly*, p. 10-15.
Doerr, M. (2004). Semantic interoperability: Theoretical Considerations. *Technical Report 345, ICS-FORTH*. October 2004. Url: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.147.4269&rep=rep1&type=pdf>

Doerr, M. (2001). Semantic Problems of Thesaurus Mapping. *Journal of Digital Information*, 2001 1(8). URL:<http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/>

Dublin Core Basic Elements. DC Metadata Element Set. Versió 1.1., 2012. Url: <http://dublincore.org/documents/dces/> [Darrer accés: 10/01/2014].

Dublin Core Qualifiers. URL: <http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>

Duke, M. (2001). Delivering OAI Records As RSS: An IMesh Toolkit Module for Facilitating Resource Sharing. *Ariadne*, October 2003, (37). Url: <http://ariadne.ac.uk/issue37/duke>

Dunsire, G. (2013a). An introduction to open Linked Data for librarians. *Presented at the National Library of Finland (Helsinki, Finland, 11 Dec 2012)*.
Url: <http://www.gordondunsire.com/pubs/pres/OpenLinkedDataLib.pptx>

Dunsire, G. (2013b). Mapping FRBR, ISBD, RDA, and other namespaces to DC for interoperability. *Presented at Kunnskapsorganisasjonsdagene (Oslo, Norway, 7-8 Feb 2013)*.
Url: www.gordondunsire.com/pubs/pres/MappingToDC.pptx

Dunsire, G. (2013c). Multilingual issues in the representation of international bibliographic standards for the Semantic web. *The Sixth Multilingual web Workshop: Making the multilingual web work (Rome, Italy, 12-13 Mar 2013)*.
Url: <http://www.w3.org/International/multilingualweb/rome/slides/19-dunsire.pdf>

Dunsire, G. (2013d). Role of national bibliographic agencies in Linked Data environment. Presented to staff of the Bibliothèque nationale de France, Paris, 25 Apr 2013.
Url: <http://www.gordondunsire.com/pubs/pres/RoleNatBibAgencies.pptx>

Dunsire, G. (2012). Linked Data, libraries and the Semantic web. *Library science talk*, 2012, p. 12-13.

Dunsire, G., et al. (2012). Linked Data Vocabulary Management: Infrastructure Support, Data Integration, and Interoperability. *Information Standards Quarterly*, Spring/Summer 2012, 24 (2-3), p.10.

Dunsire, G.; M. Willer. (2010). Initiatives to make standard library metadata models and structures available to the Semantic web. *World Library and Information Congress 76th IFLA General Conference and Assembly*, (Gothenburg, Sweden, 10-15 August 2010).

Dunsire, G.; M. Willer. (2010). UNIMARC and *Linked Data*. *IFLA Journal*, 2010, 37(4), p. 314-326.

Dutta, B. (2003). Cataloguing web Documents using Dublin Core, MARC21. *Workshop on Digital Libraries: Theory and Practice (BRTC), (Bangalore : March, 2003)*.
Url: https://drtc.isibang.ac.in/bitstream/handle/1849/24/C_Webcat_bisu.pdf?sequence=2

EndNa. Resources and networks for teaching and learning. Subject Gateway.
Url: http://apps-new.edna.edu.au/edna_retired/edna/go.html

Ellis, D. (1996). Progress and problems in information retrieval. London: Library Association, 1996.

Eppler, M. (2003). Managing information quality: Increasing the value of information in knowledge-intensive products and processes. Berlin, Germany: Springer-Verlag, 2003.

Eppler, M.J, et al. (2004). Information Quality : Organizational, Technological, and Legal perspectives. *Studies in Communication Sciences*,2004, 4/2, p. 1-16.

Eppler, M. J., Wittig, D. (2000). Conceptualizing Information Quality: A Review of Information Quality Frameworks from the Last Ten Years. *Proceedings of the 2000 Conference on Information Quality*. Cambridge, MA: Massachusetts Institute of Technology, 2000, p. 83-96.

EURASLIC 15th. World Ocean of Data & Information: Creating Value by their Organization and Management. 15th Biennial Conference of the European Association of Aquatic Sciences Libraries and Information Centres (15th EURASLIC Conference, Varna, Bulgaria, 13-15 May, 2013). Url: <http://www.euraslic.org/attachments/article/59/EURASLIC%2015.pdf>

Europeana. (2012). Annual Report and Accounts 2012. Europeana, 2012. Url: <http://pro.europeana.eu/documents/858566/99cea5ba-46ce-4037-a913-a9c81080079b>

Europeana. About us. About us. Europeana. URL: <http://www.europeana.eu/portal/aboutus.html> [Darrera consulta 01/09/2010]

Europeana Data Model (EDM) Documentation. Url: <http://pro.europeana.eu/web/guest/edm-documentation>; <http://pro.europeana.eu/edm-documentation>

Europeana Facts & Figures, April 2012. Url: <http://pro.europeana.eu/web/guest/about/facts-figures>

Europeana Thought Lab. Url: <http://pro.europeana.eu/web/guest/thoughtlab>

Europeana. Results and Resources. EuropeanaConnect. Url: <http://www.europeanaconnect.eu/results-and-resources.php>

Europeana Semantic Search (Prototype). Url: <http://eculture.cs.vu.nl/europeana/session/search>

Europeana v1.0. January 2009 - September 2011. Url: <http://pro.europeana.eu/web/europeana-v1.0>

Europeana v2.0. Url: <http://pro.europeana.eu/web/europeana-v2.0/home>

Euzenat, J., et al. (2011a). Ontology alignment evaluation initiative: six years of experience. *Journal on data semantics XV*, Springer, p. 158-192.

Euzenat, J., et al. (2011b). Final results of the ontology alignment evaluation initiative 2011. *Proc. 6th ISWC workshop on ontology matching (OM)*, 2011, p. 85-110.

Euzenat, J., et al. (2007). Results of the Ontology Alignment Evaluation Initiative 2007. Ontology Alignment Evaluation Initiative (OAEI).

Euzenat, J., et al. (2006). Results of the Ontology Alignment Evaluation Initiative 2006. Ontology Alignment Evaluation Initiative (OAEI).

Farrow, John F. (1991). A Cognitive Process Model of Document Indexing. *Journal of Documentation*, June 1991, 47, p. 149-66.

Fast, K.V., Campbell, D.G. (2001). The ontological perspectives of the Semantic web and the metadata harvesting protocol: Applications of metadata for improving web search. *Can. J. Inf. Libr. Sci.-Rev. Can. Sci. Inf. Bibl*, 2001, 26, p. 5-19.

Ferguson, J.C. (2007). Semantic web technologies: opportunity for domain targeted libraries?. *Journal of Electronic Resources in Medical Libraries*, 2007, 4, p.113-125.

Ferreira de Castro, F.; da Costa Santos, P.L.V.A. (2009). Use of Technology in Descriptive Representation: The MarcOnt Initiative Pattern of Semantic-Bibliographic Description and Its Impact. *Digital Information Environments. Ciencia da Informacao*, 2009, 38(1), p. 74-85.

Ferreira de Castro, F.; da Costa Santos, P.L.V.A. (2007). Os metadados como instrumentos tecnologicos na padronizacao e potencializacao dos recursos informacionais no ambito das bibliotecas digitais na era da web Semantica. / The metadata as technological instruments in standardization and potentialization of informational resources in the scope of the digital libraries in the Semantic web era. *Informacao & Sociedade: Estudos*, 17(2), np.

Fischer, T.; Neuroth, H. (2000). SSG-FI - special subject gateways to high quality Internet resources for scientific users. *Online Information Review*, 2000, 24(1), p. 64-68.

Flanders, J; Muñoz, T. [2012?]. An introduction to Humanities Data Curation. DH Curation Guide. Url: <http://guide.dhcuration.org/intro/> [Page update 2012-03-15];

Fjord, E. (2010). With Leaps and Bounds Toward a New Webparadigm. *REVY*, 2010, 33(2), p. 20-21.

Ford, K. (2010a). Linked Data at the Library of Congress. *SACO-At-Large meeting. American Library Association Annual Conference (Washington, DC, 2010)*

Ford, K. (2010b). Sharing Authority and Vocabulary Data at the Library of Congress. *August Federal Data Archicture Subcommittee Meeting (Arlington, VA, 2010)*.

Ford, K. (2003). Review, Changes, Future Plans, MADS/RDF. *Digital Library Federation Fall Forum (Palo Alto, CA., 2010)*.

Foskett, A.C. (1996). *The Subject Approach to Information*. 5th edition .London: Library Association Publishing, 1996.

Franco, A. (2003). Gateways to the Internet: Finding quality information on the Internet. *Library trends*, 2003, 52(2), p. 228-246

Franklin, R.A. (2003). Re-inventing subject access for the semantic web. *Online Information Review* 27(2): 94-101.

Fraser, M. (2003). Portal-to-portal : joining up content to decrease the time spent clicking as distinguished from the time spent working. Url:http://portal.humbul.ac.uk/presentations/SPPPortalTalks/spp_einiras2003.ppt

FRBR, RDA, and MARC. (2012). Library of Congress. Library of Congress (LC) RDA Training Materials, 2012.Url: <http://www.loc.gov/catworkshop/RDA%20training%20materials/LC%20RDA%20Training/LC%20RDA%20course%20table.html>

Friend of a Friend (FOAF), 2000. Url: <http://www.foaf-project.org/>

Frohmann, B. (1990). Rules of Indexing: a critique of mentalism in information retrieval theory. *Journal of Documentation*, 1990, 46 (2), p. 81-101.

Fugmann, R. (1993). *Subject Analysis and Indexing: Theoretical Foundation and Practical Advice*. Frankfurt: Indeks Verlag, 1993, 256 p. ISBN: 3-88672-500-6.

Gangemi, A., et al., (2004). A core ontology of fishery and its use in the fishery ontology service project. In EKAW04 Workshop on Core Ontologies in Ontology Engineering, Northamptonshire (UK), 8-5 October 2004. Url: <http://eprints.rclis.org/15687/>

García Delgado, P. (2001). *ROADS*. Software y conjunto de normas para la gestión de recursos electrónicos en Internet...: Dins: JBDI '2001. Segundas Jornadas de Bibliotecas Digitales (Almagro, Ciudad Real, 19 y 20 Noviembre, 2001). Url: http://infonautica.net/docs/jbidi/jbidi2001/19_2001.pdf

García Gutiérrez, A.L. (1990). *Estructura lingüística de la documentación: Teoría y Método*. Murcia: Universidad de Murcia, 1990.

García-Marco, F-J. (2011). Libraries in the digital ecology: reflections and trends. *The Electronic Library*, 2011, 29(1), p.105-120. DOI: [10.1108/02640471111111460](https://doi.org/10.1108/02640471111111460)

Gasevic, D. (2005). Searching web Resources Using Ontology Mappings. *Integrating Ontologies Workshop Proceedings* (Banff, Alberta, Canada, 2005).

GEMET Tesauro: <http://www.eionet.europa.eu/gemet>

Ghiglione, R.; Beauvois, J.L. (1980). Manuel d'analyse de contenu. Paris: Armand Colin, 1980.

Giunchiglia, F., et al. (2012). S-Match: An open source framework for matching lightweight ontologies. *Semantic web*, 2012, 3(3), p.307-317.

Glossary of terms relating to thesauri and other forms of structured vocabulary for information retrieval, Stella Dextre Clarke, Alan Gilchrist, Ron Davies and Leonard Will, Willpower Information. Url: <http://www.willpowerinfo.co.uk/glossary.htm>

Goossens, P. (2003). ELAG 2002: report of a library systems seminar on the semantic web. *Program* 37(4), p. 251-253.

Golub, K. (2006). Using Controlled Vocabularies in Automated Subject Classification of Textual web Pages, in the Context of Browsing. *TCDL Bulletin*, V2(2), 2006. URL: <http://www.ieee-tcdl.org/Bulletin/v2n2/golub/golub.html>

Golub, K., Lykke, M., (2009). Automated classification of web pages in hierarchical browsing. *Journal of documentation*, 2009, 65, p. 901-925.

Gómez Díaz, R. (2003). La evaluación en recuperación de la información. *Hipertext.net*, (1), 2003. Url: http://www.upf.edu/hipertextnet/numero-1/evaluacion_ri.html

Gonçalves, M.A. (2004). Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries. (2004). Ed. Edward A. Fox, Layne T. Watson. ACM Trans. Inf. Syst. Virginia, November 29, 2004.

Gonçalves, M.A., et al. (2007). "What is a good digital library?" : A quality model for digital libraries. *Information Processing and Management*, 2007, 43(5), p. 1416-1437.

Gradmann, S. (2010). Knowledge= Information in Context: on the Importance of Semantic Contextualisation in Europeana. Europeana White Paper 1.

Gradmann, S., et al. (2011). Design, Implementation and Evaluation of a User Generated Content Service for Europeana, Research and Advanced Technology for Digital Libraries. Springer Berlin Heidelberg, p. 477-482.

Gradmann, S. (2008). INTEROPERABILIDAD. Un concepto clave a bibliotecas digitales a gran escala persistentes. Traductor: Juan-José Boté. *Briefing Papers. Digital Preservation Europe*, 2008. Pàgina 2, secció 6. Url:

http://www.digitalpreservationeurope.eu/publications/briefs/es_interoperabilidad.pdf

Graham, S.R. (2011). Gateways to Legal Information: Best-Bet Internet Portals.

Greenleaf, G. (2009). AustLII's Business Models: Constraints and opportunities in funding free access to law, Free Access, Quality of Information, Effectiveness of Rights. *Proc. IX International Conference 'Law via the Internet'*. European Press Academic Publishing, Florence, Italy, p. 423-436.

Greenleaf, G., et al. (2010). AustLII: Thinking locally, acting globally. ALLA & NZLAA Cross Currents Joint Conference (Melbourne, 2010). Published in *Australian Law Librarian*, 2011, p.101-116, referenced as [2011] UNSWLRS 41.

Greenleaf, G., et al. (1999). With a *wysh* and a prayer: an experiment in cooperative development of legal knowledgebases. *Proceedings of the 7th international conference on Artificial intelligence and law*. Oslo, Norway: ACM, 1999.

Greenberg, J. (2003). Metadata and the World Wide Web. Dins: BATES, M. J.; MAACK, M.N. y DRAKE, M. (eds.). *Encyclopedia of Library and Information Science*. New York: Dekker, 2003, p. 1876-1888.

Greenberg, et al. (2011). HIVE: Helping Interdisciplinary Vocabulary Engineering. *Bulletin of the American Society for Information Science and Technology (ASIST)*, April/May 2011, 37 (4), p. 23-26.

Grolik, S. (2007). Information Logisitcs : Decentralization Approaches of Information Allocation in Information Exchange Networks. *Ibidem-Verlag*, Hannover.

Gruber, T. R. (1993). [A Translation Approach to Portable Ontology Specifications](#). *Knowledge Acquisition*, 1993, 5(2), p. 199-220, Url: <http://tomgruber.org/writing/ontologia-kaj-1993.htm>. Vegeu també, What is an Ontology? <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>

Guenther, R.S. (2003). MODS: The Metadata Object Description Schema. In: *Portal Libraries and the Academy*, 2003, 3(1). Url: http://muse.jhu.edu/login?uri=/journals/portal_libraries_and_the_academy/v003/3.1guenther.html

Guía Breve de web semàntica. W3C, © 1994-2005. Url: <http://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica>

Guidelines for repository implementers. Protocol Version 2.0, 2002-06-14. Document Version 2005/01/19T19:27:00Z. Url: <http://www.openarchives.org/OAI/2.0/guidelines-repository.htm>

Habing, T. (2007). METS, MODS and PREMIS, Oh My!: Integrating Digital Library Standards for Interoperability and Preservation. *NDIIPP ECHODEP Project*. ALA Summer, 2007. Url: <http://www.loc.gov/standards/mods/presentations/habing-ala07/>

Harper, C. (2006). Authority control for the semantic web. *Encoding Library of Congress subject headings. International Conference on Dublin Core and Metadata Applications, (Manzanillo, Mexico, June 20, 2008)*. Url: <http://hdl.handle.net/1794/3268>.

Harper, C; Tillett, B. (2007). Library of Congress controlled vocabularies and their application to the semantic web. *Cataloging and Classification Quarterly*, 2007, 43(3/4). Url: <https://scholarsbank.uoregon.edu/dspace/handle/1794/3269>.

Haslhofer, B., Isaac, A. (2011). Data. europeana. eu: The Europeana Linked Open Data Pilot. *International Conference on Dublin Core and Metadata Applications*, 2011, p. 94-104.

Hass Weinberg, B. (1996). Complexity In Indexing Systems : Abandonment And Failure : Implications For Organizing The Internet. *ASIS: Annual Conference Proceedings*, 1996. Url: <http://www.asis.org/annual-96/ElectronicProceedings/weinberg.html>

Heath, T. (2009). An introduction to Linked data (Tutorial). Url: <http://tomheath.com/slides/2009-02-austin-linkeddata-tutorial.pdf>

Heath, T., Bize, C. (2011). *Linked Data: Evolving the web into a Global Data Space* Morgan & Claypool, 136 p. Url: <http://linkeddatatobook.com/editions/1.0/>

Heery, R. (2000). Information gateways collaboration on content. *Online Information Review*, 2001, 42(1), p.40-45

Heery, R.; Carpenter, L.; Day, M. (2001). Renardus Project Developments and the Wider Digital Library Context. *D-Lib Magazine*, 2001, 7(4).

Hickey, T.B. (2000). Corc: System for gateway solution. *Online Information Review*, 2000, 24(1), p.49-53

Hillman, D., et al. (2010). RDA Vocabularies: Process, Outcome, USE. *D-Lib Magazine*, 2010, 16(1/2).

HILT (High-Level Thesaurus). Url: <http://hilt.cdlr.strath.ac.uk/>

Hiom, D. (2000). SOSIG : an Internet hub for the social sciences, business and law. *Online Information Review*, 2000, 24(1), p.54-58

Hiom, D. (2006). Retrospective on the RDN. *Ariadne Issue 47*, April 2006. Url: <http://www.ariadne.ac.uk/issue47/hiom>

Hohlfeld, M. (2011). Mitteilung des Vorstands zur Auflösung des vascoda e.V., 8 Dezember, 2011. Url: <http://vascoda.wordpress.com/2011/12/08/mitteilung-des-vorstands-zur-aufloesung-des-vascoda-e-v/#more-27967>

Hjørland, B. (1997). Information seeking and subject representation: An activity-theoretical approach to information science. Westport, CT: Greenwood Press, 1997.

Hjørland, B. (1992). The concept of 'subject' in information science, *Journal of Documentation*, 48, 2, p. 172-200.

Hollink, L., et al. (2008). Two variations on ontology alignment evaluation: Methodological issues. *The Semantic web: Research and Applications Lecture Notes in Computer Science*, SpringerLink, 2008, 5021, p. 388-401.

Hu, W.; Qu, Y. (2008). Falcon-AO: A practical ontology matching system. *web Semantics: Science, Services and Agents on the World Wide Web*, 2008, 6(3), p.237-239.

Hunter, P. (2003). Editorial Introduction to Issue 34: Cultivating Interoperability and Resource-Sharing. *Ariadne*, January 2003 (34). Url: <http://www.ariadne.ac.uk/issue34/editorial>

Huxley, L. (2001). Renardus: fostering collaboration between academic subject gateways in Europe. *Online Information Review*, 2001, 25(2), p.121-127

IFLA Classification and Indexing Section. Working Group on Guidelines for Multilingual Thesauri. *Guidelines for multilingual thesauri (IFLA Professional Reports, no. 115)*. The Hague: International Federation of Library Associations and Institutions, 2009. Url: <http://archive.ifla.org/VII/s29/pubs/Profrep115.pdf>

IFLA CDNL Alliance for Bibliographic Standards (ICABS). IFLA, 2005-2005. Url: <http://archive.ifla.org/VI/7/annual/icabs-sp2005-2006.pdf>

Imesh Home Page (1999-). Url: <http://www.imesh.org/>

Imesh Toolkit. Url: <http://www.imesh.org/toolkit/>

Imesh Toolkit: An Architecture and Toolkit for Distributed *Subject Gateways*. [Presentació]. Rachel Heery, R & D Team Leader, UKOLN, University of Bath 1999-2002. Url:

http://www.dli2.nsf.gov/ukworkshop/presentations/almasy_uwub/imeshtk.pdf
[Darrer accés: desembre 2011; Actualment web arxivat a: <http://www.jisc.ac.uk/whatwedo/programmes/nsf/imesh.aspx>]

L'Indexation. Jean-Michel Jolion (dir.). *Document numérique*, 4(1-2/2000). Paris: Hermes Science, 2001. 182 p. ISBN: 9782746202245.

Information Ecology: Mastering the Information and Knowledge Environment. Thomas Davenport and Laurence Prusak (ed.). New York: Oxford University Press, 1997, p. 86

INSPIRE Home Page. Url: <http://inspire.jrc.ec.europa.eu/index.cfm>

Interoperability levels for Dublin Core Metadata. DCMI, 01/05/2009. Url: <http://dublincore.org/documents/interoperability-levels/>

Instrumentos para la recogida de datos. Universidad Autonoma de Madrid. Url: http://www.uam.es/personal_pdi/stmaria/jmurillo/Met_Inves_Avan/Presentacion_es/Cuestionario_Escalas.pdf

Intute About. Electronic Libraries (eLib) Programme Origins. Url: <http://www.INTUTE.ac.uk/about.html>

Intute: Health and Life Sciences Evaluation Guidelines. Url: www.intute.ac.uk/healthandlifesciences

Intute Home Page. Url: <http://www.intute.ac.uk/>

Intute Staff. (2011). The future of Informs and the Virtual Training Suite. INTUTE Staff, April 11th, 2011. Url: <http://www.INTUTE.ac.uk/blog/2011/04/11/the-future-of-informs-and-the-virtual-training-suite/>

Isaac, A; Haslhofer, B. (2013). Europeana Linked Open Data - data.europeana.eu. *Semantic web*, 2013, 4(3), p. 291-297. Url: <http://www.informatik.uni-trier.de/~ley/pers/hd/i/Isaac:Antoine>

Isaac, A., et al. 2009. Evaluating thesaurus alignments for semantic interoperability in the library domain. *Intelligent Systems IEEE, 2009*, 24(2), p. 76-86.

ISO Standard 5964:1985. Guidelines for the establishment and development of multilingual thesauri. – 1st ed. / prepared by the Technical Committee ISO/TC 46, Documentation. [Geneva] : ISO, 1985. – iii, 61 p. Url: <http://www.iso.org/iso/en/ISOOnline.frontpage>

ISO 2709:1996, Information and documentation – Format for Information Exchange

ISO 2709:2008, Information and documentation – Format for Information Exchange

ISO 25964-1:2011: Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval

Jean-Mary, Y.R., et al. (2009). Ontology Matching with Semantic Verification. *web Semantics Online*, 2009, 7(3), p. 235-251. Url: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2825706/#!po=2.77778>

Joyce, A; Wickham, J.; Stephens, Ch. (2008). INTUTE Integration. *Ariadne*, 2008, (55).

Kahn, B. K., Strong, D.M. (1998). Product and Service Performance Model for Information Quality: An Update. In: *Proceedings of the 1998 Conference on Information Quality, Cambridge, MA, pp. 102-115, October 23-25, 1998.*

Kahn, B.K, Strong, D.M; Wang, R.Y. (1999). Information Quality Benchmarks: Product and Service Performance. *Communications of the ACM*, Forthcoming, 2002, 45(4), p.184-192.

Kalinichenko, L.A., et al. (2000). Infrastructure of the subject mediating environment aiming at semantic interoperability of heterogeneous digital library collections. *Proc. Of the 2nd Russian Scientific Conf. On Digital Libraries : Advanced Methods and Technologies*, 2000, p. 78-90. Url: <http://rcdl.ru/doc/2000/039.pdf>

Kempf, K., (2006). Vascoda and the Subject-based Gateways-the German Answer to Visibility and Accessibility in Collection Development. *Liber Quarterly*, 2006, 16.

Kim, T.P. (1994). Toward Theory of User-Based Relevance: a call for a New Paradigm of Inquiry. *Journal of the American Society for Information Science*, 1994, 46, p. 135-141.

Kirriemuir, J., et al., (1998). Cross-Searching *Subject Gateways*: The Query Routing and Forward Knowledge Approach. *D-Lib Magazine*, 1998, 1(15), 1. ISBN: 1082-9873.

Koch, T. (2006). Electronic thesis and dissertations services: semantic interoperability, subject access, multilinguality. *E-Thesis Workshop (Amsterdam : 19-20 January, 2006)*.

Koch, T. (2000). Quality-controlled subject gateways: definitions, typologies, empirical overview. *Online Information Review*, 2000, 24(1), p. 24-34.

Koch, T. (1999). Automatic classification and content navigation support for web services. DESIRE II; OCLC, June 2000. Url: www.oclc.org/oclc/research/publications/review98/koch_vizinegoetz/automatic.htm#Zthes

Koch, T. (1996). Controlled vocabularies, thesauri, classification schemes. [Online bibliography]. Url: <http://www.lub.lu.se/metadata/subjects-help.html>.

Koch, T., et al. (2003). VASCODA: A German Scientific Portal for Cross-Searching Distributed Digital Resource Collections, Research and Advanced Technology for Digital Libraries. Berlin: Springer Berlin Heidelberg, p. 257-262.

Koch, T.; Day, M. (1997). The role of classification schemes in Internet resource description and discovery. *DESIRE D3.2*, (3). Url: <http://www.ukoln.ac.uk/metadata/desire/classification/>

Koch, T., Neuroth, H.; Day, M. (2003). Renardus: Cross-browsing European subject gateways via a common classification system (DDC). McIlwaine, I. C. (ed.). *Subject retrieval in a networked world: proceedings of the IFLA Satellite Meeting (Dublin, OH, 14-16 August 2001)*, 25 ed. München: K.G. Saur, p. 25-33 (UBCIM Publications, New Series). Url: <http://www.lub.lu.se/~traugott/drafts/preifla-final.html> ; <http://www.ukoln.ac.uk/metadata/renardus/papers/ifla-satellite/ifla-satellite.pdf>

Koch, T., Neuroth, H.; Day, M. (2001). DDC mapping report. Renardus *D7.4*, 2001. URL: <http://renardus.sub.uni-goettingen.de/wp7/d7.4/>

Korfhage, R. (1997). Information Storage and Retrieval. New York: John Wiley, 1997.

Lacy, LW. (2005). Owl: Representing Information Using the web Ontology Language. Trafford Publishing; Victoria, BC, Canada: 2005.

Lancaster, W. (2003). Indexing and abstracting in theory and practice. 3rd Edition. London, Library Association, 2003.

Lancaster, W. (2002). El Control del vocabulario en la recuperación de información. 2ª ed. Corr. València: Universitat de València, 2002. 286 p.

Langridge, D.W. (1989). Subject Analysis: Principles and Procedures. London: Bowker-Saur, 1989.

Lassila, O.; McGuinness, D.L. (2001). The Role of Frame-Based Representation on the Semantic web, Knowledge Systems Laboratory Report KSL-01-02, Stanford University, 2001; Also appeared as Linköping Electronic Articles in Computer and Information Science, Vol. 6 (2001), No. 005, Linköping University, 2001. Url: http://www.ksl.stanford.edu/KSL_Abstracts/KSL-01-02.html

Laudon, K.C.; Laudon, J.P. (1998). Management Information Systems, (2nd edition), Macmillan, 1988.

Lauser, B.; Johannsen, G.; Caracciolo, C. (2008). Comparing Human and Automatic Tesauro Mapping Approaches in the Agricultural Domain. *Proc.Int. Conf. On Dublin Core and Metadata Application (Berlin : 22-26 September, 2008)*

Law, D.; Dempsey, L. (2000). A policy context - eLib and the emergence of the subject gateways. *Ariadne*, 2000 (25).

LawCite Overview. Url: <http://www.AustLII.edu.au/LawCite/doc/overview.html>

Lee, W., et al. (2002). AIMQ: methodology for information quality assessment. *Information & Management*, 40, 133-146

Library Linked Data Incubator Group Final Report. W3C Incubator Group Report, 25 October 2011. Url: <http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/>

Library Linked Data Incubator Group: Use Cases. W3C Incubator Group Report 25 October 2011. Url: http://www.w3.org/2005/Incubator/ld/XGR-ld-usecase-20111025/#Description_of_Use_Cases

Library of Congress Standards. Url: <http://www.loc.gov/standards/>

LIBRIS Quick Guide. Url: http://librishelp.libris.kb.se/help/libris_quickguide_eng.pdf

Linked Data Community. Url: <http://linkeddata.org/home>

Linked Data Definition. Wikipedia. Url: http://en.wikipedia.org/wiki/Linked_data

Linked Data Platform Working Group (Wiki). Url: http://www.w3.org/2012/ldp/wiki/Main_Page

Linked Open Data (LOD). Url: http://en.wikipedia.org/wiki/Open_Data

López-Pellizer, J. (2012). Lenguajes de representación: OWL. *Curso de Archivos y gestión de documentos ante la web semántica (Ejea de los Caballeros, Zaragoza, 16 de Julio de 2012)*. Zaragoza: Universidad de Zaragoza, 2012.

López-Huertas, MJ. (2007). Gestión del conocimiento multidimensional en los sistemas de organización del conocimiento. *VIII Congreso ISKO-ESPAÑA (León : 18, 19 y 20 de abril, 2007)*, p. 1-26

Lytras, M., M.-A. Sicilia, et al. (2005). Digital libraries in the knowledge era: Knowledge management and Semantic web technologies. *Library Management*, 2005, 26(4), p. 170-175.

MADS. Library of Congress, May 3, 2012. Url: <http://www.loc.gov/standards/mads/> [Darrera Consulta: 06/09/2012]

MADS/RDF Namespace Document. 10 May, 2012. Url: <http://www.loc.gov/standards/mads/rdf/v1.html#>

Malmsten, M. (2008). Making a Library Catalogue Part of the Semantic web. *Proc. Int'l Conf. on Dublin Core and Metadata Applications (Berlin: 22-26 September, 2008)*. Url: <http://www.kb.se/dokument/Libris/artiklar/Project%20report-final.pdf>

Manual de SKOS. Juan Antonio Pastor Sánchez i Francisco Javier Martínez Méndez (traducció). W3C, 2009. Url: <http://skos.um.es/TR/skos-primer/>

Manuals, guidelines & user guides. Australian Libraries. Url: <http://www.nla.gov.au/librariesaustralia/training-support/manuals-guides/>

MARC Standards Office. Url: <http://www.loc.gov/marc/ndmso.html>

MARC21 Format for Bibliographic Data. Library of Congress Network Development and MARC Standard Office. Versió actualitzada nº17, Setembre 2013. Url: <http://www.loc.gov/marc/bibliographic/>

MARC 21 Format for Bibliographic Data. Library of Congress, actualització Nº13, setembre 2011. Url: <http://www.bnc.cat/Professionals/Catalogacio/Publicacions-tecniques/MARC-21-2010>

MARC 21: Harmonized USMARC and CAN/MARC. Network Development and MARC Standards Office Library of Congress, October 22, 1998. Url: <http://www.loc.gov/marc/annmarc21.html>

MARC/RDF. MARC21 vocabularies from Metadata Management Associates. Open Metadata Registry: Supporting Metadata Interoperability. Url: <http://marc21rdf.info/>

Marc Standards Office. Url: <http://www.loc.gov/marc/ndmso.html>

MARCXXML. Library of Congress, April 13, 2012. Url: <http://www.loc.gov/standards/marcxml/> [06/09/2012]

Marine Metadata Interoperability Project (MMI). Content Standard References. Url: <https://marinemetadata.org/conventions/content-standards>

Marshall, C. C. and F. M. Shipman (2003). Which semantic web? *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*. Nottingham, UK, ACM.

Meadow, C.T; Boyce, B.R.; Kraft, D.H. (2000). "Multidatabase Searching and Mapping". Dins: *Text Information Retrieval Systems*. 2 ed. Estats Units, etc. : Academic press, 2000.

Mayr, P., et al. (2007). Results from a German terminology mapping effort: intra- and interdisciplinary cross-concordances between controlled vocabularies. *6th European Networked Knowledge Organization Systems*

(NKOS) Workshop at the 11th ECDL Conference (Budapest, Hungary, 21. September 2007)

Mayr, P., Petras, V. (2008). Cross-concordances: terminology mapping and its effectiveness for information retrieval. arXiv preprint arXiv:0806.3765.

McCulloch, E. (2009). HILT: Enhancing subject search through embedded web services. Dins: JISC Conference 2009. 2009: Edinburgh International Conference Centre.

Méndez, E. (1999). "RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio". *7es Jornades catalanes de documentació*, (Barcelona, 4, 5 i 6 de novembre de 1999). Url: <http://www.cobdc.org/jornades/7JCD/1.pdf>

Méndez, E. (2006). Dublin Core, metadatos y vocabularios. *El profesional de la información*, marzo-abril 2006, 15(2).

Messer-Davidow, E. (1993). *Knowledges: Historical and critical studies in disciplinaryity*. Charlottesville, VA: University Press of Virginia, 1993.

Metadata Set. VEST Registry. Agriculture Information Management Standards. FAO. Url: http://aims.fao.org/vest-registry?type_1%5B%5D=metadataset . [Darrer accés: 15/08/2012]

METS. Library of Congress, August 30, 2012. Url: <http://www.loc.gov/standards/mets/>

Miller, E. (2013). Translating the Library Catalog from MARC into *Linked Data*: An Update on the Bibliographic Framework Initiative. *NISO/DCMI Joint Webinar Series, 2013* Url: <http://www.niso.org/news/events/2013/dcmi/bibframework>

Miller, H. (1996). The Multiple Dimensions of Information Quality. *Information Systems Management*, 1996, 13(2), p.79-82.

Milstead, J. (1994). Needs for Research in Indexing. *Journal of the American Society for Information Science*, 1994, 45(8), p. 577-82.

Missingham, R. (2004). Reengineering a National Resource Discovery Service: MODS Down Under. *D-Lib Magazine*, 2004, 10(9).

MODS standard: <http://www.loc.gov/standards/mods/>

MODS user guide. Versió 3. Darrera actualització: July 12, 2013. Url: <http://www.loc.gov/standards/mods/userguide/>

MODS: uses and features. Library of Congress, July 10, 2012. Url: <http://www.loc.gov/standards/mods/mods-overview.html>

Moen, W.E. (2001a). Assessing interoperability in the networked environment: Standards, evaluation, and testbeds in the context of Z39.50. In C. R. McClure and J. C. Bertot (Eds.), *Evaluating Networked Information Services: Techniques, Policy, and Issues* (pp.85-109). Information Today

Moen, W.E. (2001b). Mapping the Interoperability Landscape for Networked Information Retrieval. *Proceedings of First ACM/IEEE-CS Joint Conference on Digital Libraries (Roanoke, VA, June 24-28, 2001)*, New York: The Association for Computing Machinery, 2001, p. 50-52.

Moens, M-F. (2000). *Automatic indexing and abstracting of document texts*. Boston: Kluwer, 2000.

Monnet: Multilingual Ontologies for Networked Knowledge Project. Funded by EU FP7. FP7-ICT-4-248458. European Commission, 2012.

Morshed, A. (2013). Role of Vocabularies for Semantic Interoperability in Enabling the Linked Open Data Publishing. *Cases on Open-Linked Data and Semantic Web Applications*, IGI Global, p. 84-104.

Morshed, A., et al. (2011). Thesaurus Alignment for Linked Data Publishing. *Proc. Int'l Conf. on Dublin Core and Metadata Application (The Hague, The Netherlands, 21-23 September, 2011)*.

Morville, P.; Rosenfeld, L. (2006). *Information Architecture for the World Wide Web: Designing Large-Scale Web Sites*, 3rd Edition. Sebastopol, CA: O'Reilly & Associates, Inc., 2006.

Müller, R., Niederée, C., Schmidt, J.W. (2000). Design Principles for Internet Community Information Gateways: MARINFO – A Case Study for a Maritime Information Infrastructure. 1st International Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT'2000), (Potsdam, Berlin, Germany: March 30 – April 2, 2000). Url: <http://www.sts.tu-harburg.de/papers/2000/MNS00.pdf>

Naumann, F., Rolker, C., (2000). Assessment methods for information quality criteria. *Proceedings of the International Conference on Information Quality (IQ)*. (Cambridge, MA, 2000), p. 148-162.

Naumann, F. Sattler, K-U. (2006). *Information Quality Fundamentals, Techniques, and Use*. EDBT Tutorial, March 28, 2006.

A National Framework for the Development of Australian Subject Gateways. Australian Subject Gateways Forum. Url: www.nla.gov.au/initiatives/sg/paper1999-07.html [Darrer accès: 30/12/2010]

Neuroth, H; Koch, T. (2001). Cross-browsing and cross-searching in a distributed network of subject gateways: Architecture, data model and classification. *Proceedings, ELAG 2001. Integrating Heterogeneous Resources*

(Prague, June 6-8, 2001). *European Library Automation Group*, 2001. Url: http://renardus.sub.uni-goettingen.de/conferences/elag2001/neuroth_koch_elag2001.html

Neuroth, H. (2002). Suche in verteilten 'Quality-controlled Subject Gateways': Entwicklung eines Metadatenprofils. Searching in 'Quality-controlled Subject Gateways': development of metadata profiles. *Bibliothek Forschung und Praxis* 26, p. 275-296.

Neuroth, H.; Koch Traugott. (2001). Cross-browsing and cross-searching in a distributed network of subject gateways: Architecture, data model, and classification. Url: <http://old.stk.cz/elag2001/Papers/HeikeNeuroth/HeikeNeuroth.html#chap2>

Neuroth, H.; Pianos, T., (2003). Vascoda: A german scientific portal for cross-searching distributed digital resource collections, *Research and Advanced Technology for Digital Libraries*. Springer, p. 257-262.

NBII Thesaurus: <http://thesaurus.nbi.gov/portal/server.pt>

O'Connor, B.C. (1996). *Explorations in Indexing and Abstracting: Pointing, Virtue, and Power*. Englewood, Colorado: Libraries Unlimited, 1996.
OAEI Initiative (Ontology Alignment__Evaluation__Initiative). URL: <http://oaei.ontologymatching.org/>

Obrst, L. (2006). *Ontologies & Databases : Similarities & Differences*. MITRE, October 12, 2006. Url: http://ontolog.cim3.net/file/work/DatabaseAndOntology/2006-10-12_DatabaseAndOntology-MiniSeries-kickoff/Ontology-Database_Similarities-Differences--LeoObrst_20061012.ppt.

Obrst, L. (2006). *The Ontology Spectrum & Semantic Models*. MITRE, January 12 & 19, 2006. Url: http://ontolog.cim3.net/file/resource/presentation/LeoObrst_20060112/Ontology_SpectrumSemanticModels--LeoObrst_20060112.ppt

Obrst, L. (2003). *Ontologies for semantically interoperable systems*, Proceedings of the twelfth international conference on Information and knowledge management. New Orleans, LA, USA: ACM, 2003.

OCLC. Mapping new LCSH with DDC numbers. Url: <http://www.oclc.org/dewey/updates/numbers/> [Darrer Accès: Gener, 2011]; Aradins: BISAC Project (2011?-).

Okalova, M. (2006). VIAF virtualny medzinarodny subor autorit. [VIAF virtual international authority file.]. *Kniznica* 7(9): 42-48.
OpenAgris, FAO. Url: <http://aims.fao.org/openagris>

Opening up Government. Url: <http://data.gov.uk> PADI. Url: <http://nla.gov.au/nla.arc-10691>

OWL Web Ontology Language Reference, Mike Dean, Guus Schreiber, Editors, W3C Recommendation, 10 February 2004. Última versió accessible a <http://www.w3.org/TR/owl-ref/> (update 2009)

PADI: <http://www.nla.gov.au/padi/>

PANDORA: Australia's Web Archive. Url: <http://pandora.nla.gov.au/abstracting.html>

Pastor, J.A., Martínez, F.J.; Rodríguez, J.V. (2009). Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. *Information Research*, 2009, 14(4) paper 422. Url: <http://InformationR.net/ir/14-4/paper422.html>

Patel, M.; Koch, T.; Doerr, M. (2004). Semantic Interoperability in Digital Library Systems. *Project nº507618 DELOS : A Network of Excellence on Digital Libraries*. European Commission, 2004.

Pavel, S.; Euzenat, J. (2013). Ontology matching: state of the art and future challenges. *Knowledge and Data Engineering, IEEE Transactions*, 2013, 25(1), p. 158-176.

Peis, E., E. Herrera-Viedma, et al. (2008). A Semantic Service Model of Selective Dissemination of Information (SDI) for Digital Libraries. *Profesional de la Informacion* 17(5): 519-525

Peset, F., Ferrer-Sapena, A., Subirats-Coll, I. (2011). Open data y Linked Open data: su impacto en el área de Bibliotecas y Documentación. *El Profesional de la Información*, 2011, 20(11), p. 165-173. Url: <http://www.elprofesionaldeinformacion.com/contenidos/2011/marzo/06.pdf>

Pinakes Directory of Subject Gateways; <http://www.hw.ac.uk/libwww/irn/pinakes/pinakes.old>

Preguntes freqüents sobre l'accessibilitat web. Direcció General d'Atenció Ciutadana. Generalitat de Catalunya, p. 17. Url: <http://www.gencat.cat/web/cat/accessibilitat/Accessibilitat-PreguntesFreqüents.pdf>

Preguntas frecuentes sobre que el lenguaje de ontologías Web (OWL). Carlos Iglesias Moro (traducció). "Frequently Asked Questions on W3C's Web Ontology Language (OWL)". W3C, 2005. Url: <http://www.w3c.es/Traducciones/es/SW/2005/owlfaq>

PREMIS. Library of Congress, July 22, 2009. Url: <http://www.loc.gov/standards/premis/>

Projecte DESIRE (1998). Url: <http://www.desire.org/> [Darrer accés: 07/01/2013]

Projecte DESIRE (1998). Quality. Url: <http://www.ukoln.ac.uk/metadata/desire/quality/> [Darrer accés: 07/01/2013]

Projecte DESIRE I: <http://www.ukoln.ac.uk/metadata/desire/phase-1/>

Projecte DESIRE II: <http://www.ukoln.ac.uk/metadata/desire/>

Policy and Standards Division Annual Report 2009. Library of Congress, 09/11/2009. Url: <http://www.loc.gov/catdir/cpso/annreppsd09.pdf> [

Policy and Standards Division Annual Report FY 2010. Library of Congress, 20/06/2011. Url: <http://www.loc.gov/catdir/cpso/annreppsd2010.pdf>

Purday, J. (2009). "Think culture: Europeana. eu from concept to construction." *Electronic Library, The* 27(6): 919-937.

RAMEAU. Url: <http://www.cs.vu.nl/STITCH/rameau/>

RDF Primer, Frank Manola, Eric Miller, Editors, W3C Recommendation, 10 February 2004. Última versió accessible a <http://www.w3.org/TR/rdf-primer/> (update 2009)

RDF. World Wide Web Consortium. Resource Description Framework (RDF). Url: <http://www.w3.org/RDF>

RDF/XML. Expressing Simple Dublin Core in RDF/XML (SUPERSEDED, SEE [DC-RDF](#) and [notes](#)). Url: [http://dublincore.org/documents/dcmes-xml/RFC-3986-Uniform-Resource-Identifiers-\(URI\)-Generic-Syntax](http://dublincore.org/documents/dcmes-xml/RFC-3986-Uniform-Resource-Identifiers-(URI)-Generic-Syntax), Tim Berners-Lee, Roy Fielding, Larry Masinter, IETF, January 2005. Disponible en <http://tools.ietf.org/html/rfc3986>

Redland Rasqal RDF Query Demonstration. Url: <http://librdf.org/query/>

Renardus: <http://renardus.sub.uni-goettingen.de/>

Reverté, C.; Sebastià-Salat, M. (2009). Collaboration Model to ensure the information access in aquatic science. *IAMSLIC Conference (27 September – 1 October, Bruges, Belgium, 2009)*.

Reverté, C.; Sebastià-Salat, M. (2009). KOS interoperability through mapping, in the aquatic sciences field. *The 8th European Networked Knowledge Organisation Systems & the 13th European Conference on Digital libraries (Corfú, Greece, 27September- 2 October, 2009)*.

Reverté, C.; Sebastià-Salat, M. (2009). Aquatic science subject gateway project as a model of interoperability. *ISKO UK Conference 2009 : Content Architecture : Exploiting and Managing Diverse Resources (London, 23 - 24 June, 2009)*. [Publicat en Aslib Proceedings]

Reverté, C. (2008). Aquatic Science Information Center in IRTA. *First Workshop of the Mediterranean Marine and Aquatic Science Libraries and Information Centers*. Cádiz, 17 - 18 November, 2008.

Reverté, C. (2006). Estandardització i passarel·les temàtiques. Projecte tesi (DEA). Doctorat en informació i documentació en l'era digital (Bienni 2006-2008). Montserrat Sebastià i Salat (tutora). Universitat de Barcelona. Facultat de Biblioteconomia i Documentació, Octubre 2006.

Richardson, S.; Powell, A. (2003). Exposing Information Resources for E-learning: An introduction to the IMS Digital Repositories Working Group. *Ariadne*, 14 January 2003, (34). Url: <http://www.ariadne.ac.uk/issue34/powell>

Rüther, M., et al. (2010). Linked environment data for the life sciences. Url: <http://arxiv.org/ftp/arxiv/papers/1012/1012.1620.pdf> [Document arxivat: arXiv preprint arXiv:1012.1620]

Rodríguez Yunta, L. (2004). Pasarelas temáticas en Internet: un modelo de directorio basado en la aplicación de técnicas documentales. *Hipertext.net*, núm. 2, 2014. Url: <http://www.upf.edu/hipertextnet/numero-2/directorio.html>

Rodríguez Yunta, L., Giménez Toledo, E. (2004). Más enllà de la usabilitat: característiques mínimes exigibles per a les interfícies de bases de dades web [en línia]. bid: textos universitaris de biblioteconomia i documentació, núm.13, 2004. Url: <http://bid.ub.edu/13rodri1.htm>

Rybarczyk, F. [s.a.] "3.4. Harvesting, indexing and automated metadata collection". Dins: *DESIRE Information Gateways Handbook*, s.a. URL: <http://www.desire.org/handbook/3-4.html> [Darrer accés: 24/10/2010]

S-Match (Open Source): <http://semanticmatching.org/s-match.html>

Salton, G. y M. J. McGill. (1983). *Introduction to Modern Information Retrieval*. New York: McGraw Hill, 1983

Saracevic, T. et al. (1988). A study of information seeking and retrieving, background and methodology. *Journal of the American Society for Information Science*, 1988, 39(3) p. 161-176

Saracevic, T. (1997). Relevance: A review of and a framework for the thinking on the notion in information science. (paper review) En *Readings in Information Science* edited by Karen Spark Jones, Peter Willet. San Francisco: Morgan Kaufmann Publisher, 1997.

Schmidt, J., Horn, A.; Thorsen, B. (2003). Australian Subject Gateways, the successes and the challenges. *World Library and Information Congress : 6th IFLA Congress and Council (Berlin, 1-9 August, 2003)*

Sharffe, F. (2006). *WP2: Ontology Management D2.6 Mapping tool Background v3*.

Shiri, A.A., et al. (2004). User evaluation of a pilot terminologies server for a distributed multi-scheme environment. *Online Information Review*, 2004, 28, p. 273-283.

Shiri, A. (2003). Schemas and ontologies: building a semantic infrastructure for the GRID and digital libraries. *Workshop Report from E-Science Institute*, Edinburgh, 2003.

Shirley, C.; Ashley S. (2006). Incorporating a virtual union catalogue into the wider information environment through the application of middleware: Interoperability issues in cross-database access. *Journal of Documentation*, 2006, 62(1), p. 120-144

Shvaiko, P.; Euzenat, J. (2013). Ontology matching: state of the art and future challenges. *Knowledge and Data Engineering, IEEE Transactions*, 2013, 25(1), 158-176

SINO documentation. AustLII. Url: http://www.AustLII.edu.au/help/sino_full.html

[SKOS Reference](#), Alistair Miles, Sean Bechhofer, Editors, W3C Recommendation, 18 August 2009. [Última versió](#) accessible a <http://www.w3.org/TR/skos-reference>

SKOS. Tutorials, Presentations & Papers. URL: <http://www.w3.org/2004/02/skos/references>

SKOS-2-HIVE WORKSHOP: Creating SKOS Vocabularies to Help Interdisciplinary Vocabulary Engineering. Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid. (Getafe, Madrid, June 19th, 2012). Url: <http://klignon.uc3m.es/hive-es/wiki/index.php/Workshop-es>

SKOS-Core Guide. Url: <http://www.w3.org/2001/sw/Europe/reports/thes/1.0/guide/>

SPARQL Query Language for RDF. 2008. Eric Prud'hommeaux, Andy Seaborne, Editors, W3C Working Draft, 15 January 2008. Url: <http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/>

Spoehr, D.; Hollink, L.; Cimiano, P. (2011). A Machine Learning Approach to Multilingual and Cross-Lingual Ontology Matching. *ISWC'11 Proceedings of the 10th international conference on The semantic web*. Berlin: Springer, 2011, Part I, p. 665-680.

State of progress in the development of guidelines to express elements of the INSPIRE metadata implementing rules using ISO 15836. INSPIRE, 2008. Pàgina 2. Url:

http://inspire.brgm.fr/Documents/MD_IR%20and%20DC_%20state%20of%20progr ess_20080507.pdf

Stvilia, B. (2007). A Framework for Information Quality Assessment. *Journal of the American Society for Information Science and Technology*, 2007, 58(12), p. 1720-1733

Stvilia, B., Gasser, L. (2008). Value based metadata quality assessment. *Library & Information Science Research*, 2008, 30(1), p. 67-74

SUB. Universitat de Gottingen. Url: <http://renardus.sub.uni-goettingen.de/>

Subject based gateways for archives. Url: <http://uow.libguides.com/content.php?pid=137632&sid=1208979>

Subject Indexing: Principles and Practice in the 90's. München: K.G. Saur, 1995.

Summers, E., et al. (2008). LCSH, SKOS and *Linked Data*. *Proc. International Conference on Dublin Core and Metadata Applications. (Berlin : 22-26 September, 2008)*. Url: <http://edoc.hu-berlin.de/conferences/dc-2008/summers-ed-25/PDF/summers.pdf>

Sutter, E. (1998). "Pour une écologie de l'information", *Documentaliste : Sciences de l'information*, 1998, 35(2), p. 83-86.

Tanenbaum, Andrew S; Van Steen, Maarten. (2006). *Distributed systems: principles and paradigms*. 2n Edition. NY: Pearson, 2006. ISBN 0-13-239227-5

Terris, G. (Book review of) Rowley, J.; Hartley, R. (2008). *Organizing knowledge: an introduction to managing access to information - 4th ed.* - Aldershot: Ashgate Publishing, 2008. ISBN 978-0-7546-4431-6 (Lang.: eng). – In: *Multimedia Information and Technology*, 34(2008)3, p.83-84.

Tillett, B.B. (2010). *Building Blocks for the Future: Making Controlled Vocabularies Available for the Semantic Web*, Getty Research Institute, Los Angeles, CA, 2010.

Tillett, B.B. (2011). Keeping Libraries Relevant in the Semantic Web with Resource Description and Access (RDA). *Serials* 24, p. 266-272.

Toth, M. (2010). Libraries in the World of the Semantic Web. *Könyvtari Figyelo*, 2010, 20, p. 413-438

Tudhope, D. and D. Cunliffe (1999). Semantically indexed hypermedia: Linking information disciplines. *Acm Computing Surveys* 31: U15-U20.

[Turtle - Terse RDF Triple Language](#), David Beckett, Tim Berners-Lee. W3C Team Submission, 14 January 2008. Última versió accessible a <http://www.w3.org/TeamSubmission/turtle/>

UDC - Universal Decimal Classification, UDC Consortium. Url: <http://www.udcc.org/>

Understanding Metadata. Bethesda: NISO Press, 2004, p.1. Url: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

UKOLN: Consorci d'Arxius, Biblioteques i Museus del Regne Unit. Url: <http://www.ukoln.ac.uk/>

van Hage, W.R; et al. (2010). The OAEI food task: An analysis of a thesaurus alignment task. *Applied Ontology*, 2010, 5(1), p. 1-28.

van Hage, W.R. (2007). OAEI 2007 environment Trilateral Thesaurus Mapping Task. Url: <http://www.few.vu.nl/~wrvhage/oeai2007/environment.html>

VASCODA Home: <http://www.vascoda.de/> [veure també Vascoda blog]

Vascoda Blog. Url: <http://vascoda.wordpress.com/>

Virtuoso SPARQL Query Editor. Url: <http://demo.openlinksw.com/sparql>

Wang, R.Y.; Strong, D. M. (1996). Beyond accuracy: what data quality means to data consumers. *J. Manage. Inf. Syst.*, March, 1996, 12 (4), p. 5-33.

Watkins, G., Joyce, A., (2009). Intute: A British Online Gateway to Slavic and East European Resources. *Slavic & East European Information Resources*, 2009, 10, p. 267-275.

Weibel, S. L. (2007). Social bibliography: a personal perspective on libraries and the Semantic Web. *Cataloging & Classification Quarterly* 43(3/4): np.

Wendler, R. (1999). LDI Update: Metadata in the Library. *Library Notes*, July/August 1999, (1286), p. 4-5

Wikimedia Commons. Url: http://commons.wikimedia.org/wiki/Main_Page

Williams, C. (2006). INTUTE: The New Best of the Web. *Ariadne*, 2006, (48).Url: <http://www.ariadne.ac.uk/issue48/williams>

WorldWideWeb: Proposal for a HyperText Project. W3C, 1990. Url: <http://www.w3.org/Proposal>

Wp12: Cross concordances of classifications and thesauri. Dins: Projecte CARMEN, 2000. URL: <http://www.bibliothek.uni-regensburg.de/projects/carmen12/index.html.en>

Xsearch Overview. Url: <http://www.easexp.com/xsearch/>

XML-Schema. W3C, 2012/06/22 Url: <http://www.w3.org/XML/Schema.html>

Yee, M.M. (2009). Can Bibliographic Data be Put Directly onto the Semantic Web?. *Information Technology and Libraries*, 2009, 28, p. 55-80.

Zeng, M.L.; Chan, L.M. (2006). Metadata interoperability and standardization-A study of methodology, Part II. *D-Lib Magazine*, 2006, 12, p. 1082-9873.

Zeng, M.L.; Chan, L.M. (2004). Trends and issues in establishing interoperability among knowledge organization systems / Maria Lei Zeng and Lois Mai Chan. *Journal of the American Society for Information Science & Technology*, 2004, 55 (5), p. 377-395.

Zhiping, Z. and Z. Qinglai (2006). Advances in semantic web services. *Journal of the China Society for Scientific and Technical Information* 25(4): 462-467.

Zhu, J., et al. (2008). DYNIX: A novel meta-search engine for metadata based cross search, *Applications of Digital Information and Web Technologies, 2008. ICADIWT 2008. First International Conference on the. IEEE*, p. 204-209.

Zhu, H.; Wang, R.Y. (2007). An Information Quality Framework for Verifiable Intelligence Products. Book Chapter. Eds. Chan, Y., Talburt, J.R., and Talley, T., Forthcoming, Springer. Hongwei Zhu and Richard Wang. URL: <http://mitiq.mit.edu/Documents/Publications/Papers/2007/An%20IQ%20Framework%20for%20Verifiable%20Intelligence%20Products.pdf>

Zthes. URL: <http://zthes.z3950.org/>

11. GLOSSARI

ANSI/NISO: American National Standards Initiative Institute / National Information Standards Organization

API: Application Programmers Interface

CIP⁵⁹: és una evolució i refinament de Whois+++, protocol d'Internet per la cerca d'informació de recursos a les xarxes. CIP ofereix un camí als servidors d'informació per conèixer els continguts d'altres servidors d'informació per l'intercanvi d'índexs d'informació. Un cop els índexs són intercanviats, un servidor pot mirar en el seu propi índex per respondre a una interrogació, o mirar en els índexs rebuts des de altres servidors i veure si la interrogació potser resposta d'alguna manera. D'altra banda CIP és també un protocol d'indexació que defineix mètodes per crear i intercanviar índexs d'informació entre servidors d'informació. Distribueix cerques a través de diferents instàncies d'un tipus concret de cercador per crear un directori global. CIP pot unir bases de dades individuals amb dades distribuïdes en dipòsits de dades o repositoris.

Datasets: grups o conjunts de dades

DC: Dublin Core

DDC: Dewey Decimal Classification

DOI: Digital Object Identifier System

DTD: Document Type Definition

Etiquetes alternatives (*alternative label*): un concepte pot estar relacionat amb múltiples etiquetes (per exemple, etiquetes d'idiomes), però només un d'ells pot està associat a un idioma com a etiqueta preferida o terme preferit. Les etiquetes restants relacionats amb el concepte es coneixen com etiquetes alternatives.

FAO: Food and Agriculture Organization of the United Nations

FRBR: Functional Requirements for Bibliographic Records

HILT: High Level Thesaurus

HTML: HyperText Markup Language

⁵⁹ The Encyclopedia of Networking and Telecommunications. Ed. Tom Sheldon. McGraw-Hill's, 2001. 1296 pp. ISBN:978-0072120059

Interdisciplinarietat: DRAE⁶⁰ o defineix com “l’estudi o activitat que es realitza amb la cooperació d’altres disciplines”. Segons Lopez-Huertas (2007), es tracta de l’interrelació orgànica dels conceptes de diverses disciplines fins al punt d’unir-se en una espècie de “nova-unitat” que subsumeix en un nivell superior les aportacions de cada una de les disciplines particulars.

JISC: Joint Information Systems Committee

KOS: Knowledge Organization System

LCSH: Library of Congress Subject Headings

Linked Data: El terme *Linked Data* és definit per la pròpia comunitat⁶¹ com el fet de connectar dades web relacionades que no estan vinculades prèviament.

M2M: Machine to machine interaction

MADS: Metadata Authority Description Schema

Mapeig (Mapping): és pot definir com l’acció produïda pels processos de mapeig. Normalment aquests processos poden ser manuals (mapeig i/o alineament realitzat per una/varies persones), automàtic (acció realitzada per una màquina) i/o semiautomàtic (acció combinada per l’home i la el software/sistema de mapeig). ANSI/NISO ho defineix com un conjunt de correspondències entre categories, esquemes de noms o termes controlats. Aquestes correspondències o mapejos s'utilitzen per transformar dades o consultes d'un vocabulari per al seu ús amb un altre vocabulari.

MARC: Machine-Readable Cataloging

METS: Metadata Encoding and Transmission Standard

MODS: Metadata Object Description Schema

Multidisciplinarietat: el DRAE o defineix com a que “abarca o afecta a varies disciplines”. Segons Lopez-Huertas (2007)⁶²Es caracteritza perquè afecta a l’estudi d’un objecte d’una disciplina per varies disciplines a la vegada i també perquè suposa que el coneixement d’un objecte d’estudi que pertany a una disciplina es profunditza pel recolzament multidisciplinari de les demés.

OAI-PMH: The Open Archives Initiative Protocol for Metadata Harvesting

OCLC: Online Computer Library Center

⁶⁰ Diccionario de la lengua española. Accessible a: <http://www.rae.es/drae/>

⁶¹ Linked Data community. Accessible a: <http://linkeddata.org/home> [Última consulta: 30/12/2012]

⁶² López-Huertas, María José. 2007. Gestión del conocimiento multidimensional en los sistemas de organización del conocimiento. *VIII Congreso ISKO-ESPAÑA (León : 18, 19 y 20 de abril, 2007. Pp. 1-26*

Ontologia: una ontologia es podria definir com un vocabulari controlat específic per a la seva utilització en màquines o que és intel·ligible per les màquines. En la W3C⁶³, aquest concepte es defineix com a un “vocabulari, emprat en informàtica, per ser entès per màquines i que són especificats amb suficient precisió com per permetre diferenciar termes i citar-los de manera precisa”. A més a més una ontologia permet definir els termes a utilitzar per descriure i representar una àrea del coneixement, i potser utilitzada tant per aplicacions com persones o bases de dades que necessiten compartir un domini de informació. La diferència entre els vocabularis controlats tradicionals, és que les ontologies codifiquen tant el domini d'un coneixement com el coneixement que s'estén dels dominis, és a dir, permetent la reutilització del coneixement.

OPAC: Online Public Access Catalog

OWL: *Web Ontology Language* o Lenguatge d'ontologies web. Està dissenyat per a ser usat en aplicacions que necessiten processar el contingut de la informació en lloc de únicament representar informació per als humans. OWL facilita un millor mecanisme de interoperabilitat de contingut web que els mecanismes admesos per XML, RDF, i esquema RDF (RDF-S) proporcionant vocabulari addicional juntament amb una semàntica formal. OWL té tres subllenguatges, amb un nivell d'expressivitat creixent: OWL Lite, OWL DL, i OWL Full.s

Procés d'alineament: la OAEI, defineix el procés d'alineament com una funció (f) des de la qual un parell d'ontologies (o i o') s'han d'alinejar, i donen com a resultat una alineació (A), que amb un conjunt de paràmetres (p) i un conjunt de recursos (r), donen com a resultat una nova alineació A' que prové d'aquest parell d'ontologies: $A' = f(o, o', A, p, r)$

RDA: Resource Description Access

RDF: Resource Description Framework

RDN: Resource Discovery Network

Relació associativa (associative relationship): és una relació que s'estableix entre dos o més termes que formen part d'un vocabulari controlat, i que permet relacionar un terme amb altres termes que estan relacionats amb o associats amb ell. L'abreviatura que representa el concepte en un vocabulari controlat és RT o terme relacionat.

Relació d'equivalència (equivalence relationship): és una relació entre dos o més termes d'un vocabulari controlat que condueix a un o més termes que es poden utilitzar en lloc del terme al qual es fa la referència. Normalment en el

⁶³ Preguntas frecuentes sobre que el lenguaje de ontologías web (OWL). Carlos Iglesias Moro (traducció). “Frequently Asked Questions on W3C's web Ontology Language (OWL)”. W3C, 2005. Uri: <http://www.w3c.es/Traducciones/es/SW/2005/owlfaq>

vocabulari controlat s'identifica perquè el concepte comença amb la paraula SEE o USE.

Relació jeràrquica: tipus de relació que s'estableix entre dos o més termes dintre d'un vocabulari controlat, representant des d'un terme més ampli (genèric) a un terme més específic, o un tipus de relació que exemplifica un tot i una part del tot. Aquest tipus de relació queda representat pels conceptes “broader term” (BT) o “narrower term” (NT).

Semantic web: Una iniciativa de col·laboració liderada pel W3C, el web semàntic proporciona un marc comú que facilita l'intercanvi de dades i la reutilització de dades a través d'aplicacions, empreses, comunitats llinants.

SKOS-Core: Suporta el llenguatge de descripció RDF, orientat als sistemes d'organització del coneixement (KOS), com ara tesaurus, glossaris, vocabularis controlats, taxonomies i sistemes de classificació.

SKOS: “Simple Knowledge Organization System” o Sistema Simple de Organización del Conocimiento⁶⁴. Model per a la representació de l'estructura bàsica i el contingut d'esquemes de conceptes de vocabularis controlats tradicionals com tesaurus, esquemes de classificació, llistes d'encapçalaments de matèria, taxonomies, folksonomies i altres vocabularis controlats similars. És una aplicació de RDF (Resource Description Framework), que permet la creació i publicació de conceptes en la web, així com vincular-los amb dades en aquest mateix mitjà i fins i tot integrar-los en altres esquemes de conceptes.

SOAP: Simple Object Access Protocol

SPARQL: Llenguatge de consulta per a RDF

SQL: Structured Query Language

SRU: Search and Retrieve via URL.

SRW: Search and Retrieve Web Service

Subject Gateways (passarel·les temàtiques): Al llarg de la història han estat anomenades també com: *subject-based information gateways (SBIGs)*, *subject based gateways*, *subject index gateways*, *virtual libraries*, *clearinghouse*, *subject trees*, *pathfinders*, *quality controlled subject gateways*. Les passarel·les temàtiques són senzillament un sistema d'informació de qualitat que dona accés d'una forma fàcil als recursos d'informació d'una xarxa especialitzada en una matèria determinada. El tipus més simples de passarel·les temàtiques són conjunts de pàgines web que contenen llistes d'enllaços a recursos. Les

⁶⁴ Manual de SKOS. Juan Antonio Pastor Sánchez i Francisco Javier Martínez Méndez (traducció). W3C, 2009. Url: <http://skos.um.es/TR/skos-primer/>

passarel·les temàtiques com a terme va ser popularitzat en el Programa de biblioteques Electròniques Regne Unit (e-lib).

Terme específic (narrower term): és un terme que està subordinat a un altre terme o diversos termes d'una jerarquia. En els tesaurus, l'indicador de relació d'aquest tipus d'expressió és l'abreviatura NT.

Terme genèric (broader term): és un terme en el qual hi ha subordinats altres termes (un o més termes) dintre una jerarquia. En un tesaurus, l'indicador de relació del termes es representa per l'abreviatura BT.

Terme preferit (Preferred term): definit molts cops com a un simple descriptor que forma part d'un vocabulari controlat. ANSI/NISO (2005), ho defineix com un, dos o més sinònims o variants lèxiques seleccionades com a un terme per a la seva inclusió en un vocabulari controlat. Vegeu també el terme no preferit o *Non-prefered term*.

Tesaurus: en ANSI/NISO Z39.19-2005 és definit com un vocabulari controlat disposat en un ordre conegut i estructurat de manera que les diverses relacions entre els termes es mostren de forma clara i identificats pels indicadors de parentiu estandarditzats. Indicadors de relació s'han d'utilitzar recíprocament. El seu propòsit és promoure la coherència en la indexació d'objectes de contingut, especialment per a informació post-coordinada dels sistemes de recuperació i emmagatzematge de la informació, i per facilitar la navegació i la recerca mitjançant la vinculació de termes amb els termes. Els tesaurus també poden facilitar la recuperació d'objectes de contingut en recerca de text lliure.

Top term: El terme més ampli en una jerarquia de vocabulari controlat, a vegades indicat per l'abreviatura TT.

UDC: Universal Decimal Classification

UKOLN: UK Office for Library & Information Networking, centre especialitzat en la gestió de la informació digital, la prestació d'assessorament i serveis per a la biblioteca, informació, educació i comunitats del patrimoni cultural, amb base a la Universitat de Bath.

UNESCO Tesaurus: United Nations Educational, Scientific and Cultural Organization subject scheme.

USDA: Departament d'Agricultura d'Estats Units (U.S. Department of Agriculture)

Vocabulari Controlat⁶⁵: Una llista dels termes que han estat enumerats explícitament. Aquesta llista està controlada per i està disponible en una

⁶⁵ Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. ANSI/NISO Z39.19-2005. ISBN:1-880124-65-3. Url:

autoritat de registre vocabulari controlat. Tots els termes d'un vocabulari controlat ha de tenir, defineix, no redundants inequívoca

web Semàntic: és un conjunt d'activitats desenvolupades pel World Wide Web Consortium, que es basen en la creació de tecnologies per publicar dades llegibles per aplicacions informàtiques. Aquesta idea promoguda per Tim Berners-Lee (1998), es basa amb la incorporació de metadades semàntiques i ontològiques (vocabularis controlats) que descriuen el contingut, el significat i la relació entre les dades publicades al World Wide Web, i evidentment, en la seva extensió es pot aplicar a qualsevol sistema d'informació que sigui publicat al web i que vulgui incorporar aquesta semàntica.

Wordmap: Una aplicació de programari de gestió de taxonomies disponibles al mercat que dona suport a la gestió de múltiples vocabularis controlats.

XML: eXtensible Markup Language

Z39.50: Una norma internacional que especifica un protocol basat en client-servidor per a la cerca i recuperació d'informació per a les bases de dades remotes.

Zthes: és un model abstracte per representar i buscar tesaurus i especificar com aquest model pot ser implementat utilitzant els protocols Z39.50 i SRW.

12. ANNEXES

Annexes Capítol 1. Evolució dels sistemes d'informació de qualitat

Annex 1.1. Llistat de *Subject Gateways* consultades

Identificació de la Passarel·la Temàtica	Activitat (2013)
ADAM: art and design. Accessible a: http://www.adam.ac.uk/adam/ → Passa a formar part de la passarel·la Intute (Arts and Humanities)	Inactiva (des de juliol 2011)
AERADE (Aerospace Resources on the Internet). Accessible a: http://aerade.cranfield.ac.uk/index.php	Activa (juliol 2013)
AHDS: arts and humanities. Accessible a: http://ahds.ac.uk/ → Passa a formar part de la passarel·la Intute (Arts and Humanities)	Inactiva
AUSTLII. Accessible a: http://www.austlii.edu.au/	Activa (agost 2013)
Biz/ed: business and economics. ACCESSIBLE A: http://www.bized.ac.uk/	Activa (juliol 2013)
BUBL (1997-2013): resources in many academic subject areas selected for library use. Accessible a: http://www.bubl.ac.uk/link/	Inactiva (darrera consulta, maig 2013)
Business information on the Internet. Accessible a: http://www.dis.ac.uk/business	Inactiva
CAIN (Conflict Archive on the INternet): information and source material on 'the Troubles' and politics in Northern Ireland from 1968 to the present. Accessible a: http://cain.ulst.ac.uk	Activa (juliol 2013)
ChemDex: chemistry (1993-). Accessible a: http://www.chemdex.org/	Activa (juliol 2013)
EdWeb: educational reform and information technology (1994-2004). Accessible a: http://www.edwebproject.org/	Inactiva (des de gener 2004)
EEL: engineering (1994-?) ⁶⁶ . Accessible a: http://www.ub2.lu.se/eel/ (URL Inactiva)	Inactiva
EEVL: engineering (199?-). Accessible a: http://www.eevl.ac.uk ; part dels seus recursos s'han traslladat al portal TechXtra que està actiu. Accessible a: http://www.techxtra.ac.uk/	Activa (juliol 2013)
ELDIS: development and the environment (199?). Accessible a: http://ntT.ids.ac.uk/eldis/	Activa (juliol 2013)
GEM (Gateway to Educational Materials): educational	Activa (juliol 2013)

⁶⁶ EELS va ser una de les primeres Subject Gateways publicades a Internet. Veure EELS - Engineering Electronic Library Sweden. Dins: Gateways: Archives & Libraries into the Next Millennium. 18th Polar Libraries Colloquy, Winnipeg, Manitoba, Canada, 12-17 June 2000. Accessible a: http://pure.ltu.se/portal/files/2037544/EELS_Winnipeg_PF.pdf

resources. (1996-). Accessible a: http://thegateway.org/	2013)
Geo-Information Gateway: geography, geology, the environment. Accessible a: http://www.geog.le.ac.uk/ct/info.html	Inactiva
History:historical studies. Accessible a: http://www.hrinfo.ac.uk/	Inactiva
Human Languages Page: language-learning and linguistics. Accessible a: http://www.june29.com/HLP/	Inactiva
HUMBUL: humanities. Accessible a: http://users.ox.ac.uk/~humbul/	Inactiva
InfoLaw * legal resources. Accessible a: http://www.infolaw.co.uk/	Inactiva
Internet Directory for Botany. Accessible a: http://www.helsind.!!/kmus/botmenu.html	Inactiva
INTUTE (2006-juliol 2011). Accessible a: www.intute.ac.uk/	Inactiva (des de juliol 2011)
The Math Forum: mathematics. ACCESSIBLE A: http://forum.swarthmore.edu/library/	Inactiva
MCS: media and communication studies. Accessible a: http://www.aber.ac.uk/~dgc/gate.html	Inactiva
NetEc: economics. Accessible a: http://netec.mcc.ac.uk/NetEc.html	Inactiva
NOVAGate: forestry, food, veterinary, and agricultural sciences http://novagate.nova-university.org/	Inactiva
OMNI: medicine and healthcare. Accessible a: http://www.omni.oc.uk	Inactiva
Philosophy Around The web: philosophy studies. Accessible a: http://users.ox.ac.uk/~worc0337/phil}index.html	Inactiva
Philosophy in Cyberspace: philosophy. Accessible a: http://www-personal.monash.edu.au/~dey/phil/	Inactiva
PICK; library/information science. Accessible a: http://www.ober.ac.uk/~tplwww/e/	Inactiva
Port: maritime studies. Accessible a: http://www.port.nmm.ac.uk/	Inactiva
Psci-com: science. Accessible a: http://omni.oc.uk/psci-com/	Inactiva
Psych web: psychology. Accessible a: http://www.psywww.com/	Inactiva
RUDI: urban design. Accessible a: http://rudl.herts.ac.uk/	Inactiva
SciCentral: science. Accessible a: http://www.scicentral.com/index.html	Inactiva
SOSIG: social sciences. Accessible a: http://www.sosig.ac.uk (1999?-2006) → A posteriori evoluciona amb el portal INTUTE (2006-juliol 2011). Accessible a: www.intute.ac.uk/	Inactiva des de juliol 2011)
TIPTOP: physics. Accessible a: http://physicsweb.org/TIPTOP/	Inactiva
VASCODA. Accessible a: www.vascoda.de → http://vascoda.wordpress.com/	Inactiva

World Wide Arts Resources: the arts. Accessible a: http://wwar.com/	Inactiva
--	----------

Annex 1.2. Taula d'iniciatives i projectes de biblioteques i *linked data* sobre dades bibliogràfiques

Tots els casos presentats són consultables a través de la W3C.

Nº	Títol Projecte i web	Descripció	Formats
1	AGRIS (1975-); <i>linked data</i> (2012-)	El catalogador normalitza la semàntica d'elements dels registres d'entrada de conjunts d'elements estàndard locals. Els indexadors utilitzen registres d'entrada per buscar un índex web de recursos relacionats, i produeix un conjunt de paraules clau relacionades pertinents sobre la base de vocabularis estàndards i descripcions d'autoritats. AGRIS es centra en l'explotació de la riquesa semàntica de les seves dades mitjançant la creació d'un conjunt de dades obert en ciències agrícoles.	XML i RDF
2	Bibliographic Network	L'usuari veu totes les expressions, manifestacions i elements relacionats amb la feina que els resulta interessant. S'utilitzen tècniques de dades relacionades que permeten que els registres bibliogràfics es publiquen com a gràfics d'informació, utilitzant estàndards web per facilitar les necessitats de descobriment dels usuaris. Treball existent: The Open Library Blog. Accessible a: API with RDF/XML output available -	FRBR (XML/RDF)
3	Community Information Service	Servei de dades enllaçades que permet seleccionar dades de fonts d'informació d'entrada per determinar quins llibres, articles, fotografies, vídeos, etc. s'han publicat recentment que poden ser d'interès per a la comunitat. Les dades amb una llicència de codi obert i ús d'estàndards oberts, permet la reutilització de les dades i redueix les barreres a la participació en la preservació i intercanvi d'informació. Exemple, <i>linked data</i> at OCLC Community. Accessible a: http://www.oclc.org/data/community.en.html	MARC XML i RDF

4	Data BNF (Bibliothèque nationale de France)	Tecnologies de dades relacionades per reunir dades de diverses fonts amb un model de dades escalable i interoperable i es millora la publicació dels recursos en el catàleg online. I els usuaris finals troben recursos sobre un tema o context més fàcilment.	SKOS, FOAF, RDA i DC (Dublin Core)
5	Use Case Identification And Deduplication Of Library Records	Est tracta d'una xarxa de biblioteques que unifica els seus registres, fent coincidir en un de sol. Els usuaris busquen dintre els catàlegs de la biblioteca, o de la web per cercar llibres, i rebre un sol registre amb enllaços a diferents còpies, en lloc de les desenes de descripcions similars d'un mateix llibre que es poden trobar en cercadors que no utilitzen la Linked Data	Marc21 i Linked Data formats (no s'especifica)
6	Use Case linked data and legacy library applications	Els usuaris, inclosos els de les aplicacions poden beneficiar-se de l'addició de les noves aplicacions ("Linked Data").	Linked Data formats (no s'especifica)
7	Use Case Open Library Data	Els usuaris que es troben referències a llibres en línia, dintre una varietat d'entorns, són capaços d'enllaçar a una font d'accés al llibre, per exemple, una versió a text complet disponible a través del "Internet Archive".	ONIX data, FRBD Data, FRBR/FRAD entities i MARC
8	Use Case Pode	L'usuari final pot buscar continguts d'un autor agrupats pel resum del treball, amb la informació extra rellevant que és afegida des d'altres fonts de dades, i es pot navegar per la col·lecció a través de la Classificació Decimal Dewey. Es converteixen dades bibliogràfiques de biblioteques a formats del <i>linked data</i> .	Bibliographic Library data (DC, etc.), FRBR, RDF, SKOS i OWL
9	Use Case Polymath Virtual Library	Els usuaris poden buscar i recuperar dades biogràfiques des de textos digitals, i altres fonts d'informació de o sobre un autor específic, i afegir comentaris, ressaltar les dades d'interès, complementar la informació, i navegar per altres autors que provenen del mateix període, etc. ús del "Linked Dada" per millorar els processos d'obtenció de enllaços des de diferents recursos i la difusió de diferents tipus de recursos,	MARC, MARC21, RDA, RDF, SKOS i OWL
10	Use Case Regional Catalog	Els usuaris poden buscar en totes les biblioteques alemanyes alhora (el <i>linked data</i> facilita aquest procés), i poden rebre	MARC, FRBR, FRAD, RDA,

		informació sobre possibles obres destacades i sobre les biblioteques que tenen disponibles les obres, i rebre informació sobre quina és la biblioteca més propera a l'usuari.	ISBD
11	Use Case Talis Prism 3	Els usuaris poden buscar llibres i altres recursos per paraula clau simple o índexs específics, refinar els resultats, i ampliar la seva recerca per la navegació de temes o autors. El servei de <i>linked data</i> ofereix cerques amb SPARQL i poderoses capacitats de cerca del text complet.	DC, MARC, MARC 21, i Linked Data formats (no s'especifica)
12	Use Case Migrating Library Legacy Data	Registres de metadades heretades de biblioteques de col·leccions, i conjunts de metadades les quals poden optimitzar la utilitat de les aplicacions d'usuari orientades a el web semàntic.	DC, MARC, RDF/XML, SKOS i OWL

Annex 1.3. Taula d'iniciatives i projectes de biblioteques i *linked data* sobre dades d'autoritats

Nº	Iniciatives	Descripció	Formats
1	AuthorClaim	El servei de registres AuthorClaim té com a objectiu vincular els autors a nivell acadèmic, amb els registres sobre les obres que han escrit. "Linked Data" es pot utilitzar per generalitzar encara més les aplicacions bàsiques, i fomentar la reutilització dels resultats de les aplicacions. L'aplicació, amb uns 13 anys d'antiguitat, contribueix a la identificació dels autors.	Linked Data formats (no s'especifica)
2	Use Case Authority Data Enrichment	"Linked Data" aplicat a dades de control d'autoritat, s'espera que es puguin reutilitzar els conjunts de dades externes mitjançant la vinculació en lloc de fer processos de còpia i fusió.	RDF, OWL, SKOS i RDA
3	FAO Authority Description Concept Scheme	Una gestió més eficient d'ús de diversos idiomes d'un concepte de l'esquema d'autoritat de la FAO. Això és possible a través de l'ús de la URI i l'assignació de les relacions entre els conceptes.	RDF i OWL (FOAF i BIBO data), SKOS, DC i RDA
4	International Registry for Authors (Iralis)	Un dels principals pilars de la recuperació d'informació científica són els noms dels autors. El problema de la inconsistència en les formes de noms d'autor existeix des de fa molts anys. Aquest sistema permet un adequat reconeixement de signatura i sense confusió, gràcies a les tecnologies de <i>linked</i>	RDF i OWL (FOAF i BIBO data) i SKOS

		<i>data</i> on es crea una URI específica per a cada autor.	
5	linked data Service of the German National Library	<i>linked data</i> proporciona un marc adequat per a la publicació de les dades rellevants a la Biblioteca Nacional d'Alemanya i la seva vinculació amb altres fonts de dades d'interès	Linked Data formats (no s'especifica)
6	Virtual International Authority File (VIAF)	L'objectiu del projecte VIAF és facilitar la recerca en tots els idiomes arreu del món, fent que sigui un recurs internacional. VIAF combina els fitxers d'autoritats de noms d'una sèrie d'institucions en un servei de noms d'una sola autoritat. L'enfocament VIAF <i>linked data</i> proporciona coneixements i experiències útils sobre com aplicar els principis de <i>linked data</i> per registres d'autoritat. Al 2009 hi havia 18 arxius personals d'autoritats de nom, de 15 organitzacions participants en VIAF. Al 2011 hi ha 21 fitxers d'autoritats de noms personals, corporatius i conferències de 18 organitzacions participants. I el 4 d'abril de 2012, VIAF ha passat de ser un prototip d'investigació d'OCLC per ser un servei d'OCLC.	VIAF, OWL, SKOS, FOAF, UMBEL, FRBR entities and elements (RDA) i DC terms

Annex 1.4. Taula d'iniciatives i projectes de biblioteques i *linked data* sobre alineament de vocabularis

Nº	Iniciatives	Descripció	Formats
1	AGROVOC Thesaurus of the Food and Agricultural Organization of the UN in Rome (FAO) (1980-)	Creació en 1980, Introducció de OWL al 2005. Esquema de conceptes multilingüe (SKOS) sobre terminologia del camp de l'agricultura, pesca, alimentació i ciències afins com el medi ambient. Aquests conceptes s'utilitzen per etiquetar i descobrir recursos, resultat de la cerca a través de diversos idiomes. Amb l'ús de <i>linked data linked data</i> , permet crear equivalències entre termes explícits d'AGROVOC i termes en vocabularis agrícoles mantinguts per altres organitzacions. Projectes existents: AGROVOC Concept Server Work Bench, AGROVOC WebServices (ambdós	OWL, SKOS, DC i DCTerms

		<p>contenen AGROVOC Concept Scheme i utilitzen OWL i SKOS-XL com a mecanisme d'indexació), EUROVOC <i>linked data linked data</i>. AGROVOC Accessible a: http://aims.fao.org/standards/agrovoc</p>	
2	<p>Browsing And Searching In Repositories With Different Thesauri</p>	<p>Cerca i Navegació de diferents catàlegs de biblioteques. La característica més important és que hi ha correlacions entre diferents tesaurus i sistemes de categorització, és a dir, hi ha intercanvi d'informació a través de tesaurus. La disponibilitat de dades obertes i enllaçades (linkables), fa que sigui més fàcil la integració d'altres sistemes de categorització o tesaurus. Un exemple de l'aplicació es pot consultar en la Biblioteca Nacional d'Alemanya <i>Schlagwortnormdatei</i> on es relacionen la Classificació "Dewey Decimal Classification" i els encapçalaments de matèries "Library of Congress Subject Headings" (Alemanya, setembre 2011). Accessible a: http://www.dnb.de/EN/Service/DigitaleDiens te/LinkedData/linkedata_node.html</p>	<p>SKOS i Dublin Core</p>
3	<p>Civil War Data 150</p>	<p>Projecte de col·laboració per compartir i connectar dades nord-americanes relacionades amb la guerra civil a través de les institucions locals, estatals i federals durant la commemoració del 6è centenari de la Guerra Civil (abril de 2011). CWD150 utilitza la tecnologia de <i>linked data linked data</i> per crear connexions basades en els identificadors i la taxonomia de la guerra civil, en particular els regiments, batalles, camps de batalla, els oficials, soldats i mariners.</p>	<p>XML, MARCXML, EAD, RDF i DC</p>
4	<p>Component Vocabularies</p>	<p>L'objectiu és permetre que les metadades siguin linkades a vocabularis establerts. Amb la tecnologia <i>linked data linked data</i> es pot aconseguir aquest objectiu, mitjançant l'assignació d'URI als termes de vocabulari. El sistema local pot</p>	<p>SKOS, MAD , RDF</p>

		emmagatzemar en memòria cau els termes de vocabulari, és a dir, mantenir una taula d'URI's i els termes corresponents, i actualitzar-la periòdicament. Els registres d'autoritat s'exposen en RDF. Descripcions bibliogràfiques, és a dir, enllaços a les descripcions RDF es relacionen amb el RDF d'autoritats per a noms, títols i matèries. Exemple Servei de <i>linked data</i> de la <i>Library of Congress</i> (Autoritats i Vocabularis). Accessible a: http://id.loc.gov/ (2009-)	
5	Language Technology	S'aplica en àrees com la traducció automàtica, el resum automàtic, recerca (web), la correcció ortogràfica, etc. Un recurs important que s'utilitza amb freqüència en la tecnologia del llenguatge és WordNet. WordNet (amb representacions RDF) és una base de dades lèxica, disponible en anglès i amb els seus homòlegs en molts altres idiomes. Un altre cas és el reconeixement de noms d'entitats (NER). Un enfocament tradicional cap NER és l'aplicació d'un diccionari geogràfic, és a dir, un diccionari amb informació sobre llocs, persones i institucions. Una de les aplicacions és el cas de Muddy, que utilitza noms d'entitats extretes de DBpedia. Accessible a: http://muddy.it/benefits	RDF
6	Subject Search	En la pràctica bibliotecària els encapçalaments de matèria s'emmagatzemen en forma de cadenes de text en els registres bibliogràfics i representen els registres a la web en HTML, on poden ser indexats pels motors de cerca a la web. Amb els principis de <i>linked data</i> s'ajuda a les biblioteques a utilitzar de forma més eficaç els processos descobriment i reutilització de recursos web. S'utilitza l'HTTP, URI i OWL per identificar i oferir millors recursos modelats	HTML, RDF, SKOS, OWL

		d'ús consolidat pels éssers humans, màquines i agents semàntics. Cas d'ús: <i>Library of Congress Vocabularies & Authorities</i> : http://id.loc.gov/ ; Primera publicació al respecte apareix al 2008 (Summers, et al.).	
7	Vocabulary Merging (Subject Interoperability)	<p>Les tècniques de mapeig i les normes que donen suport (com ara propietats SKOS mapping, <i>OWL sameAs</i> i <i>equivalentClass</i>) segueixen sent en gran part experimentals, fins i tot en el terreny de dades enllaçades. L'objectiu d'aquest tipus de sistemes és crear un punt de cerca únic en consorcis de serveis per a la descoberta de recursos que impliquin múltiples organitzacions i grans agregacions de metadades. A través de la tecnologia <i>linked data</i> es podria proporcionar una infraestructura subjacent mitjançant mapeig semàntic o la fusió dels conceptes a través de vocabularis on un usuari pogués realitzar la cerca d'un terme d'una manera inequívoca i contextualitzada. Hi ha diversos mètodes de LD descrits per a la federació de vocabularis. Projectes relacionats:</p> <p>The HILT (High-Level Thesaurus) project: provide mappings between subject terms in different schemes, developed a hub-and-spoke mapping architecture using the Dewey Decimal Classification (DDC) as the hub, implemented the architecture in RDF/SKOS, developed an API to the pilot service, demonstrated the utility of the service in several operational subject indexing and resource discovery applications, and recommended the development of a distributed approach based on <i>linked data</i> (2008).</p> <p>The MACS (enabling large-scale multilingual access to subjects) project created one-to-one mappings between terms in the English Library of Congress</p>	RDF, SKOS, OWL

		<p>Subject Headings (LCSH), French RAMEAU, and German Schlagwortnormdatei (SWD) vocabularies. The LCSH-RAMEAU mappings have been released as <i>linked data</i> to accompany the <i>linked data</i> versions of LCSH and RAMEAU, in the framework of STITCH (2009)</p> <p>ISO 25964 Thesauri and Interoperability with other Vocabularies. Part 2 will cover interoperability between thesauri and other vocabularies such as classification schemes, taxonomies and ontologies. It will provide guidance on mapping practice and architecture.</p> <p>Quite off the library framework, but long experience in vocabulary integration is provided in the biomedicine and health domain by UMLS</p> <p>Pervasive work on ontology mapping, see e.g. Jérôme Euzenat and al.</p> <p>National Diet Library's mapping of NDLSH to LCSH using SKOS.</p>	
8	<p>Bridging</p> <p>OWL and UML (Original Case Details)</p>	<p>El primer principi "Linked Data", diu: "Utilitzar URI com noms per a les coses." No obstant això, a fi d'evitar inconsistències i duplicitats, els noms han de ser manejats de manera sistemàtica i racionalitzar-ho en un model conceptual adaptable que es basa en els casos d'ús i es gestiona a través d'un metamodel de diversos llenguatges. les hi va arreglar amb sensibles llengües metamodel. En aquest cas d'ús, es poden consultar diagrames de classes UML que serveixen per explorar, reutilitzar dades basades en OWL. La comunitat UML ha desenvolupat un metamodel pera la definició d'ontologies (ODM) per ajudar a la conversió entre UML i OWL.</p>	<p>OWL, UML, ODM, RDF/XML</p>

Annex 1.5. Taula d'iniciatives i projectes de dades heterogenies

N o	Iniciatives	Descripció	Formats
1	Archipel	Disponibilitat dels vocabularis de metadades de dominis específics, expressats amb "Linked Data", aplicats en el projecte flamenc Archipel, que preten millorar la qualitat de mapeig de vocabularis.	MARC, CDWA, SPECTRUM, P/META, DC i Linked Data formats (no s'especifica)
2	Digital preservation	Ús de <i>linked data</i> per proporcionar un entorn global per a la descripció dels objectes digitals i les seves propietats amb menys duplicació d'esforços entre els proveïdors de metadades de preservació.	RDF+SPA RQL i metadades de preservació (<i>Library of Congress vocabularies</i> , OAI-ORE (structure of complex data objects), DOAP (to describe software agents), PRONOM (file formats, and encoding

			types), PREMIS
3	Europeana	Projecte Europeu (biblioteques, arxius i museus) que utilitza la tecnologia <i>linked data</i> per millorar la interoperabilitat semàntica entre diferents models de metadades, enriquint les metadades existents, millorant les dades de collita en els enllaços entre els objectes, i millorar les funcions de descobriment d'objectes i d'accés, emmagatzemament de dades en enllaços entre objectes i millorar les funcions de descoberta i accés als recursos d'informació.	Europeana Data Model (EDM), DC terms, SKOS, OAI-ORE, CIDOC-CRM, FRBR, i vocabularis de valor com VIAF, Geonames, ULAN, TGN, AAT, DDC, UDC, Iconclass, dbPedia
4	LOCAH	LOCAH (Map Based Data Visualisation), forma part del <i>Archives Hub</i> , que és un servei nacional que ofereix una gran quantitat d'informació rica interdisciplinària sobre arxius del Regne Unit. El Projecte LOCAH està investigant la creació de vincles entre el Hub i altres fonts de dades, incloent dbpedia, BBC, LCSH, i altres. <i>linked data</i> és una forma d'ampliar enormement els beneficis de la recerca, ajudant als usuaris a descobrir materials contextualment relacionats mitjançant la creació de vincles entre les col·leccions d'arxius disponibles i altres fonts que sovint estan molt disperses a la web.	DC Terms, Bibliontology (BIBO), FOAF, SKOS, RDFS i FRBR
5	Photo museum	Projecte finlandès, que estudia l'ús del <i>linked data</i> per fer front als problemes tècnics causats per la heterogeneïtat de dades i recursos d'informació disponibles a la xarxa. Les col·leccions fotogràfiques, molt populars a Internet, s'expressen en diverses formes i formats i están	ISAD(G), I RDF entre d'altres (no especificats)

		disponibles també en diverses bases de dades que descriuen de forma diferent els seus continguts (descripcions físiques dels materials i conservació, el contingut, els agents, els contractes i els drets de propietat intel·lectual, etc.)	
6	Radio station archive digitization	Projecte centrat en arxius d'emissores de radio (molt extensos, antics i no digitalitzats) i amb metadades molt poc consistents i fetes a mida. Amb l'ús del Linked Data, l'objectiu és crear referències de diferents esdeveniments d'àudios, per permetre la recerca federada sobre aquests arxius, i en general, afegir valor al procés de digitalització així com augmentar impacte en els potencials resultants (registres)	FRBR and flat Dublin Core i RDF entre d'altres
7	Recollection	Projecte del programa de preservació de la <i>Library of Congress</i> . Es tracta d'una iniciativa per desenvolupar una estratègia nacional per recopilar, arxivar i preservar les quantitats creixents de continguts digitals dipositades en diversos sistemes i institucions. Es vol utilitzar el Linked Data, per crear una plataforma bàsica per a bibliotecaris i conservadors de col·leccions web (Recollection), i com a font de dades per augmentar aquestes col·leccions. Els usuaris potencials de la informació poden descobrir més fàcilment i analitzar aquestes dades en una varietat de noves formes.	XML MODS, DDC, LCSH, people data (e.g. VIAF), entre d'altres formats i estàndards Linked Data o LOD (LINKED Open Data)
8	Ontology of Cantabria's cultural heritage	Projecte de la Universitat de Cantabria, per a la publicació de dades sobre patrimoni de Cantabria. Ús del "Linked Data" per contribuir en aconseguir una major quantitat i qualitat d'informació a la xarxa sobre el patrimoni cultural de Cantabria	CIDOC CRM + FRBRoo, MARC 21, EAD, DC and METS.

Annex 1.6 Citacions

Nº	Cas d'ús	Descripció	Formats
1	Citation of scientific datasets	Ús del "Linked Data" per poder proporcionar una forma estàndard per connectar els conjunts de dades científiques provinents de les publicacions científiques que els suporten. Les dades relacionades poden proporcionar orientació sobre com assignar identificadors, com a enllaç de dades, descripcions, publicacions, autors, etc. a través de vocabularis controlats.	OPM, BIBO, CITO
2	Enhanced publications	Les biblioteques i altres dipòsits emmagatzemen articles d'investigació en revistes, tesis, informes, memoràndums, capítols de llibres, etc. L'aplicació de "Linked Data" podria ajudar a crear una col·lecció de material de recerca que està junt i fer possible que els processos de recerca siguin més comparables i transparents.	Dublin Core Terms, Dublin Core Types (Dataset, Image, Sound, ...), FOAF, ORE, etc.
3	Mapping scholarly debate	A través del projecte <i>Bibliographica</i> , es vol capturar les relacions entre obres i autors, amb dades molt més riques que les que es troben normalment disponibles en les biblioteques. L'objectiu final és capturar l'evolució del pensament i el debat acadèmic en relació a les obres escrites i altres publicacions i els seus autors. Aquesta evolució i debat podrien ser representats com un graf dirigit i publicat sota els principis de Linked Data.	Linked Data formats (no s'especifica)

Annex 1.7. Objectes Digitals

Nº	Cas d'ús	Descripció	Formats
1	Collecting material related to	LUCERO (Linking University Content for Education and Research Online) Project de l'Open University està investigant en l'ús	RDF, SPARQL, SKOS,

	courses at the Open University	de les tecnologies "Linked Data" i treballant amb un prototipus sobre la vinculació i accés a les dades d'interès per als estudiants i investigadors, per tal de que un estudiant/investigador pugui accedir a tots els materials dels cursos de la Universitat (llibres, DVD, CD, programes de televisió, recursos educatius oberts, etc).	Bibliotolog y (BIBO), MARC, DC Terms, etc.
2	Digital text repository	Ús del <i>linked data</i> en repositoris digitals per tal de que es puguin aprofitar la quantitat de metadades de les que disposen i alhora d'altres proveïdors de dades, per tal de facilitar als usuaris finals les tasques de descoberta, selecció i vinculació semantic de la informació. Existeixen diversos projectes al respecte: <i>Gutenberg Project</i> , <i>Wikisource</i> , i el <i>Ben-Yehuda Project</i> .	MARC 21, Dublin Core terms, FOAF (RDF, OWL, SKOS)
3	Enhanced publications	Citat anteriorment en l'Annex 6 (cas 2). S'engloba també en aquesta secció, perquè el projecte fa referència a objectes digitals.	Dublin Core Terms, Dublin Core Types (Dataset, Image, Sound, ...), FOAF, ORE, etc.
4	Editing reports on new academic documents	<i>Editing Reports on New Academic Documents</i> (Ernad), software que pretén ajudar als editors de publicacions acadèmiques d'una matèria en concret, a decidir si els documents s'ajusten a la categoria específica definida. Amb la tecnologia "Linked Data" es pot ajudar al desenvolupament d'un sistema de classificació de matèries d'una forma molt dinàmica, i proporcionar una base per a la representació normalitzada del coneixement.	Linked Data formats (no s'especifica)
5	NDNP (National	Es tracta d'una aplicació web, feta amb la col·laboració de la <i>National Endowment for</i>	RDF/XML, OAI-ORE,

	Digital Newspaper Program	<i>the Humanities</i> (NEH), i la <i>Library of Congress</i> (LC), per tal de proporcionar un millor accés als diaris dels Estats Units publicats entre 1836 i 1922. Aquest sistema permet als usuaris buscar i veure més de 2,5 milions de pàgines digitalitzades. Els principis "Linked Data" han proporcionat la base per al disseny de l'espai identificador de la <i>Chronicling America</i> .	DC Terms, Bibliographic Ontology, EXIF FRBR
6	NLL digitized map archive	La versió digital de la Col·lecció Cartogràfica de la Biblioteca Nacional de Letònia (LLN), conté col·leccions digitalitzades de mapes històrics (S.XVI al S.XVIII). Les metadades dels mapes estan disponibles com a dades relacionades (Linked Data) amb RDFa o utilitzant HTTP 303, que apunten a diverses representacions dels recursos i les seues metadades. A més a més, els mapes estan vinculats a altres recursos relacionats (per exemple, noms de llocs o dades d'autoritat sobre les persones).	RDF, RDFa, SKOS (for place name, person thesaurus), W3C Geo vocabulary, Dublin Core
7	Publishing 20th Century press archives	Els Arxius del segle 20 de premsa de la Biblioteca Nacional Alemanya d'Economia (ZBW) és una gran col·lecció de retalls de diaris sobre persones, empreses, matèries i productes, que s'estén des de 1826 fins a 2005, organitzat per carpetes temàtiques. Prové de metadades de OAI-ORE, descrites també en RDFa per tal d'enriquir amb enllaços a altres tipus de dades.	ORE, DC Terms, SKOS, rdaGr2 (profession or Occupation), exif (resolution), p20vocab

Annex 1.8. Projectes d'agrupacions de col·leccions

Nº	Cas d'ús	Descripció	Formats
1	AuthorClaim	Citat anteriorment en l'Annex 3 (cas 1). S'engloba també en aquesta secció, perquè el projecte fa referència a agrupacions de col·leccions a través de les quals es vincula	Linked Data formats (no s'especifica)

		l'autor amb l'obra existent en la col·lecció.	
2	Collection-level description	Fa referència a la descripció d'una col·lecció amb metadades FRBR (àmbit arxius, biblioteques i museus). La definició i publicació d'aquestes descripcions de les col·leccions com a <i>linked data</i> , procedents de múltiples i heterogènies fonts, permet la seva vinculació amb altres conjunts de dades (geogràfiques, universitats, matèries primeres, etc.), proporcionant un conjunt divers de serveis que es poden personalitzar a través de proveïdors de servei i usuaris finals.	RDA, FRBR, FRAD, Dublin Core vocabularies, MARC21, UNIMARC, ISBD
3	Community information service	Organitzacions acadèmiques de diferents mides (grups de recerca, departaments universitaris, societats acadèmiques, grups d'interès, etc) amb un fort interès en el manteniment de la consciència i la qualitat de la informació en el seu domini, opten per publicar en obert aquesta informació per a la seva comunitat acadèmica i al públic en general. Amb el <i>linked data</i> es podria combinar amb una llicència oberta que permeti la reutilització de la informació per a tals fins, i donar suport a les API, estàndards de dades i programari client per reduir la barrera a la participació en l'intercanvi i gestió/custòdia (<i>curation</i>) de la informació. la curatela de la informació i el compartir. Alguns exemples són els programaris de col·leccions bibliogràfiques com <i>BibSonomy</i> , <i>BibServer</i> , <i>BibApp</i> , <i>Acadèmic Obre</i> . Tots aquests sistemes es beneficien de la fàcil disponibilitat de les dades enllaçades publicades per biblioteques a través d'una API. I els serveis de custòdia de dades d'una comunitat com VIVO (http://www.vivoweb.org/) o BKNPeople (http://people.bibkn.org/) o BKN Project: (http://www.bibkn.org/).	Linked Data formats (no s'especifica)
4	Digital resources with access restrictions	Ús de tecnologies <i>linked data</i> per donar suport l'accés a recursos digitals centrats amb l'usuari, en lloc de centrar-ho amb institucions, i determinar si l'usuari té accés a un element o no, donades les seves subscripcions personals i privilegis de la biblioteca. Un exemple és D-FOAF, sistema distribuït de gestió d'identitats que s'aplica en xarxes socials. A través de la informació inherent en les xarxes socials es pot utilitzar,	

		analitza algorismes per a la gestió d'identitats, autoritzacions i comprova drets d'accés distribuïts.	
5	Library address data	Ús del <i>linked data</i> amb dades sobre adreces que poden ser utilitzades per enllaçar dades amb diferents escenaris possibles. Per exemple, <i>The Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen</i> (HBZ) prové d'un conjunt de dades amb identificadors de biblioteques (URI). aquestes dades es vinculen amb fons de dades bibliogràfiques o aplicacions mòbils basades en geolocalització que proporcionen hores d'obertura i informació sobre l'edifici o el personal de la biblioteca, etc. Accessible a: http://www.hbz-nrw.de/	DC Terms, FOAF, geo data, hcard, vcard, RDFa, RDFS
6	Nearest physical collection	Ús de tecnologies <i>linked data</i> per proporcionar d'una forma estàndard i més estructurada, una navegació a través d'un catàleg col·lectiu que permeti trobar registres específics i la seva geolocalització. Un cas d'exemple és el servei d'informació Scotland's Information service , el qual combina el nivell de descripció (metadades) de la col·lecció de la Scottish Collections Network (SCONE) service amb el servei Google Maps service , utilitzant l'aplicació Google Maps API . A més a més, <i>Scotland's Information</i> ofereix el servei Nearest library service , que permet mostrar totes les ubicacions de les biblioteques d'Escòcia. I un usuari final pot allunyar o apropar el mapa a qualsevol lloc i identificar visualment quines biblioteques té més properes.	DC Terms i Linked Data formats (no s'especifica)

Annex 1.9. Usos Socials

Nº	Cas d'ús	Descripció	Formats
1	Community information service	Citat en l'Annex 8 (cas 3). En aquest cas es fa referència a l'ús de dades enllaçades (LD) en serveis d'informació socials (comunitats socials) que proporcionen dades en llicència oberta. Mitjançant l'ús d'estàndards de la WS es permet la reutilització d'informació i la reducció de barreres en la gestió, preservació i intercanvi d'informació. Per exemple, Bibsonomy: sistema social per compartir "etiquetes" i llistes de publicacions	Linked Data formats (no s'especifica)

		<p>(bibliografia) a través del web semàntic. O bé el projecte <i>Bibliographic Knowledge Network</i> (BKN), creat al 2008 i finançat per la <i>National Science Foundation</i>, per tal de desenvolupar eines i serveis per estimular la formació d'organitzacions virtuals en les comunitats científiques (grups de conferències, grups de recerca, etc.) i poder filtrar els documents pertinents de diversos fluxos d'entrada, seleccionar i millorar la qualitat de les dades bibliogràfiques relacionats amb l'organització, etc. Un altre exemple és l'Open Bibliography (JISC, 2010-), sistema d'informació mundial que exposa informació bibliogràfica com a "Linked Open Data". Utilitza els estàndards i eines existents del web semàntic (RDF, SPARQL), patrons de dades enllaçades i ontologies obertes acceptades (FoaF, Bibo, DC, FRBR, etc). Les dades han de sortir dels catàlegs de biblioteques i dels sumaris de revistes d'accés obert, el projecte ha comptat amb la col·laboració dels catàlegs de les biblioteques de la <i>Cambridge University Library</i> i de la <i>British Library</i> i de l'editor científic IUCr (<i>open-access peer-reviewed journal from the International Union of Crystallography</i>)</p>	
2	Crosslinking environment data and the library	<p>Casos d'ús que fan referència a entrecreuar informació bibliogràfica amb altres tipus de dades relacionades. Per exemple, l'Agència Federal de Medi Ambient d'Alemanya (UBA) té una llarga tradició d'organització del coneixement que combina recursos de la biblioteca amb molts sistemes d'informació basats en la web que presenten dades d'observació i d'anàlisis de resultats. UBA té l'objectiu d'entrecreuar informació bibliogràfica amb dades d'observació relacionades amb el medi ambient, i vincular cada tipus d'informació amb el seu vocabulari de referència. Amb les dades vinculats, i el seu sistema de classificació, reforçat amb un vocabulari de referència (tesaurus, nomenclàtor geogràfic i un diccionari cronològic) les dades de la biblioteca poden estar interrelacionades.</p>	<p>Dublin Core, Darwin Core, Statistical Core Vocabulary, RDF, SKOS, RDF, etc. (Geonames Ontology, Event Ontology, etc.)</p>
3	Crowdsourced catalog	<p>A través de <i>linked data</i> es permeten que usuaris voluntaris d'un sistema d'informació/comunitat social millorin les</p>	<p>MARC, CSV, XML i Linked</p>

		dades bibliogràfiques, així com crear nous registres bibliogràfics. Exemples: Biblios.net ; Flickr Commons ; Librarything ; Open Library o Galaxy Zoo (comunitat astronoma que permet que els seus usuaris classifiquin galxies)	Data formats (no s'especifica)
4	Mendeley research networks for linking researchers and publications	En el procés de publicació científica, una publicació impresa està separada del seu context (dades i enllaços que estan en l'entorn informacional de l'investigador. Mendeley proposa afegir un context posterior a les publicacions a través de les metadades socials així com els enfocaments algorítmics per la vinculació de documents, gràcies a <i>linked data</i> i l'anàlisi gràfica d'enllaços (per exemple, a través dels lectors, co-citacions, etc.)	Linked Data formats (no s'especifica), entre d'altres formats de tipus bibliogràfic
5	Open Library data (Home: http://openlibrary.org/)	És una base de dades bibliogràfica de grans dimensions (al voltant de 25 milions d'articles), amb metadades de llibres electrònics. Conté més d'un milió de llibres electrònics provinents de dades enllaçades a bases de dades bibliogràfiques. L'objectiu és que l'usuari pugui vincular, de forma directa, les dades d'una biblioteca oberta, i enllaçar-les amb un treball específic que conté informació sobre manifestacions del mateix treball que són accessibles a text complet. Vocabularis associats:	RDF, FOAF, FRBR, FRBR/FRA D, RDVOCAB i DC Terms
6	Ranking search results by popularity using circulation data	Millorar el rànquing dels materials del catàleg de la biblioteca amb els resultats de cerca. <i>linked data</i> pot ajudar a la millora els cercadors, per exemple, per realitzar cerca per rànquings de popularitat. Amb les dades enllaçades, es poden vincular les dades de la circulació de documents de les biblioteques (recollides, préstecs, etc.) amb les dades bibliogràfiques per determinar quin és l'element més popular d'una biblioteca. Les dades poden ser consultades amb SPARQL.	RDF, MARC, RDA, FR family, entre d'altres formats bibliogràfics
7	SEO	Aplicació del <i>linked data</i> amb dades estructurades procedents de catàlegs de biblioteques per tal de que aquestes dades siguin fàcilment cercades i puguin ser millor indexades i puntuades (rankings) pels motors de cerca web.	RDF, RDFa, entre d'altres (no especifica)
8	Support this	És un mecanisme similar al botó "Like" de Facebook, però "support this book button"	RDFa/BIBO

	book button	connecta les pàgines dels catàlegs en línia de biblioteques i llocs web relacionats amb les biblioteques del catàleg Gluejar, on aquests "vots" indiquen el suport a l'alliberament d'una versió amb llicència CC d'un llibre o llibre electrònic. Quan un usuari fa clic a un botó d "ajuda", el sistema recollirà les metadades de la pàgina web del seu interès. Amb el <i>linked data</i> , el sistema agafaria les metadades i les utilitzaria per vincular a una obra o llistat d'obres del catàleg Gluejar. Es proposa l'ús dels formats RDFa/BIBO + DC	, DC
9	Social annotation	Cada cop més la gent crea anotacions dels recursos que estan llegint, com llibres o articles d'investigació, etc. I en aquest cas d'estudi es planteja l'ús del <i>linked data</i> per permetre als usuaris descriure i publicar metadades d'anotació perquè pugui ser reutilitzades per altres.	RDF, RDFa i FRBR
10	Peer to Peer bookswapping	Són limitades les oportunitats i opcions actuals per compartir legalment ebooks restringits. I en aquest cas d'ús es planteja l'exploració dels límits de les opcions de préstecs de llibres electrònics existents a través del <i>linked data</i> . Les tecnologies de dades enllaçades permetrien l'ús de llicències CC (Creative Commons) per lectura per mecànica, permeten l'agregació de les dades de fonts distribuïdes com les xarxes socials o pàgines web individuals. Es planteja el <i>linked data</i> com a facilitador de creació de serveis innovadors de préstec.	Linked Data formats (no s'especifica)
11	Publish data for lightweight application development	Actualment, els recursos de dades i APIs són rarament exposats pels serveis de biblioteca, i poden requerir permisos especials. Les tecnologies i principis del <i>linked data</i> poden ajudar a exposar les dades que s'utilitzaran en aplicacions de tercers. Després, els usuaris poden proporcionar les seves pròpies aplicacions, que poden ser compartides i difoses sense por de la desaparició de dades o de problemes de drets. Treballs existents: JISC's Open Biblio project i SHERPA RoMEO	Linked Data formats (no s'especifica)
12	Social recommendations	Us del <i>linked data</i> per connectar les dades socials (xarxes socials, etc.) sobre classificacions i recomanacions de recursos en l'aplicació de cerca que proporcionen els	RDF entre d'altres formats

	catàlegs online de biblioteques. En general, els catàlegs no ofereixen el rànquing de resultats de cerca per rellevància per a l'usuari, així doncs a través de les dades enllaçades, els catàlegs de biblioteques podrien proporcionar als usuaris recomanacions o rànquings de cerca basats en la informació disponible per popularitat i/o per l'activitat de l'usuari.	web i/o formats del Linked Data (no especifica)
--	--	---

Annexes Capítol 2. Comparativa de Models de Sistemes d'Informació distribuïts de qualitat

Annex 2.1. Atributs associats a la informació de qualitat

1. **Actualització:** data/edat i actualitat del contingut/servei, etc.
2. **Accessibilitat (web)**⁶⁷: capacitat d'una pàgina web o aplicació per facilitar als usuaris (independentment dels seus nivells de discapacitat física o tecnològica) l'accés a la mateixa o als mateixos continguts.
3. **Accessibilitat (disponibilitat):** paràmetre referent a la col·lecció, és a dir, nivell de representació i estructura de les dades (per exemple, ús de metadades estàndards de descripció), drets d'accés a la informació, etc.
4. **Autoria:** Si la font atribuïble a un autor o una organització de bona reputació
5. **Completesa:** Quocient del nombre d'elements de resposta i el nombre d'elements del món real. També: la cobertura, abast, granularitat, la integralitat, la densitat i extensió. A més a més, s'inclou la inclusió de metadades per a la descripció de la informació.
6. **Comprensibilitat:** Grau en que la informació pugui ser compresa per l'usuari. Nivell de profunditat i de detall de la informació, si es cobreixen tots els aspectes del tema, si la informació és incompleta, etc.
7. **Contingut:** es refereix al contingut web, textual, visual, auditiu, provinent de diferents unitats d'informació, i que es pot abordar a través del web. Això pot ser contingut no estructurat, com articles, o contingut estructurat, com els registres de base de dades.
8. **Disponibilitat:** Percentatge de temps que una font d'informació està al sistema, s'inclou també l'accessibilitat, la fiabilitat, la recuperabilitat i la representabilitat
9. **Disponibilitat:** Percentatge de temps que una font d'informació està al sistema, s'inclou també l'accessibilitat, la fiabilitat, la recuperabilitat i la representabilitat
10. **Estructura/Comprensibilitat:** nivell de profunditat i de detall de la informació, si es cobreixen tots els aspectes del tema, si la informació és incompleta, etc.
11. **Estructura/Consistència:** fa referència al nivell de representació consistent en què l'estructura de la informació s'ajusta a la d'altres fonts d'informació, per exemple nivell d'integritat, homogeneïtat, consistència semàntica, compatibilitat, etc. (perspectiva de la recuperació de la informació)
12. **Estructura de referència:** Una estructura de referència és un conjunt jeràrquicament organitzat d'elements de contingut.
13. **Exhaustivitat:** a nivell de la recuperació de la informació es considera la probabilitat de que tots els articles pertinents siguin recuperats. Per tant, és una variable dependent de la precisió.
14. **Integritat:** si la informació és actual i si està actualitzada, la durabilitat de la informació (si pot perdura en el temps o no, etc.) i si existeix un adequat manteniment dels continguts.

⁶⁷ Serrano Mascaraque, E., 2009. Accesibilidad vs usabilidad web: evaluación y correlación. Investigaciónn bibliotecológica 23, 61-103.

15. **Integritat:** si la informació és actual i si està actualitzada, la durabilitat de la informació (si pot perdura en el temps o no, etc.) i si existeix un adequat manteniment dels continguts.
16. **Precisió:** quocient entre el nombre de valors correctes proporcionats pel sistema d'informació i el nombre total de valors que conté el sistema.
17. **RDF⁶⁸:** Resource Description Framework, és un estàndard del W3C que ha estat desenvolupat per al modelatge del contingut web, especialment per millorar la recuperació del contingut web. Originalment RDF s'entén com un model de metadades, però avui dia s'està aplicant a formes més àmplies de modelatge. RDF juga un paper clau en la web semàntica, també coneguda com a web 3.0.
18. **Rellevància:** Grau en què la informació compleix les necessitats dels usuaris. També s'inclou la precisió de domini, la redundància mínima, aplicabilitat i utilitat. Per alguns autors amb la rellevància inclouen el concepte de pertinença, la mesura de com un document s'ajusta a una necessitat informativa, també anomenada rellevància subjectiva, en contraposició de com una pregunta s'ajusta a un document (rellevància objectiva)⁶⁹.
19. **Seguretat:** Grau en què la informació es transmet privada des de l'usuari a la font d'informació. També s'inclou: la privacitat i la seguretat d'accés.
20. **SKOS⁷⁰:** Simple Knowledge Organization System, és un format estàndard del W3C basat en RDF, que serveix per a la representació dels sistemes d'organització del coneixement (vocabularis controlats) en la web semàntica. Es pot dir, que és una versió moderna i orientada al web de la ISO 1986 2788⁷¹ per a desenvolupament de tesaurs estàndards.
21. **Xarxa de referència:** Una xarxa de referència és una xarxa que sorgeix quan els elements de contingut estan connectats no només a la seva pròpia estructura de referència, sinó que també estan vinculats directament a altres elements de contingut.
22. **XML⁷²:** Extensible Markup Language, llenguatge estàndard per estructurar el contingut web

Annex 2.2. Definició d'escapes de valors proposades per a un model IQ

Annex 2.2.1. Escala d'avaluació sumatòria

Valor	Definició
1 = baix	Menys de la meitat de valors no es contempen (< 40%)
3 = mig	La meitat dels valors es contempen (40-70%)

⁶⁸ RDF: Url: www.w3c.org/rdf/

⁶⁹ Gómez Díaz, R., 2003. La evaluación en recuperación de la información Hipertext.net. Url: http://www.upf.edu/hipertextnet/numero-1/evaluacion_ri.html

⁷⁰ SKOS. Url: www.w3c.org/2004/02/skos/

⁷¹ ISO 2788 Guidelines for monolingual thesauri. Url: http://www.iso.org/iso/catalogue_tc/catalogue_detail.htm?csnumber=7776

⁷² XML. Url: www.w3c.org/xml

5 = alt/molt alt	Més de la meitat dels valors es contempen (> 70%)
------------------	---

*Els sistemes d'informació de qualitat, ja en si tenen unes premisses que impliquen nivells alts de qualitat, i per tant, no s'ha contemplat el valor nul en aquesta escala de valors

Annex 2.2.2. Escala d'avaluació per distància (absència/presència d'elements)

Valor*/Puntuació	Definició
No = 0	Absència de l'atribut: considerat un valor negatiu respecte la IQ
Si = 1	Presència de l'atribut: considerat un valor positiu respecte la IQ

*Excepció: Quan el valor d'un paràmetre correspongui a una resposta negativa, independentment del valor (Si/No), la puntuació de l'element serà 0 i a l'inversa, si el valor és positiu independentment del valor de la resposta (Si/No), la puntuació serà 0.

Annex 2.3. Qüestionari d'avaluació

Atributs ^{73,74}	Escala de valor d'IQ		
	Avaluació Qualitativa/Quantitativa	Puntuació	Total
1. Dimensió Sistema (Interrogació)	Descripció/Qüestions⁷⁵	[Si=1; No=0]; [Nivell d'IQ=1,3,5]	
Autoria (1,2,3)	Es coneix l'autoria i l'origen del recurs? Si/No		
	Apareixen dades de contacte (adreça, telèfon, etc.)? Si/No		
	Hi ha bústia de suggeriments, comentaris, etc.? Si/No		

⁷³ Naumann, F., Rolker, C., 2000. Assessment methods for information quality criteria, IQ. Citeseer, pp. 148-162.

⁷⁴ 1999. A list of quality selection criteria: a reference tool for Internet subject gateways, Selection criteria for quality controlled information gateways UKOLN Metadata Group. Url: <http://www.ukoln.ac.uk/metadate/desire/quality/report-2.html>

⁷⁵ Nota: algunes respostes poden ser respostes amb si/no i d'altres amb la descriptiva qualitativa (DQ). Les preguntes específiques són adaptades del model DESIRE.

<p>Comprensibilitat (1)</p>	<p>Grau en que la informació és compresa per l'usuari (per exemple, multilingüisme): alt</p> <ul style="list-style-type: none"> - La informació és completa (de la informació recuperada)? Si - El títol és informatiu? Si - Es dona un resum? Si - Altres? 			
<p>Estructura / Consistència en la representació de la informació (1)</p>	<p>Grau de representació consistent en què l'estructura de la informació s'ajusta a la d'altres fonts, per exemple nivell d'integritat, homogeneïtat, consistència semàntica, compatibilitat, etc. (DQ)</p> <ul style="list-style-type: none"> - Visualització immediata dels resultats sense passos intermedis? Si/No - Qualitat de la referència bibliogràfica (ús d'estàndards)? Si/No - Possibilitat de selecció del criteri d'ordenació dels resultats? Si/No - Possibilitat de selecció dels registres de sortida (impressió, gravació o enviament via e-mail)? Si/No - Possibilitat de limitar o augmentar el nombre de resultats de la cerca en cada pantalla? Si/No - Dona com a resultat els termes de cerca? - Opcions extra de visualització dels resultats (per exemple: agrupacions o índex KWIC)? Si/No 			

Integritat (1): integritat de la informació, del lloc web i del sistema	- La informació proporcionada pel sistema és actual i està actualitzada? Si/No			
	- Existeix un manteniment adequat dels continguts informatius? Si/No			
	S'amplia informació sobre el recurs (per exemple, recursos/enllaços relacionats, biblioteca o base de dades dipositària, etc.)? Si/No			
	- Hi ha enllaços morts? Si/No			
	- Hi ha enllaços a llocs on s'ha mogut o hi ha redireccionament a un nou lloc web? Si/No			
	- Les pàgines web del sistema tenen la data amb la última data de revisió? Si/No			
	- Apareix el número de versió dels recursos? Si/No			
	- Apareix la data de la darrera actualització del recurs? Si/No			
	- Hi ha una descripció de les freqüències de l'actualització dels recursos? Si/No			
	- L'organització/persona que allotja el recurs sembla tenir el compromís amb el manteniment i l'estabilitat dels recursos? Si/No			
- El lloc s'actualitza/manté amb freqüència / manteniment? Si/No				
-Fiabilitat del sistema (DQ)				
Disponibilitat/A ccessibilitat (1,2,3)	-Ofereix serveis per facilitar l'accés a informació disponible en diferents formats, suports, etc.? Si/No			

[nivell d'accessibilitat i recuperabilitat]	<p>-Té serveis que faciliten la conversió/extracció de la informació amb formats diferents? Si/No</p> <p>-Ofereix serveis per a que un usuari pugui accedir a la informació de forma indirecta? (serveis d'alerta, butlletins, guies temàtiques, etc.) Si/No</p> <p>- Informació sobre existència de limitacions o condicions d'accés (per exemple, data límit d'accés, si es requereix quota, etc,?Si/No</p> <p>- Ofereixen servei d'Obtenció de Documents?Si/No</p> <p>- Existència de missatges clars quan la connexió no es possible?</p>			
Seguretat (1) [Privacitat i seguretat d'accés]	<p>Grau en què la informació es transmet de forma privada a l'usuari. (DQ)</p> <p>Hi ha polítiques de privacitat, protocols d'intercanvi d'informació, etc.)? Si/No</p>			
Precisió (1,2, 3)	<p>-Veure si la informació recuperada en la interrogació (RI) és o no precisa?</p> <p>I si la presentació de la informació és adequada en termes de precisió.</p> <p>-La informació es pot verificar amb exactitud?Si/No</p> <p>-El recurs, mostra cites bibliogràfiques o referències per confirmar la informació? Si/No</p> <p>-La presentació de la informació en la RI és correcta (sense faltes</p>			

	<p>ortogràfiques/gramaticals)?Si/No</p> <p>-Hi ha prevalença d'errors ortogràfics? Si/No</p>			
Rellevància (1,2,3)	<p>Grau en què la informació recuperada compleix les necessitats dels usuaris.</p> <p>- El document rellevant respon a la pregunta/interrogació realitzada al sistema (rellevància objectiva)?Si/No</p> <p>- El document recuperat s'ajusta a la necessitat informativa de l'usuari? (rellevància subjectiva o pertinència). Si/No</p> <p>-Precisió del domini, aplicabilitat, utilitat, etc.?Si/No</p> <p>- Hi ha accés de navegació per matèries (tesaurus, etc.)?Si/No</p> <p>- Hi ha possibilitat de cercar termes genèrics i específics (navegabilitat per estructura jeràrquica)?Si/No</p> <p>- Fiabilitat de la cerca per classificació/descriptors de termes específics?</p> <p>- Consistència dels termes d'indexació (un concepte/tema ha d'aparèixer d'una forma única)?</p> <p>- Suggestions de cerca per termes/conceptes alternatius?</p> <p>- Capacitat per discriminació de termes?Si/No</p>			
2. Dimensió Serveis (Tipologia)	Descripció/Qüestions	[Si=1; No=0];		
		[Nivell d'IQ=1,3,5]		

<p>Actualització (2,3)</p>	<ul style="list-style-type: none"> - Serveis actualitzats? Si/No - Apareix data d'actualització? Si/No - Informació sobre noves versions/serveis/aplicacions? Si/No - Tutorials actualitzats? Si/No - etc. 			
<p>Autoria (2,3)</p> <p>[tota la informació disponible sobre l'autor dels recursos d'informació i la seva reputació]</p>	<ul style="list-style-type: none"> - És la font atribuïble a un autor o una organització de bona reputació? Si/No - Es coneix l'origen del recurs? Si/No - Apareix el nom complet de l'autor? Si/No - Es coneix l'autor/institució o editor del recurs? Si/No - Hi ha una URL a un servidor conegut (universitat, centre de recerca, etc.) Si/No - L'autor ja estat citat abans? Si/No - El recurs proporciona altres cites/referències de l'autor? Si/No - Hi ha disponible informació sobre l'autor (CV., adreça, e-mail, etc.)? Si/No 			
<p>Completesa (2,3), [Cobertura, Abast, Granularitat, Densitat i Extensió], aquí s'inclou també</p>	<ul style="list-style-type: none"> - Quocient del nombre de serveis inclosos i el nombre de serveis bàsics existents (cerques, alertes, butlletins, etc.). - Nivell de descripció dels recursos inclosos en els serveis i dels propis serveis? - És a dir, quina és la informació bàsica que ha de contenir el 			

l'ús de Metadades	servei? -Nivell d'especificitat, els serveis s'adeqüen a l'audiència, etc.?			
Disponibilitat / Accessibilitat (1,2,3) [Accessibilitat, recuperabilitat, etc.]	El lloc conté informació que no està fàcilment disponible: es difícil trobar trobar dades similars? Està disponible a través d'altres fonts d'informació? El recurs de facilitar l'accés a informació que no s'obté fàcilment a altres formats / no va dur a terme a moltes biblioteques? El recurs de proporcionar accés a la informació que els usuaris poden no ser capaços de trobar / accedir a les biblioteques?			
Rellevància (1, 2, 3) [precisió, redundància, aplicabilitat i utilitat]	Grau en què la informació recuperada compleix les necessitats dels usuaris. Quant a serveis es refereix, s'aplica a la perspectiva de l'usuari, sobretot referent al serveis d'informació que puguin proporcionar els sistemes (serveis d'alerta, serveis de cerca, butlletins, guies temàtiques, etc.) Aquest atribut potser difícil d'avaluar completament en alguns dels sistemes perquè alguns serveis són privats o de pagament. (DQ)			
Precisió (1,2, 3)	Aquest atribut és limita als serveis d'informació/documentació (obtenció de documents, serveis d'alertes, butlletins de notícies, guies temàtiques, etc.), cerca i navegació. Es poden incloure les mateixes interrogacions que en la			

	<p>dimensió del sistema:</p> <ul style="list-style-type: none"> - Informació obtinguda a través del servei és precisa i adequada?Si/No - La informació es pot verificar amb exactitud?Si/No - Correctesa de la informació rebuda (sense faltes ortogràfiques, gramaticals o tipogràfiques), etc.? 			
3. Dimensió Col·lecció	Descripció/Qüestions	[Si=1; No=0];		
		[Nivell d'IQ=1,3,5]		
Autoria (2,3)	<ul style="list-style-type: none"> - És la font/recurs atribuïble a un autor o una organització de bona reputació? - És la font atribuïble a un autor o una organització de bona reputació?Si/No - Es coneix l'origen del recurs? Si/No - Apareix el nom complet de l'autor?Si/No - Es coneix l'autor/institució o editor del recurs? Si/No - Hi ha una URL a un servidor conegut (universitat, centre de recerca, etc.) Si/No - L'autor ja estat citat abans?Si/No <p>El recurs proporciona altres cites/referències de l'autor? Si/No</p> <p>-Hi ha disponible informació sobre l'autor (CV., adreça, e-mail, etc.)? Si/No</p>			

<p>Actualització (2,3)</p>	<ul style="list-style-type: none"> - Nivell i/o polítiques d'actualització de la col·lecció. (DQ) - Hi ha una declaració sobre la freqüència d'actualització? Si/No - Si es tracta d'un recurs estàtic (no actualitzat) és la informació útil per a l'usuari/audiència? Si/No - Com sensible al temps és la informació continguda en el sistema, i com es relaciona amb la freqüència d'actualització? (DQ) 			
<p>Completesa (2,3), [Cobertura, Abast, Granularitat, Densitat i Extensió], aquí s'inclou també l'ús de Metadades</p>	<p>Quocient del nombre d'elements inclosos en la col·lecció i el nombre d'elements del món real.</p> <p>I nivell de descripció dels recursos (ús de metadades, especificitat dels recursos, etc.):</p> <ul style="list-style-type: none"> - A quin nivell es poden seleccionar els recursos catalogats? - Quina és la quantitat mínima d'informació necessària per crear una descripció dels recursos al catàleg? - És a dir, quina és la informació bàsica que ha de contenir un recurs per a ser seleccionat? (aquí s'inclourà el criteri de que utilitzi els 15 elements mínims de descripció, de l'estàndard de metadades Dublin Core) 			
<p>Disponibilitat/Accessibilitat (1,2,3) [Nivell de</p>	<ul style="list-style-type: none"> - Què hi ha disponible fora de la col·lecció? Si/No - Quin és el valor relatiu o el recurs en comparació amb altres disponibles en aquest tema? Si/No - Existeixen materials similars 			

fiabilitat]	/millors sobre la matèria d'interès?			
Precisió (1,2, 3)	<p>Quocient entre el nombre de valors correctes recuperats pel sistema i el nombre total de valors continguts en el sistema. (Valor no determinant, perquè les col·leccions dels sistemes analitzats són diferents i no són comparables, a més a més, de no tenir accessibilitat al 100% de totes les col·leccions). (DQ)</p> <p>-La informació es pot verificar amb exactitud? Si/No</p> <p>-El recurs, mostra cites bibliogràfiques o referències per confirmar la informació? Si/No</p> <p>-La presentació de la informació en la RI és correcta (sense faltes ortogràfiques/gramaticals)? Si/No</p> <p>-Hi ha prevalença d'errors ortogràfics? Si/No</p>			
Rellevància (1, 2,3) [precisió, redundància, aplicabilitat i utilitat]	<p>Atribut subjectiu, no valorable col·lectivament perquè és l'usuari final qui valora la col·lecció segons les seves necessitats. Aspectes valuables</p> <p>- Precisió del domini, aplicabilitat, utilitat, etc.? Si/No (audiència vs. Col·lecció)</p> <p>- Hi ha accés de navegació per matèries (thesaurus, etc.)? Si/No</p> <p>- Hi ha possibilitat de cercar termes genèrics i específics (navegabilitat per estructura</p>			

	<p>jeràrquica)? Si/No</p> <ul style="list-style-type: none"> - Fiabilitat de la cerca per classificació/descriptors de termes específics? Si/No - Consistència dels termes d'indexació (un concepte/tema ha d'aparèixer d'una forma única)? Si/No - Suggestions de cerca per termes/conceptes alternatius? Si/No - Capacitat per discriminació de termes? Si/No 			
--	---	--	--	--

*(Dimensions: 1=Interrogació; 2= Serveis i 3=Col·lecció)

Annex 2.4. Cas INTUTE

2.4.1. Interfícies

- Interfície (consulta, juliol 2011)

intute Helping you find the best websites for study and research

Intute closed in July 2011
- see FAQ for details -

Home | Web resources | Internet training | All services | Support for... | About us | FAQs

Browse Web resources by subject

- [Agriculture, food, and forestry](#)
- [Architecture and planning](#)
- [Biological sciences](#)
- [Business and management studies](#)
- [Communication and media studies](#)
- [Creative and performing arts](#)
- [Education and research methods](#)
- [Engineering](#)
- [Geography and environment](#)
- [Humanities](#)
- [Law](#)
- [Mathematics and computer science](#)
- [Medicine including dentistry](#)
- [Modern languages and area studies](#)
- [Nursing, midwifery and allied health](#)
- [Physical sciences](#)
- [Psychology](#)
- [Social sciences](#)
- [Veterinary medicine](#)

Search Web resources

[Advanced search](#) | [Help](#) | [Add this search box to your website](#)

Virtual Training Suite
The [Virtual Training Suite](#) provides free Internet tutorials on using the Web for education and research, written by qualified lecturers and librarians.

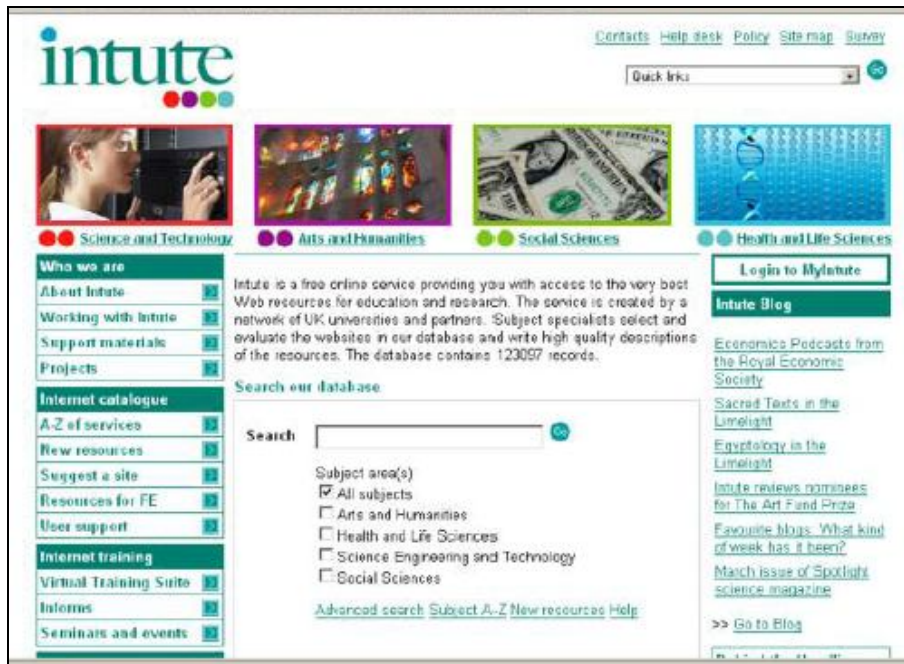
Intute blog

[Intute plans for the future - 2010 and beyond](#)
12 Apr 2010 [Comments](#)

[Goodbye and thanks](#)

Intute closed in July 2011. Please note that the site has now been archived and we are not adding any further resources to the catalogue or making any changes to existing content.

- Interfície (consulta desembre 2008)



2.4.3. Cerques INTUTE

- Equació de cerca: (Microalgues i toxines) OR (Microalgae and toxins)

(ALGAL TOXIN STANDARDS) OR (TOXIN STANDARDS AND ALGAE)
 OR (MICROALGAE AND TOXIN STANDARDS) OR (TOXIC
 MICROALGA* AND STANDARDS) OR (BIOLOGICAL POISONS AND
 ALGAE) OR (BIOLOGICAL POISONS AND MICROALGAE)

http://www.intute.ac.uk/cgi-bin/advancedsearch.pl?field=All&term1=(ALGAL+TOXIN+STANDARDS)+OR+(TOXIN+STANDARDS+AND+ALGAE)+OR+(MICROALGAE+AND+TOXIN+

Esta página está escrita en inglés . ¿Quieres traducirla? Traducir No Opciones

intute Helping you find the best websites for study and research

Home Web resources Internet training All services Support for... About us Feedback

Home > Web resources > Advanced search > Results

Advanced search results - (ALGAL TOXIN STANDARDS) OR (TOXIN STANDARDS AND ALGAE) OR (MICROALGAE AND TOXIN STANDARDS) OR (TOXIC MICROALGAE AND STANDARDS) OR (BIOLOGICAL POISONS AND ALGAE) OR (BIOLOGICAL POISONS AND MICROALGAE)

Search web resources

(ALGAL TOXIN STANDARDS) OR (TOXIN STANDARDS AND ALGAE) OR (MICROALGAE)

[Search](#) [Advanced search](#) | [Help](#)

Subject(s): agriculture, biologicalsciences Resource type(s): All

[Save this search](#)

1 - 25 of 341 records

Order by: Relevance | [Alphabetical](#) 1 | 2 | 3 | 4 | 5 | 14 Next

Algal blooms
http://portal.water.wa.gov.au/portal/page/portal/WaterQuality/Publications/Water
One of a series of fact sheets produced by the Waters and Rivers Commission of the Government of Western Australia aimed at the general public. This fact sheet focuses on **algal blooms** as a public health concern and an ecological problem in wetlands, waterways (rivers and estuaries) and oceans of south west Western Australia. Topics featured include the role of **algae** in waterways, alien species that travel around within the ballast water of ships, eutrophication, phytoplankton and the possible dangers of **toxin-producing blue-green algae**.

MyIntute
Hello Carmen RevetÁ©. Here is an overview of your MyIntute account.
Saved records: 3
Saved searches: 1
[Go to MyIntute](#)
[Logout](#)

- Equació de cerca 2: (Microalgues i toxines) OR (Microalgae and toxins) AND STUDIES

http://www.intute.ac.uk/cgi-bin/advancedsearch.pl?field=All&term1=(FOODBORN*+OR+FOODBORNE+ILLNESS+OR+FOODBORNE+DISEASES+OR+FOOD+POISONING)+AND+(TOXIC+ALGAE+OR+TOXIC+MICROALGAL+OR+TOXIC+MICROALGAE+OR+ALGAL+TOXICITY)+AND+(PUBLIC+HEALTH+OR+FOOD+SAFETY)+AND+(EPIDEMIOLOGICAL+STUDIES+OR+EPIDEMIOLOGICAL+STUDY+OR+STUDIES)

Esta página está escrita en inglés . ¿Quieres traducirla? Traducir No Opciones

intute Helping you find the best websites for study and research

Home Web resources Internet training All services Support for... About us Feedback

Home > Web resources > Advanced search > Results

Advanced search results - (FOODBORN* OR FOODBORNE ILLNESS OR FOODBORNE DISEASES OR FOOD POISONING) AND (TOXIC ALGAE OR TOXIC MICROALGAL OR TOXIC MICROALGAE OR ALGAL TOXICITY) AND (PUBLIC HEALTH OR FOOD SAFETY) AND (EPIDEMIOLOGICAL STUDIES OR EPIDEMIOLOGICAL STUDY OR STUDIES)

Search web resources

(FOODBORN* OR FOODBORNE ILLNESS OR FOODBORNE DISEASES OR FOOD POISON

[Search](#) [Advanced search](#) | [Help](#)

Subject(s): agriculture, biologicalsciences, education, geography, veterinary Resource type(s): All

[Save this search](#)

1 - 3 of 3 records

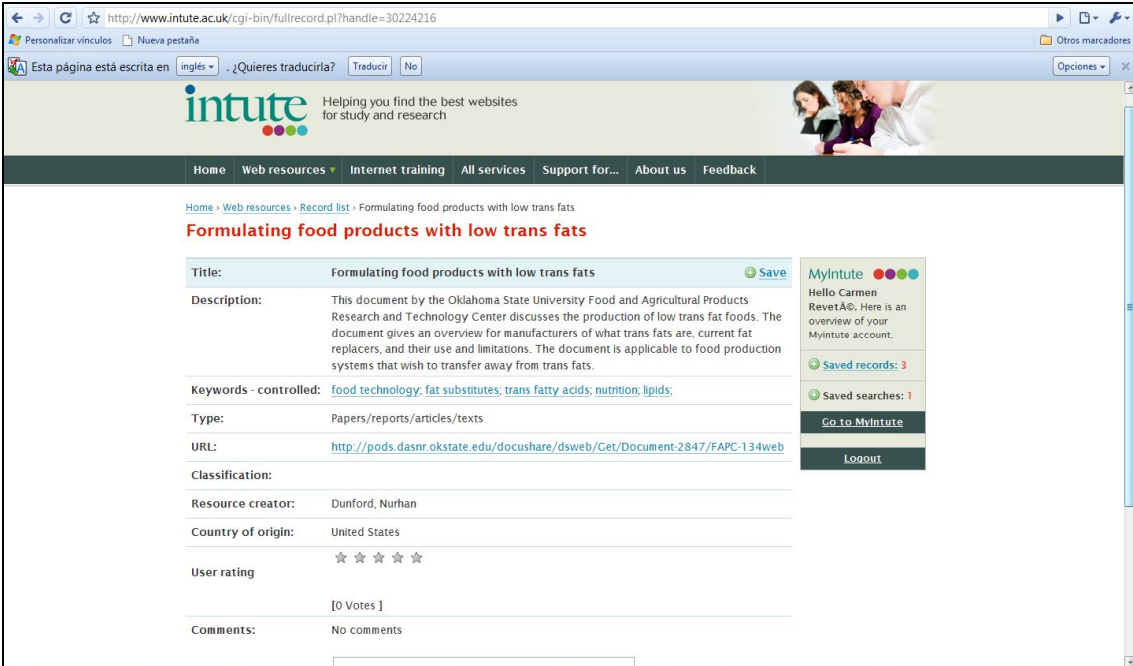
Order by: Relevance | [Alphabetical](#)

National toxicology program
http://ntp-server.niehs.nih.gov/
The National Toxicology Program (NTP) is a United States interagency organisation involving, among others, the National Institute of Environmental Health Sciences and the National Institute of Occupational Health and Safety. It develops scientific information about potentially **toxic chemicals** which can be used for the protection of **public health** and prevention of chemically induced **diseases**. This home page contains extensive testing information including searchable **abstracts of studies underway and completed for all test species of short-term toxicity and long-term**

MyIntute
Hello Carmen RevetÁ©. Here is an overview of your MyIntute account.
Saved records: 3
Saved searches: 0
[Go to MyIntute](#)
[Logout](#)

2.4.4. Registros de Metadades

- Exemple 1



The screenshot shows a web browser window displaying the Intute website. The URL is <http://www.intute.ac.uk/cgi-bin/fullrecord.pl?handle=30224216>. The page title is "Formulating food products with low trans fats". The Intute logo is at the top left, with the tagline "Helping you find the best websites for study and research". A navigation bar includes links for Home, Web resources, Internet training, All services, Support for..., About us, and Feedback. The main content area shows the record details for "Formulating food products with low trans fats".

Title: Formulating food products with low trans fats [Save](#)

Description: This document by the Oklahoma State University Food and Agricultural Products Research and Technology Center discusses the production of low trans fat foods. The document gives an overview for manufacturers of what trans fats are, current fat replacers, and their use and limitations. The document is applicable to food production systems that wish to transfer away from trans fats.

Keywords - controlled: [food technology](#); [fat substitutes](#); [trans fatty acids](#); [nutrition](#); [lipids](#);

Type: Papers/reports/articles/texts

URL: <http://pods.dasnr.okstate.edu/docushare/dsweb/Cet/Document-2847/FAPC-134web>

Classification:

Resource creator: Dunford, Nurhan

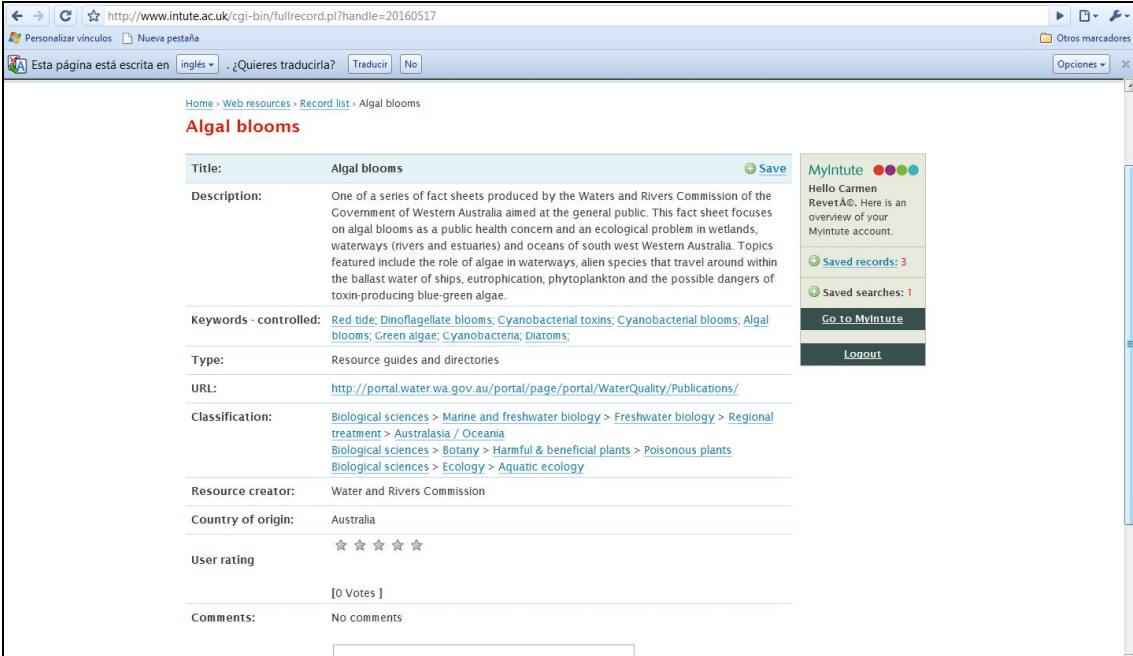
Country of origin: United States

User rating: ☆ ☆ ☆ ☆ ☆ [0 Votes]

Comments: No comments

On the right side, there is a "MyIntute" user profile for "Hello Carmen RevetA©" with an overview of the account, 3 saved records, and 1 saved search. There are buttons for "Go to MyIntute" and "Logout".

- Exemple 2



The screenshot shows a web browser window displaying the Intute website. The URL is <http://www.intute.ac.uk/cgi-bin/fullrecord.pl?handle=20160517>. The page title is "Algal blooms". The Intute logo is at the top left, with the tagline "Helping you find the best websites for study and research". A navigation bar includes links for Home, Web resources, Internet training, All services, Support for..., About us, and Feedback. The main content area shows the record details for "Algal blooms".

Title: Algal blooms [Save](#)

Description: One of a series of fact sheets produced by the Waters and Rivers Commission of the Government of Western Australia aimed at the general public. This fact sheet focuses on algal blooms as a public health concern and an ecological problem in wetlands, waterways (rivers and estuaries) and oceans of south west Western Australia. Topics featured include the role of algae in waterways, alien species that travel around within the ballast water of ships, eutrophication, phytoplankton and the possible dangers of toxin-producing blue-green algae.

Keywords - controlled: [Red tide](#); [Dinoflagellate blooms](#); [Cyanobacterial toxins](#); [Cyanobacterial blooms](#); [Algal blooms](#); [Green algae](#); [Cyanobacteria](#); [Diatoms](#);

Type: Resource guides and directories

URL: <http://portal.water.wa.gov.au/portal/page/portal/WaterQuality/Publications/>

Classification: [Biological sciences > Marine and freshwater biology > Freshwater biology > Regional treatment > Australasia / Oceania](#)
[Biological sciences > Botany > Harmful & beneficial plants > Poisonous plants](#)
[Biological sciences > Ecology > Aquatic ecology](#)

Resource creator: Water and Rivers Commission

Country of origin: Australia

User rating: ☆ ☆ ☆ ☆ ☆ [0 Votes]

Comments: No comments

On the right side, there is a "MyIntute" user profile for "Hello Carmen RevetA©" with an overview of the account, 3 saved records, and 1 saved search. There are buttons for "Go to MyIntute" and "Logout".

2.4.5. Extracció resultats: e-mail, html, .txt i javascript

The screenshot shows the MyIntute website. At the top, the 'intute' logo is displayed with the tagline 'Helping you find the best websites for study and research'. A navigation menu includes 'Home', 'Web resources', 'Internet training', 'All services', 'Support for...', 'About us', and 'Feedback'. Below the menu, there are links for 'MyIntute home', 'Edit your profile', and 'Logout'. A notice states that email alerts are currently not working and provides an RSS feed link: <http://www.intute.ac.uk/rss/intute/latest.xml>. The main content area is divided into two sections: 'My bookmarks' and 'My Alerts'. The 'My bookmarks' section shows a list of bookmarked items with options to export them in 'Email', 'Plain Text', 'HTML', or 'javascript' formats. It also includes a 'Tags for my bookmarks' section with a 'No tag' button and a list of tags. The 'My Alerts' section provides a list of subject areas with checkboxes to select or unselect them.

Annex 2.5. Cas VASCODA

2.5.1. Cerques VASCODA

Es comprova que ni el sistema és prou avançat ni la col·lecció és suficientment especialitzada. Tampoc es poden fer cerques extenses (vegeu cerca 3 i 4). Només es pot fer cerca de màxim 20 paraules. L'Avantatge d'aquest portal és el servei que ofereix de resultats externs "External search results".

- Cerca 1

The screenshot shows the VASCODA search interface. The browser address bar displays 'http://www.vascoda.de/vascoda'. The page title is 'vascoda The Internetportal for Scientific Information'. The search bar contains the query 'ALGAE OR MICROALGAE AND STANDARDS AND TOXIN'. The search results section shows '0 Search results for: (ALGAE OR MICROALGAE) AND STANDARDS AND TOXIN'. It also indicates that the search term has been extended to synonyms, related terms, and translations, and that there are external search results available in subject portals. A 'Tips' section provides advice on how to improve search results.

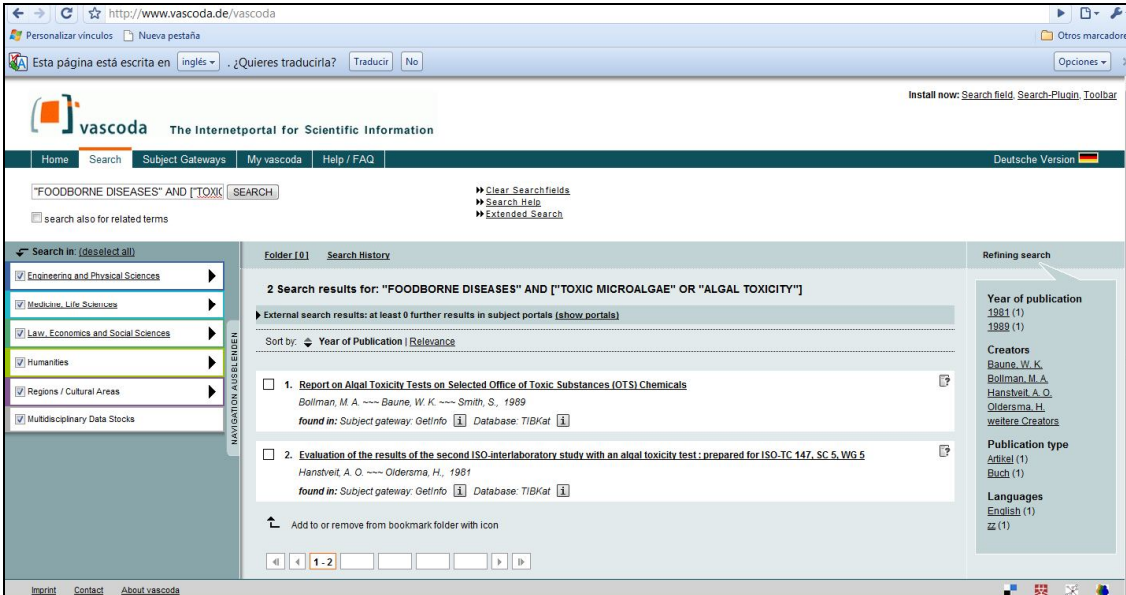
- Cerca 2

The screenshot shows the Vascoda search interface. The search bar contains the query: `FOODBORN* AND TOXIC ALGAE AND (PUBLIC HEALTH OR FOOD SAFETY) AND EPIDEMIOLOGICAL STUDIES`. The search results section displays: **0 Search results for: FOODBORN* AND { TOXIC AND ALGAE } AND ({ PUBLIC AND SAFETY } OR (HEALTH OR FOOD)) AND (EPIDEMIOLOGICAL AND STUDIES)**. Below this, it states: **Your search term has been extended to synonyms, related terms and translations (display/hide)**. It also mentions: **External search results: at least 0 further results in subject portals (show portals)**. A 'Tips' section provides advice on using external search results and related terms.

- Cerca 3

The screenshot shows the Vascoda search interface with an error message. The search bar contains the query: `EPIDEMIOLOGICAL STUDY OR STUDIES) (FOODBORN* OR FOODBORNE ILLNES...`. The search results section displays an error: **Error: HTTP-Status-Code 413 Request Entity Too Large**. Below the error, it states: **The Query has too many Searchwords - it must be smaller than 20 words, but it has had 37 searchwords**. A message explains: **By selecting or dropping subject groups, subject / regions and subject portals on the left hand side you can limit the search area.** It also provides a link to [Search help](#).

2.5.2. Registres Vascoda



The screenshot shows the Vascoda search interface. The search query is "FOODBORNE DISEASES" AND ["TOXIC MICROALGAE" OR "ALGAL TOXICITY"] and the results are sorted by Year of Publication. Two results are displayed:

- 1. Report on Algal Toxicity Tests on Selected Office of Toxic Substances (OTS) Chemicals**
Bollman, M. A. --- Baune, W. K. --- Smith, S., 1989
found in: [Subject gateway](#); [Getinfo](#) [Database: TIBKat](#)
- 2. Evaluation of the results of the second ISO-interlaboratory study with an algal toxicity test: prepared for ISO-TC 147, SC 5, WG 5**
Hanstveit, A. O. --- Olderama, H., 1981
found in: [Subject gateway](#); [Getinfo](#) [Database: TIBKat](#)

The interface includes a navigation sidebar on the left with categories like Engineering and Physical Sciences, Medicine, Life Sciences, Law, Economics and Social Sciences, Humanities, Regions / Cultural Areas, and Multidisciplinary Data Stocks. A 'Refining search' panel on the right allows filtering by Year of publication (1981 (1), 1989 (1)), Creators (Baune, W. K., Bollman, M. A., Hanstveit, A. O., Olderama, H., weitere Creators), Publication type (Artikel (1), Buch (1)), and Languages (English (1), zz (1)).

Nota: l'extracció de resultats a través de Vascoda, només hi ha opció per enviar-ho via correu electrònic. I sense subscripció no es poden analitzar la resta d'opcions ofertes.

Annex 2.6. Cas AustLII

2.6.1. Interfície General

The screenshot displays the AustLII website interface. At the top, there is a navigation bar with links for 'About AustLII', 'Media Releases', 'Contact', 'Translate', 'Copyright & Usage', 'Privacy', 'Disclaimers', 'Feedback', and 'Help'. The main header features the AustLII logo and the text 'Australasian Legal Information Institute' and 'A joint facility of UTS and UNSW Faculties of Law'. A search bar is located below the header, with links for 'Advanced Search', 'LawCite', and 'Help'. The main content area is divided into several sections: 'Cases & Legislation' with a sub-menu for 'Australia' and 'New Zealand (NZLII)'; 'Libraries' with a list of legal subjects; 'Catalog' with a sub-menu for 'Australia' and 'New Zealand'; 'News & Database Additions' with a list of recent news items; 'Update Status' with links for 'Update Status for Case Law', 'Update Status for Journals', and 'Update Status for Legislation'; 'AustLII Contributors' with a list of contributors; 'Contribute \$\$\$' with links for 'All Contributors' and 'By Jurisdiction'; 'Research Projects' with a list of projects; and 'Special Features & Tools' with links for 'AustLII User Guide', 'AustLII Apps', 'AustLII Toolbar', 'AustLII Research Seminars', 'Subscribe: AustLII News', and 'More...'. At the bottom, there is a footer stating 'AustLII participates in the free access to law movement, a grouping of the following organisations:'.

www.austlii.edu.au

Search the web (Babylon)

About AustLII | Media Releases | Contact | Translate | Copyright & Usage | Privacy | Disclaimers | Feedback | Help

Australasian Legal Information Institute
A joint facility of UTS and UNSW Faculties of Law

[Advanced Search] [LawCite] [Help] Search

554 databases from all Australasian jurisdictions

Last updated 3 September 2013

News & Database Additions

- [Law via the Internet Conference 2013](#), Jersey, Channel Islands, 26-27 September 2013
- [Construction Forestry Mining & Energy Union v Mamoet Australia Pty Ltd \[2013\] HCA 36](#) (14 August 2013)
- [Legal Services Board v Gillespie-Jones \[2013\] HCA 35](#) (14 August 2013)
- [Historic South Australian Law Report Series now available for free access](#) (PDF) (7 August 2013)
- [South Australian Law Reports 1862-1920](#)
- [State Reports \(South Australia\) 1921-1950](#)
- [New Zealand Advertising Standards Authority 2001-](#)
- [New Zealand Real Estate Agents Authority 2010-](#)
- [New Zealand Lost Cases 1842-1882](#)
- [New Zealand Alcohol Regulatory and Licensing Authority 2013-](#)
- [New Zealand Review Authority 2011-](#)
- [New Zealand Supreme Court Transcripts 2011-](#)
- [New Zealand Law Students' Journal 2010-](#)
- [Fair Work Commission 2013-](#)
- [Fair Work Commission - Full Bench 2013-](#)
- [Fair Work Commission - Enterprise Agreement 2013-](#)
- [Fair Work Commission - General Manager and Delegates 2013-](#)
- [Law Reports \(New South Wales\) 1880-1900](#)
- [State Reports \(New South Wales\) 1901-1950](#)
- [Forest Practices Tribunal of Tasmania 2002-](#)

Past announcements...

Latest Australasian Catalog Additions

- [Australian Planning & development Law Library](#)
Link added: >> By Subject >> [Planning & Development](#)
- [Past additions...](#)

Update Status

- [Update Status for Case Law](#)
- [Update Status for Journals](#)
- [Update Status for Legislation](#)

AustLII Contributors

[AustLII Annual Report 2012 \(2011 2010 2009 2008 2007\)](#)
[Law Council encourages contributions](#) (PDF)

Contribute \$\$\$

- [All Contributors](#)
- [By Jurisdiction](#)
- [Education](#)
- [Business & Industry](#)
- [Community](#)
- [Sociologists, Grant Decisions](#)
- [Courts, Tribunals & Regulators](#)
- [Government Agencies](#)
- [Law Societies & Bar Associations](#)
- [Legal Profession](#)

Research Projects

- [Australasian Legal History Library](#)
- [Australian Legal Scholarship Library](#)
- [Improving Online Case Law](#)
- [Point-in-Time Legislation](#)
- [SINO Search Engine](#)
- [Victoria Model Jurisdiction](#)
- [More...](#)

Special Features & Tools

- [AustLII User Guide](#)
- [AustLII Apps](#)
- [AustLII Toolbar](#)
- [AustLII Research Seminars](#)
- [Subscribe: AustLII News](#)
- [More...](#)

AustLII participates in the [free access to law](#) movement, a grouping of the following organisations:

2.6.2. Interfície de Cerca Avançada

Advanced Search

You are here: [AustLII](#) >> Advanced Search

To search for legal material not on AustLII use [WorldLII](#).

Enter search query: Search Reset

Find this Boolean query [\[Help\]](#) [\[Search operation\]](#)

Select the AustLII Database(s) to search:

- All AustLII Databases
- All Legislation Databases
- All Case Law Databases
- All Journals Databases
- All Secondary Materials Databases (includes journals)
- Commonwealth: All Primary Materials
- Commonwealth: All Legislation
- Commonwealth: All Cases
- Commonwealth: Bills
- Commonwealth: Bills Explanatory Memoranda
- Commonwealth: Bills Digests
- Commonwealth: Consolidated Acts
- Commonwealth: Consolidated Regulations
- Commonwealth: Legislation Tables
- Commonwealth: Numbered Acts

Find this Boolean query

- using autosearch
- this Boolean query
- any of these words
- all of these words
- this phrase
- this legislation name
- this case name
- this document title

2.6.3. Cerques: resultats

- Exemple de presentació dels resultats per ordre de rellevància (interfície per defecte)

environment" and (conserv* or preserv* or protection or *impact

AustLII Databases - Documents found: 167805 for (environment* and (conserv* or preserv* or protection or * impact analysis" or control or maintenance or management or custody or supervision))

By Citation Frequency By Database By Date **By Relevance** By Title

Page 1 of 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | >>>

Results per page: 10 20 50 100

1. pdf "Environment Protection Amendment (Beverage Container Deposit and Recovery Scheme) Bill 2009 - Research Brief Number 3" [2009] VicBillsRR 6 (1 May 2009) [100%]
(From Victorian Bills Research Resources; 56870 KB) [LawCite](#)
2. pdf Brouwer, G E --- "Investigation into the handling of drug exhibits at the Victoria Police Forensic Services Centre (1.12MB)" [2009] VicOmbPRp 11 (1 December 2009) [52%]
(From Victorian Ombudsman Parliamentary Reports; 71220 KB) [LawCite](#)
3. pdf Brouwer, G E --- "Conditions for persons in custody (1.22MB)" [2006] VicOmbPRp 4 (1 July 2006) [35%]
(From Victorian Ombudsman Parliamentary Reports; 60932 KB) [LawCite](#)
4. pdf "Transport Legislation Amendment (Ports Integration) Bill 2010 - Research Brief Number 7" [2010] VicBillsRR 11 (1 May 2010) [24%]
(From Victorian Bills Research Resources; 47638 KB) [LawCite](#)
5. pdf Brouwer, G E --- "Ombudsman investigation into the probity of the Kew Residential Services and St Kilda Triangle developments (1.71MB)" [2010] VicOmbPRp 5 (23 June 2010) [22%]
(From Victorian Ombudsman Parliamentary Reports; 88960 KB) [LawCite](#)
6. pdf "Tobacco Amendment (Protection of Children) Bill 2009 - Research Brief Number 4" [2009] VicBillsRR 8 (1 July 2009) [22%]
(From Victorian Bills Research Resources; 53686 KB) [LawCite](#)
7. pdf "Domestic Animals Amendment (Dangerous Dogs) Bill 2010 - Research Brief Number 6" [2010] VicBillsRR 9 (1 May 2010) [21%]
(From Victorian Bills Research Resources; 68688 KB) [LawCite](#)
8. pdf Brouwer, G E --- "Crime statistics and police numbers report (3.97MB)" [2009] VicOmbPRp 1 (11 March 2009) [13%]

- Exemple de presentació de resultats per freqüència de citació

The screenshot shows the AustLII search results page. The search query is "environment* and (conserv* or preserv* or protection or *impact". The results are sorted by Citation Frequency, showing the 10000 most relevant results. The first result is "Commonwealth v Tasmania ("Tasmanian Dam case") [1983] HCA 21; (1983) 158 CLR 1 (1 July 1983)". Other results include "Graham Barclay Oysters Pty Ltd v Ryan", "Western Australia v Ward", "Environment Protection Authority v Caltex Refining Co Pty Ltd", "Western Australia v Ward (includes summary dated 3 Mar 2000 & corrigendum dated 23 Mar 2000)", "THE BELL GROUP LTD (IN LIQ) -v- WESTPAC BANKING CORPORATION", "Thomas v Mowbray", and "Cubillo v Commonwealth".

2.6.4 Registres AustLII

The screenshot shows the AustLII page for the Supreme Court of New South Wales - Court of Criminal Appeal. The case title is "Environment Protection Authority v Truegain Pty Ltd [2013] NSWCCA 204 (30 August 2013)". The page includes a breadcrumb trail: "You are here: AustLII >> Databases >> Supreme Court of New South Wales - Court of Criminal Appeal >> 2013 >> [2013] NSWCCA 204". A red box highlights the breadcrumb trail and a note: "Opció de descàrrega de registre, només apareix a la pàgina del registre individual". The page also shows the case title with a green arrow pointing to "Environment Protection" and another green arrow pointing to "Authority v Truegain Pty Ltd". The last updated date is "2 September 2013".

2.6.5. Extracció d'Informació

Com la imatge anterior indica, la opció d'extracció de la informació, només es pot du a terme de forma individual per cadascun dels registres. Depenent del registre resultant, la exportació potser directa, és a dir, si el registre està en format pdf es pot procedir a guardar el registre des de les opcions de Windows habituals (Arxiu → “Desar o Guardar como”), o bé, si el registre d'interès té format html, es selecciona el títol del recurs resultant d'interès, i s'accedeix al document complert, i com apareix en la imatge superior, apareix un menú a sobre del document on hi ha l'opció de “Download”, que permet exportar el document. En cap cas, hi ha l'opció d'exportar la referència/cita bibliogràfica.

Annex 2.7. Cas Libris

2.7.1. Interfície d'Inici

The screenshot shows the LIBRIS homepage. At the top, there's a navigation bar with 'LIBRIS' logo and links for 'HELP | PÅ SVENSKA | PREFERENCES ✓ | MY LIBRARIES | CLEAR HISTORY'. Below this is a secondary navigation bar with 'Start', 'Extended search', 'Index A-Ö', 'Boolean', 'Subdatabases', and 'Search history'. A search bar is present with a 'Search' button and a 'My libraries' link. The main content area is divided into three columns: 'Explore' (with a magnifying glass icon), 'Customize' (with an 'A' icon), and 'What is LIBRIS?' (with a question mark icon). Each column contains descriptive text and links. Below these columns are several book covers, including 'Emil in Lönneberga 50 years'. A horizontal navigation bar below the covers lists categories like 'Swedish audio books in Libris', 'Emil in Lönneberga 50 years', 'New free e-resources', and 'New Swedish'. At the bottom, there's a footer with five columns of links: 'About LIBRIS', 'Help', 'On your site', 'Customize', and 'Search facilities'.

2.7.2. Cerques

- **Cerca simple:** “foodborne* (alga* OR plants OR cianófitas OR Cyanophyta OR cianobacteria OR Cyanobacteria)”

LIBRIS HELP | PÅ SVENSKA | PREFERENCES ✓ | MY LIBRARIES | CLEAR HISTORY






Start Extended search Index A-Ö Boolean Subdatabases Search history

foodborne* (ALGA* OR Plants OR cianofitas OR Cyanophyta OR cianob) My libraries

Search: foodborne* (ALGA* OR Plants OR cianofitas OR Cyanophyta OR cianobacteria OR Cyanobacteria)

Result 1-8 of 8 [1] 2 3 4 5 6 7 ... 17

Sort by: relevancy Group by: none Hits per page: 10

1.		Algal toxins in seafood and drinking water / edited by Ian R. Falconer 1993 Book	1 library
2.		Andersson, H. Christer (author) Calystegine alkaloids in solanaceous food plants / by H. Christer Andersson 2002 Book	6 libraries
3.		Barceloux, Donald G. (author) Medical toxicology of natural substances [Elektronisk resurs] : foods, fungi, medicinal herbs, plants, and venomous animals / Donald G. Barceloux. 2008 E-book	1 library
4.		Lundén, Janne (author) Persistent Listeria monocytogenes contamination in food processing plants / Janne Lundén 2004 Book Thesis	1 library
5.		Thougaard, Herluf (author) [Mikrobiologi. Svenska] Grundläggande mikrobiologi med livsmedelsapplikationer / Herluf Thougaard, Verner Varlund, Rene Møller Madsen ; översättning och svensk bearbetning: Anna Blücher 2007. - 2., uppdaterade och utök. uppl.	37 libraries

My libraries

No libraries selected

Refine your search

Libraries

- Chalmers Tekniska ... (2)
- Göteborgs universi ... (2)
- Högskolan Dalarna (1)
- Högskolan i Halmst ... (1)
- Högskolan i Kristi ... (1)
- show more...

Type

- book (8)
- e-resource (3)
- theses (1)

Author

- Falconer, Ian R. (1)
- Andersson, H. Christ ... (1)
- Barceloux, Donald G. (1)
- Blücher, Anna, 1948- (1)
- Guerrero-Legarreta, ... (1)
- show more...

Language

- English (7)
- Swedish (1)

Year

Subject (SAB)

• Visualització en format de llista curta

LIBRIS HELP | PÅ SVENSKA | PREFERENCES ✓ | MY LIBRARIES | CLEAR HISTORY

Start Extended search Index A-Ö Boolean Subdatabases Search history

foodborne* (ALGA* OR Plants OR cianofitas OR Cyanophyta OR cianobac) My libraries

Search: foodborne* (ALGA* OR Plants OR cianofitas OR Cyanophyta OR cianobacteria OR Cyanobacteria)

Result 1-8 of 8 [1] 2 3 4 5 6 7 ... 17

Sort by: relevancy Group by: none Hits per page: 10

1.	Algal toxins in seafood and drinking wat...	1993	1 lib.
2.	Andersson, H. Christer Calystegine alkaloids in solanaceous foo...	2002	6 lib.
3.	Barceloux, Donald G. Medical toxicology of natural substances...	2008	1 lib.
4.	Lundén, Janne Persistent Listeria monocytogenes contam...	2004	1 lib.
5.	Thougaard, Herluf Grundläggande mikrobiologi med livsmedel...	2007	37 lib.
6.	Health and the food-chain / scientific e...	2000	3 lib.
7.	Handbook of poultry science and technolo...	2010	1 lib.
8.	Russell, Scott M. Controlling Salmonella in poultry produc...	2012	3 lib.

Select all

My libraries

No libraries selected

Refine your search

Libraries

- Chalmers Tekniska ... (2)
- Göteborgs universi ... (2)
- Högskolan Dalarna (1)
- Högskolan i Halmst ... (1)
- Högskolan i Kristi ... (1)
- show more...

Type

- book (8)
- e-resource (3)
- theses (1)

Author

- Falconer, Ian R. (1)
- Andersson, H. Christ ... (1)
- Barceloux, Donald G. (1)
- Blücher, Anna, 1948- (1)
- Guerrero-Legarreta, ... (1)
- show more...

Language

- English (7)
- Swedish (1)

Year

Subject (SAB)

- **Cerca avançada (o estesa): 4 resultats però més pertinents, ús de camp de cerca de “vocabularis controlats” o “subject”**

- Cerca booleana

- **Cerca per índexs:** cerca i navegació per llistes alfabètiques d'autors, temes, títols de llibres, etc.

LIBRIS [HELP](#) | [PÅ SVENSKA](#) | [PREFERENCES](#) | [MY LIBRARIES](#) | [CLEAR HISTORY](#)

Start | Extended search | **Index A-Ö** | Boolean | Subdatabases | Search history

Enter term and choose index

Term:

Choose index:

- author (person)
- author (person)**
- subject
- book title
- journal title
- series title
- corporate body
- conference
- host publication
- reviewed work

[Browse](#)

Index

Author (person):
Titles by a certain person

Subject:
Titles about a subject or a certain person, corporate body, place etc.

Book title:
Books, maps, journal articles etc.

Journal title:
Journals, year books, report series, newspapers etc.

Series title:
Books pertaining to a certain report series etc.

[show more...](#)

en by August Strindberg?
term field and choose the

What is published about Selma Lagerlöf?
Enter Lagerlöf, Selma in the term field and choose Subject index.

Which records and movies are recorded by ABBA?
Enter ABBA in the term field and choose index. Corporate body

- **Cerca per sub-bases de dades:** LIBRIS no només conté les col·leccions de les biblioteques acadèmiques i de recerca de Suècia, sinó que també conté el conjunt de la Bibliografia Nacional de Suècia i una selecció de bibliografia especialitzada en diverses àrees produïdes per les biblioteques LIBRIS. En aquesta opció de cerca es pot cercar en subconjunts de la base de dades LIBRIS fent clic a "Subdatabases". Cada bibliografia és responsable de seleccionar el material que cobrirà (articles de revistes, llibres, enregistraments de so, etc.)

LIBRIS HELP | PÅ SVENSKA | PREFERENCES | MY LIBRARIES | CLEAR HISTORY

Start Extended search Index A-Ö Boolean Subdatabases Search history

[What are subdatabases?](#) ?

Swedish National Bibliographic Databases, Swedish Hand Press Publications ...

<p>The Swedish National Bibliographic Databases ⓘ</p> <p>Books, periodicals, maps, printed music ...</p> <p>➤ Over 500 000 hits to search form</p>	<p>Swedish Hand Press Publications ⓘ</p> <p>Swedish Bibliography 1600-1829 and the Swedish Public Print -1833</p> <p>➤ Over 70 000 hits to search form</p>
<p>Suecana Extranea ⓘ</p> <p>Translations from Swedish and foreign literature about Sweden</p> <p>➤ Over 50 000 hits to search form</p>	<p>Suecia Antiqua et Hodierna ⓘ</p> <p>355 digitized plates from Erik Dahlbergh's 17th century work about Sweden</p> <p>➤ Under 1 000 hits to search form</p>

The humanities, the social sciences ...

<p>KVINNSAM ⓘ</p> <p>➤ Over 100 000 hits to search form</p>	<p>Swedish Historical Bibliography ⓘ</p> <p>➤ Over 90 000 hits to search form</p>
<p>Arblina ⓘ</p> <p>➤ Over 75 000 hits to search form</p>	<p>Swedish Jewish Bibliography ⓘ</p> <p>➤ Over 60 000 hits to search form</p>
<p>Bibliography of Swedish Music Literature ⓘ</p> <p>➤ Over 25 000 hits to search form</p>	<p>Swedish Sami Bibliography ⓘ</p> <p>➤ Appr. 25 000 hits to search form</p>

[show more...](#)

Local and regional

<p>Bothnica ⓘ</p> <p>➤ Over 50 000 hits to search form</p>	<p>Småland, Öland & Blekinge ⓘ</p> <p>➤ Over 20 000 hits to search form</p>
<p>Dala Bibliography ⓘ</p> <p>➤ Over 10 000 hits to search form</p>	<p>Bohus Bibliography ⓘ</p> <p>➤ Over 5 000 hits to search form</p>
<p>Gotlandica ⓘ</p> <p>➤ Over 4 000 hits to search form</p>	<p>Västmanland Bibliography ⓘ</p> <p>➤ Over 4 000 hits to search form</p>

Authors and persons

<p>The Vilhelm Moberg Bibliography ⓘ</p> <p>➤ Over 6 000 hits to search form</p>
<p>Birgitta Bibliography ⓘ</p> <p>➤ Appr. 5 000 hits to search form</p>
<p>The Linné Bibliography ⓘ</p> <p>➤ Appr. 2 000 hits to search form</p>
<p>Litteraturbanken ⓘ</p> <p>➤ Under 1 000 hits to search form</p>

2.7.3. Registres

- Exemple de registre bibliogràfic

1 of 2 ◀ Previous record ▶ Next record ▶ ▶ To hitlist

Overview Details

Bacteriological and epidemiological studies of campylobacter spp. in Swedish broilers / Ingrid Hansson

Hansson, Ingrid, 1962- (author)
 ISBN 978-91-576-7362-6
 Uppsala : Dept. of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, 2007
 English 56 s.
 Series: Acta Universitatis agriculturae Sueciae, 1652-6880 ; 2007:63

[Read full text](#) (Full text)

Related links:
<http://epsilon.slu.se/...> (Abstract)

Book Thesis (Dis. (sammanfattning) Uppsala : Sveriges lantbruksuniv., 2007)

► **Subject headings**

SAVE CITE EMAIL ▶ Permalink

Get it Other editions

Loan | Interlibrary loan/request

Title available at 6 libraries. [Show map](#) • [Expand list](#)

Search further

More titles by ⓘ

- Hansson, Ingrid, 196 ...

More titles about ⓘ

- campylobacter jejuni
- broiler chickens
- epidemiology
- disease surveys
- disease transmission
- biological contamina ...
- show more...

Series ⓘ

- Acta Universitatis a ...
- More parts

Search outside LIBRIS

Title in Google Book Search: ⓘ

- Bacteriological and...

Extend your search to: ⓘ

- Google
- Google Book Search
- Google Scholar
- Scirus
- LibraryThing

- Visualització ampliada de l'exemple de registre bibliogràfic

1 of 2 | Previous record | Next record | To hitlist

Overview | Details

Bacteriological and epidemiological studies of campylobacter spp. in Swedish broilers / Ingrid Hansson

Hansson, Ingrid, 1962- (author)
 ISBN 978-91-576-7362-6
 Uppsala : Dept. of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, 2007
 English 56 s.
 Series: Acta Universitatis agriculturae Sueciae, 1652-6880 ; 2007:63

[Read full text](#) (Full text)

Related links:
<http://epsilon.slu.se> (Abstract)

Book Thesis (Diss. (sammanfattning) Uppsala : Sveriges lantbruksuniv., 2007)

Subject headings

Subject headings

- campylobacter jejuni (agrovoc)
- broiler chickens (agrovoc)
- epidemiology (agrovoc)
- disease surveys (agrovoc)
- disease transmission (agrovoc)
- biological contamination (agrovoc)
- foodborne diseases (agrovoc)
- transport of animals (agrovoc)
- slaughtering (agrovoc)
- environmental factors (agrovoc)
- analytical methods (agrovoc)
- laboratory diagnosis (agrovoc)
- sweden (agrovoc)

Classification

- 636.5 (DDC)
- Qdfi (kssb/8)

SAVE | CITE | EMAIL | Permalink

Get it | Other editions

Loan | [Interlibrary loan/request](#)

Title available at 6 libraries. [Show map](#) • [Expand list](#)

► **Libraries in northern Sweden** (1) [ADD AS FAVOURITE](#)

► **Libraries in Mid-Sweden** (2) [ADD AS FAVOURITE](#)

Search further

More titles by

- Hansson, Ingrid, 196 ...

More titles about

- campylobacter jejuni
- broiler chickens
- epidemiology
- disease surveys
- disease transmission
- biological contamina ...
- show more...

Series

- Acta Universitatis a ...
- More parts

Search outside LIBRIS

Title in Google Book Search:

- Bacteriological and...

Extend your search to:

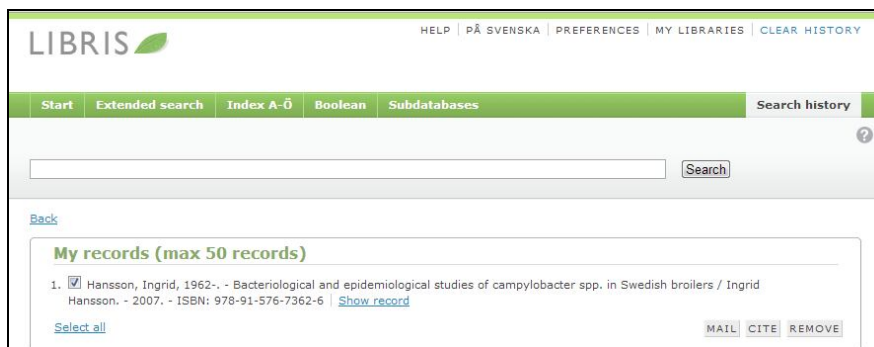
- Google
- Google Book Search
- Google Scholar
- Scirus
- LibraryThing

2.7.4. Extracció d'Informació

Opció disponible tant en la interfície de registres resultants d'una cercacom en la interfície individual d'un registre:

SAVE | CITE | EMAIL | Permalink

- **Extracció mitjançant “save”**: ens porta a les opcions de cite i e-mail



- **Extracció via e-mail**

The screenshot shows the 'Send mail' form. The title is 'Send mail'. Below it, the instruction reads: 'To mail the selected record, fill in the empty form below'. The form contains several fields: 'Send the record as:' with a dropdown menu showing 'Plain text' selected; 'Subject' with a text input field containing 'Hints from LIBRIS'; and 'Comment' with a large text area. At the bottom right of the form is a 'Send' button.

- **Extracció per cita bibliogràfica**

LIBRIS HELP | PÅ SVENSKA | PREFERENCES | MY LIBRARIES | CLEAR HISTORY

BIBLIOTEK
National Library
of Sweden

Start Extended search Index A-Ö Boolean Subdatabases Search history

[Back](#)

Citations

Create reference of the selected record

- Choose format
- Cut and paste
- or
- Click on Save as file

Format: Plain text F-8 character

All format encoding

- Plain text
- Harvard
- APA
- Vancouver
- RIS**
- BibTex
- RefWorks
- Bokref/Tidskriftsref
- MARC-XML

[Save as file](#)

Cut and paste

Plain text
Hansson, Ingrid, 1962-. - Bacteriological and epidemiological studies of campylobacter spp. in Swedish broilers / Ingrid Hansson. - 2007. - ISBN: 978-91-576-7362-6

Harvard
Hansson, Ingrid (2007). *Bacteriological and epidemiological studies of campylobacter spp. in Swedish broilers*. Diss. (sammanfattning) Uppsala : Sveriges lantbruksuniv., 2007
Available on Internet: <http://epsilon.slu.se/200763.pdf>

Oxford
Hansson, Ingrid, *Bacteriological and epidemiological studies of campylobacter spp. in Swedish broilers*, Dept. of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, Diss. (sammanfattning) Uppsala : Sveriges lantbruksuniv., 2007, Uppsala, 2007
<http://epsilon.slu.se/200763.pdf>

APA
Hansson, I. (2007). *Bacteriological and epidemiological studies of campylobacter spp. in Swedish broilers*. Diss. (sammanfattning) Uppsala : Sveriges lantbruksuniv., 2007. Uppsala.

Vancouver
Hansson I. Bacteriological and epidemiological studies of campylobacter spp. in Swedish broilers [dissertation]. Uppsala: Dept. of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences; 2007. Available from <http://epsilon.slu.se/200763.pdf>


RIS

Annex 2.8. Cas Europeana

2.8.1 Interfície

La imatge de fons i missatges d'entrada canvien a diari (vegeu Annex 7.2.1)


Home My Europeana Choose a language

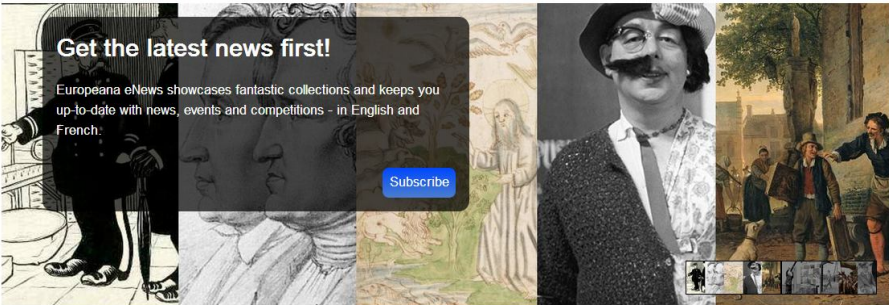

 [Search](#) [Help](#)

Get the latest news first!

Europeana eNews showcases fantastic collections and keeps you up-to-date with news, events and competitions - in English and French:

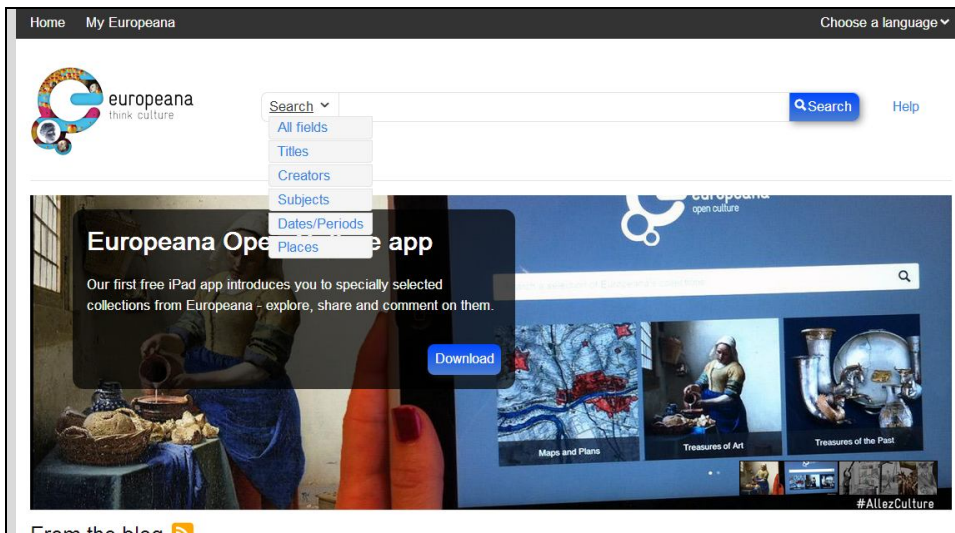
[Subscribe](#)

From the blog 

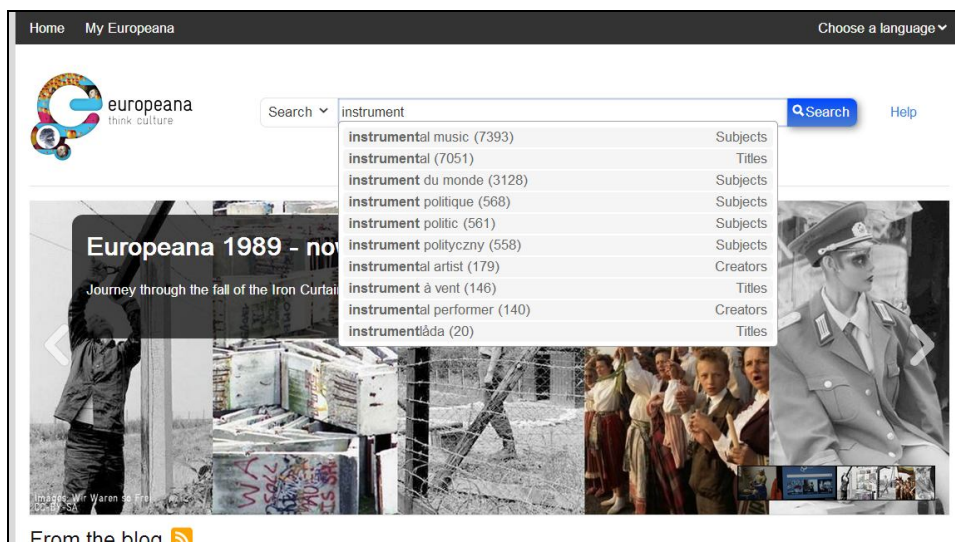


2.8.2. Cerques

- **Motor de Cerca simple: exemple d'opcions de cerca per camps**



- **Motor de cerca simple: exemple d'auto-realització**



- **Prototipus de cerca semàntica: exemple cerca terme "Spain"**

This is a [research prototype](#) of Europeana's semantic search engine.
Enter a search term, for example: [Egypt](#), [Rembrandt](#), [window](#).

Spain

artifact [view all 4 results](#)

- Grim Spain**
prent Goulding, Frederick
- Abriss der Festung Gulich von Sp...**
prent Anonymous
- Afternoon, Sahagún, Spain**
prent Bone, Muirhead (Sir)

concept [view all 9 results](#)

- Spain**
Europe
- capital of Spain**
Espana
- castle in Spain**
dream [castle in Spain](#)

location

- Kingdom of Spain**
inhabited region

person [view all 17 results](#)

- Spaink, H.**
- Spaink, Frieda**
- Spaink, Herman**
HCBK Stroom

Louvre
11,327 artworks

home disclaimer datacloud acknowledgments

W3C RDF

2.8.3. Registres

- Tipus de registre vídeo

Search Subjects environment and nature Search Help

Return to search results < Previous Next >

DELTEBRE: CLOSING CEREMONY OF BIRD DAY

Alternative Title: DELTEBRE: CLAUSURA DEL DIA DEL OCELLS

Description: Reportatge sobre la campanya "Dia dels ocells" Europa 87" que s'enmarca dins de l'any europeu del medi ambient.; Report on "Europe '87 Bird Day" as part of the 1987 European Year of the Environment.

Contributor: Jordi Ferrerons (PRESENTER)

Geographic coverage: Katalonien;
<http://sws.geonames.org/3336901/>

Date: 1987; <http://semium.org/time/1987>

Publication date: 23/05/1987

Type: Video; News;
<http://www.eionet.europa.eu/gemet/concept/8964>

Format: 00:01:34; 4:3; Colour; Stereo

Subject: Environmental Issues; Natural World; Conservation; Environment and Nature;
<http://www.eionet.europa.eu/gemet/concept/1700>

Identifier: [EUS_F7B4E582F8324A8C887CAE5850C75270](https://nbn-resolving.org/EUS_F7B4E582F8324A8C887CAE5850C75270)

Is part of: TELENOTÍCIES CAP DE SETMANA VESPRE: WEEKEND EVENING NEWS

Language: Catalan; Valencian

Rights: Els continguts de la pàgina de TVC estan protegits i subjectes a les lleis de propietat intel·lectual i industrial. Queda prohibida l'explotació, reproducció, distribució, transformació, execució pública, comunicació pública i qualsevol altra forma d'explotació dels continguts, sense autorització de TVC. La persona que accedeixi als continguts només els pot visualitzar per a finalitats personals i educatives no comercials. Per a més informació podeu adreçar-vos al Servei de còpies i venda d'imatges d'arxiu, imatges@tv3.cat. The content of the pages submitted by TVC are protected by and subject to TVC copyright and industrial property rights. The exploitation, reproduction, distribution, transformation, public execution, public communication and any other form of exploitation of the content is expressly prohibited without authorization from TVC. Access to the content is for the sole purpose of personal and educational non-commercial viewing. For further information, please contact the Image Copies and Sales Archive Service at: imatges@tv3.cat

Publisher: TVC

Data provider: TVC

Provider: EUScreen Project

Providing country: Spain

[Auto-generated tags](#)

Search also for:

- Title**
DELTEBRE: CLOSING CEREMONY OF BIRD DAY (1)
DELTEBRE: CLAUSURA DEL DIA DEL OCELLS (1)
- Who**
Jordi Ferrerons (PRESENTER) (20)
- What**
Video (67522)
News (11291)
Environmental Issues (324)
Natural World (2004)
Conservation (4191)
Environment and Nature (1428)
- Provider**
TVC (1882)
EUScreen Project (21065)

View item at TVC

Share

Cite on Wikipedia

Translate details

Select language

Powered by Microsoft Translator

- Registre d'artefacte: cerca per “Gioconda” o “Mona Lisa”


- Registre bibliogràfic simple del portatretrat de la Mona Lisa

Property	Value
Creator	DI SER PIERO DA VINCI ; VINCI Léonard de (dit) LEONARDO; Léonard de Vinci
Date	Vers 1503 - 1506
Location	France; Ile-de-France; Paris; 1 e étage; Denon; Peintures; Salle 6; Salle de la Joconde
Material	Bois (peuplier)
Style/Period	Italie
Subject	accoué; assis; cours d'eau; de trois-quarts; femme; fond de paysage; loggia; montagne; plaine; pont; portrait; rocher; route; à mi-corps; peintures; Italien; Mona Lisa
Technique	bois; peinture à l'huile
Title	PORTRAIT DE MONA LISA (1479-1528) ; DITE LA JOCONDE; <i>Portrait de Lisa Gherardini</i> ; épouse de Francesco del Giocondo, dite <i>Monna Lisa</i> , la <i>Gioconda</i> ou la <i>Joconde</i>
Type	peinture; tableau; DITE LA JOCONDE; Département des Peintures; LEONARDO DI SER PIERO DA VINCI; VINCI Léonard de (dit) ; <i>PORTRAIT DE MONA LISA (1479-1528)</i>

- Registre bibliogràfic complet del portatretrat de la Mona Lisa amb exemple de dades relacionades per tipus d'objecte, “tableau”:

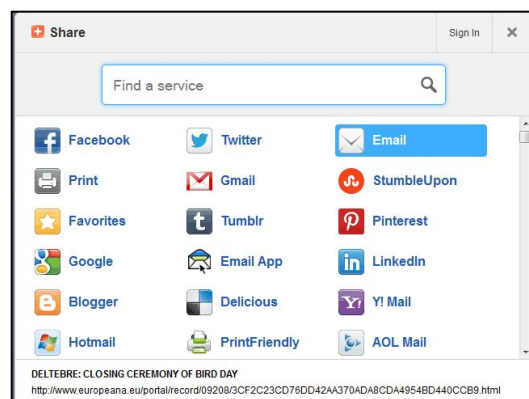
Property	Value	Source
Creator	<ul style="list-style-type: none"> DI SER PIERO DA VINCI LEONARD DE (d4) LEONARDO 	<ul style="list-style-type: none"> loane_joconde_works_creators.rdf
Object type	<ul style="list-style-type: none"> tableau 	<ul style="list-style-type: none"> loane_joconde_works_annotations.rdf
Domain	<ul style="list-style-type: none"> peinture 	<ul style="list-style-type: none"> loane_joconde_works_annotations.rdf
Style/Period	<ul style="list-style-type: none"> Italie 	<ul style="list-style-type: none"> loane_joconde_works_places.rdf
Current location	<ul style="list-style-type: none"> France Île-de-France Paris 	<ul style="list-style-type: none"> loane_joconde_works_places.rdf
Depicted subject	<ul style="list-style-type: none"> accoudé assis cours d'eau de trois-quarts femme fond de paysage loggia montagne plaine pont portrait rocher route à mi-corps italien Mona Lisa 	<ul style="list-style-type: none"> loane_joconde_works_annotations.rdf loane_joconde_works_literals.rdf
Technique	<ul style="list-style-type: none"> bois peinture à l'huile 	<ul style="list-style-type: none"> loane_joconde_works_annotations.rdf loane_joconde_works_literals.rdf
TICO	<ul style="list-style-type: none"> DITE LA JOCONDE LEONARDO DI SER PIERO DA VINCI VINCI Leonard de (d4) : PORTRAIT DE MONA 	

Property	Value	Source
type	<ul style="list-style-type: none"> Concept 	<ul style="list-style-type: none"> joconde.rdf
has broader	<ul style="list-style-type: none"> peinture 	<ul style="list-style-type: none"> joconde.rdf
additional note	<ul style="list-style-type: none"> Différent de tableau au sens de "support d'enseignement" (tableau noir). Cette acception du terme n'est actuellement illustrée par aucune notice. 	<ul style="list-style-type: none"> joconde.rdf
has exact match	<ul style="list-style-type: none"> tableaux (sculpture) tableaux (effemenceux) 	<ul style="list-style-type: none"> loane_joconde_enrich_map_match.asf.rdf
is a schema	<ul style="list-style-type: none"> http://culture.multimedia.nina.joconde/ 	<ul style="list-style-type: none"> joconde.rdf
preferred label	<ul style="list-style-type: none"> tableau 	<ul style="list-style-type: none"> joconde.rdf
has related	<ul style="list-style-type: none"> art de la peinture dessin peintre peinture 	<ul style="list-style-type: none"> joconde.rdf
scope note	<ul style="list-style-type: none"> Oeuvre d'art, résultat de l'activité de peinture 	<ul style="list-style-type: none"> joconde.rdf

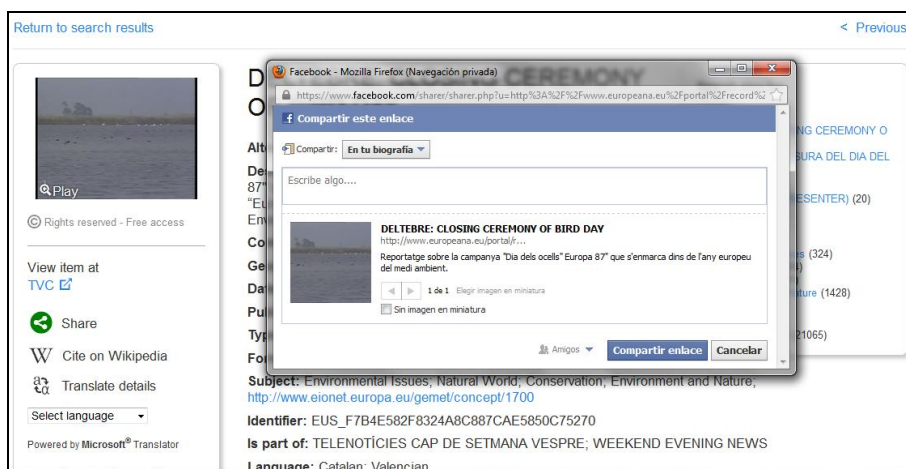
Property	Subject	Source
Object type	 <ul style="list-style-type: none"> MISE AU TOMBEAU L'HOMME AU VERRE DE VINI 	<ul style="list-style-type: none"> loane_joconde_works_annotations.rdf

2.8.4. Extracció d'Informació

L'extracció de la informació no s'implementa en la cerca semàntica. El registre d'informació només s'extrau amb l'opció d'impressió, en la resta d'opcions només s'exporta l'enllaç del registre seleccionat. Com a opcions d'exportació hi ha algunes d'habituals com l'opció d'impressió i enviament via e-mail, i la resta d'opcions permeten compartir el registre amb un gran nombre de xarxes socials:



- **Opció d'extracció d'informació a través de xarxes socials: exemple de Facebook**



Annex 2.9. Qüestionaris d'avaluació de sistemes d'informació de qualitat: anàlisi de dimensions d'IQ

2.9.1. Qüestionari Intute (2006 – juliol 2011)

Aquesta passarel·la tanca al 2011, i la darrera informació actualitzada en aquesta avaluació és en data de juliol 2011. No obstant això, aquest fet no afecta a l'estudi de comparació i avaluació de sistemes d'informació distribuïts de qualitat, ja que els sistemes s'avaluen des del punt de vista de la seva arquitectura de la informació i els seus paràmetres d'informació de qualitat, i no pas des de la seva col·lecció, que com ja s'ha explicat anteriorment, la col·lecció no potser comparable plenament entre els sistemes ja que la cobertura temàtica dels sistemes és molt diferent. En el cas d'Intute, exceptuant els serveis, la informació continguda a Intute està arxivada i les col·leccions anteriors al juliol del 2011 estan accessibles.

Atributs ^{76,77}	Escala de valor d'IQ		
	Avaluació Qualitativa/Quantitativa	Puntuació	Total
1. Dimensió	Descripció/Qüestions ⁷⁸	Si=1; Nivell	

⁷⁶ Naumann, F., Rolker, C., 2000. Assessment methods for information quality criteria, IQ. Citeseer, pp. 148-162.

⁷⁷ 1999. A list of quality selection criteria: a reference tool for Internet subject gateways, Selection criteria for quality controlled information gateways UKOLN Metadata Group. Url: <http://www.ukoln.ac.uk/metadata/desire/quality/report-2.html>

Sistema (Interrogació)		No=0	IQ= 1,3,5]	
Autoria (1,2,3)	Es coneix l'autoria i l'origen del recurs? Si Altres elements addicionals: - Apareixen dades de contacte (adreça, telèfon, etc.)? No - Hi ha opció a suggeriments, comentaris, valoració, etc.? Si - Apareixen cites/referències de l'autor? No (però no es rellevant) - Altres? NO	1 0 1 0	50%	3
Comprensibilitat (1)	Grau en que la informació és compresa per l'usuari (per exemple, multilingüisme): alt - La informació és completa (de la informació recuperada)? Si - El títol és informatiu? Si - Es dona un resum? Si	1 1 1 1	100%	5
Estructura / Consistència en la representació de la informació (1)	Grau de representació consistent en què l'estructura de la informació s'ajusta a la d'altres fonts, per exemple nivell d'integritat, homogeneïtat, consistència semàntica, compatibilitat, etc. (IQ) - Visualització immediata dels resultats sense passos intermedis? Si - Qualitat de la referència bibliogràfica (ús d'estàndards)? Si - Possibilitat de selecció del criteri d'ordenació dels resultats? Si (alfabètic, rellevància i per novetat) - Possibilitat de selecció dels registres	1 1 1 1 1	85%	5

⁷⁸ Nota: algunes respostes poden ser respostes amb si/no i d'altres amb la descriptiva qualitativa (DQ). Les preguntes específiques són adaptades del model DESIRE.

	<p>de sortida (impressió, gravació o enviament via e-mail)?Si (versió subscripció/personalització=MyIntute)</p> <p>- Possibilitat de limitar o augmentar el nombre de resultats de la cerca en cada pantalla?Si (permet limitar)</p> <p>- Dona com a resultat els termes de cerca? Si</p> <p>- Opcions extra de visualització dels resultats (per exemple: agrupacions o índex KWIC)?No</p>	0		
Integritat (1): integritat de la informació, del lloc web i del sistema	- La informació proporcionada pel sistema és actual i està actualitzada?Si	1	75%	5
		1		
	- Existeix un manteniment adequat dels continguts informatius?Si	1		
	S'amplia informació sobre el recurs (per exemple, recursos/enllaços relacionats, biblioteca o base de dades d'oposició, etc.)?Si (enllaç a al font original)	0		
		1		
		1		
	- Hi ha enllaços morts?Si	0		
	- Hi ha enllaços a llocs on s'ha mogut o hi ha redirecció a un nou lloc web? Si	1		
		0		
	- Les pàgines web del sistema tenen la data amb la última data de revisió? Si	1		
		1		
	- Apareix el número de versió dels recursos?No (tot i que no és un element negatiu)	1		
- Apareix la data de la darrera actualització del recurs? Si				
- Hi ha una descripció de les freqüències de l'actualització dels				

	<p>recursos?Si/No</p> <p>- L'organització/persona que allotja el recurs sembla tenir el compromís amb el manteniment i l'estabilitat dels recursos?Si/No</p> <p>- El lloc s'actualitza/manté amb freqüència / manteniment?Si</p> <p>-Fiabilitat del sistema: Si</p>			
<p>Disponibilitat / Accessibilitat (1,2,3)</p> <p>[nivell d'accessibilitat i recuperabilitat]</p>	<p>-Ofereix serveis per facilitar l'accés a informació disponible en diferents formats, suports, etc.?Si</p> <p>-Té serveis que faciliten la conversió/extracció de la informació amb formats diferents? No</p> <p>-Ofereix serveis per a que un usuari pugui accedir a la informació de forma indirecta? (serveis d'alerta, butlletins, guies temàtiques, etc.) Si</p> <p>- Informació sobre existència de limitacions o condicions d'accés (per exemple, data límit d'accés, si es requereix quota, etc.?Si</p> <p>- Ofereixen servei d'Obtenció de Documents?No</p> <p>- Existència de missatges clars quan la connexió no es possible?Si</p>	<p>1</p> <p>0</p> <p>1</p> <p>1</p> <p>0</p> <p>1</p>	67%	3
<p>Seguretat (1)</p> <p>[Privacitat i seguretat d'accés]</p>	<p>Grau en què la informació es transmet de forma privada a l'usuari. IQ= 5 (opció de personalització, cerca i emmagatzemament privat de l'usuari, amb MYIntute)</p> <p>Hi ha polítiques de privacitat, protocols d'intercanvi d'informació, etc.)? Si</p>	<p>1</p> <p>1</p>	100%	5
Valoració				

Màxima (30)	Puntuació total paràmetres de RI	26 (87% IQ) = 5		
2. Dimensió Serveis (Tipologia)	Descripció/Qüestions	[Si=1; No=0]; [Nivell d'IQ=1,3,5]		
Actualització (2,3)	<ul style="list-style-type: none"> - Serveis actualitzats? No tots - Apareix data d'actualització? Si - Informació sobre noves versions/serveis/aplicacions? Si - Tutorials actualitzats? Si - etc. 	0	75%	5
		1		
		1		
		1		
Autoria (2,3) [informació referent als serveis i la informació que aquests proporcionen en cada servei]	<ul style="list-style-type: none"> - És la font atribuïble a un autor o una organització de bona reputació? Si - Es coneix l'origen del recurs? Si - Apareix el nom complet de l'autor? Si - Es coneix l'autor/institució o editor del recurs? Si (exemple, Guies temàtiques)⁷⁹ - Hi ha una URL a un servidor conegut (universitat, centre de recerca, etc.) Si - L'autor ja ha estat citat abans? No es coneixen dades El recurs proporciona altres cites/referències de l'autor? No (però no es rellevant) -Hi ha disponible informació sobre l'autor (CV., adreça, e-mail, etc.)? No (però no es rellevant) 	1	57%	3
		1		
		1		
		1		
		0		
		0		
		0		
Completesa	-Quocient del nombre de serveis			

⁷⁹ Exemple de Servei de Guies temàtiques: Biological Sciences. Subject guides. Intute, March 2008. Url: <http://www.intute.ac.uk/supportdocs/biologicalsciences.pdf>

(2,3), [Cobertura, Abast, Granularitat, Densitat i Extensió]	inclosos i el nombre de serveis bàsics existents (cerques, alertes, butlletins, etc.) ⁸⁰ : 16% són serveis bàsics	1	75%	5
	-Nivell de descripció dels recursos inclosos en els serveis i dels propis serveis? Si	1		
	- Quina és la informació bàsica que ha de contenir el servei (metadades)? Mitjà-Alt, inclou metadades mínimes en tots els serveis, però no els 15 elements DC ⁸¹ (3)	0		
	-Nivell d'especificitat, els serveis s'adeqüen a l'audiència, etc.? Si	1		
Disponibilitat / Accessibilitat (1,2,3) [Accessibilitat, recuperabilitat, etc.]	El lloc conté informació que està fàcilment disponible? Si	1	100%	5
	Està disponible a través d'altres fonts d'informació? Si	1		
	El recurs facilitar l'accés a informació de diferents formats? i aquests són accessibles per tots els usuaris? Si (per exemple cas RSS, incloure aplicació per instal·lar lector de RSS).	1		
		1		
	El recurs de proporcionar accés a la informació que els usuaris poden no ser capaços de trobar / accedir a les biblioteques? No (positiu)	1		
Rellevància (1, 2, 3) [precisió, redundància, aplicabilitat i utilitat]	Grau en què la informació recuperada compleix les necessitats dels usuaris. Quant a serveis es refereix, s'aplica a la perspectiva de l'usuari, sobretot referent al serveis d'informació que puguin proporcionar els sistemes (serveis d'alerta, serveis de cerca, butlletins, guies temàtiques, etc.)? Si	1	100%	5

⁸⁰ Serveis bàsics o mínims: s'interpreten serveis mínims els motors de cerca i navegació, serveis de notícies i arxiu (sigui quin sigui el seu mitjà o suport), serveis que permetin guardar continguts/cerques, exportar resultats, etc., i servei d'alertes sobre el contingut (independentment del mitjà i suport: e-mail, RSS, etc.), i serveis com "About us" (informació sobre la responsabilitat del sistema) o "Help" (ajuda).

⁸¹ Dublin Core Metadata Initiative (DC). Url: <http://dublincore.org/>

	Aquest atribut potser difícil d'avaluar, perquè alguns serveis són privats o de pagament. En aquest cas els serveis de subscripció són gratuïts, l'adequació i utilitat dels serveis a l'audiència és alt, i hi ha molts serveis addicionals (54%). Mancança però d'un servei d'alt valor afegit per al món de la recerca com és l'exportació dels resultats en formats bibliogràfics (.enl, .ris, etc.)	1		
Precisió (1,2, 3)	Aquest atribut és limita als serveis d'informació/documentació (obtenció de documents, serveis d'alertes, butlletins de notícies, guies temàtiques, etc.), cerca i navegació. Es poden incloure les mateixes interrogacions que en la dimensió del sistema? No (són serveis més simples) - Informació obtinguda a través del servei és precisa i adequada? Si - La informació es pot verificar amb exactitud? Si - Correctesa de la informació rebuda (sense faltes ortogràfiques, gramaticals o tipogràfiques), etc.?	0 1 1 1	80%	5
Valoració Màxima (30)	Puntuació total paràmetres serveis	28 (93% IQ) = 5		
3. Dimensió Col·lecció	Descripció/Qüestions	[Si=1; No=0]; [Nivell d'IQ=1,3,5]		
Autoria (2,3)	- És la font/recurs atribuïble a un autor o una organització de bona reputació? Si	1 1 1		

	<ul style="list-style-type: none"> - Es coneix l'origen del recurs? Si - Apareix el nom complet de l'autor?Si - Es coneix l'autor/institució o editor del recurs? Si (normalment) - Hi ha una URL a un servidor conegut (universitat, centre de recerca, etc.) Si - L'autor ja ha estat citat abans?Es desconeix - El recurs proporciona altres cites/referències de l'autor?No -Hi ha disponible informació sobre l'autor (CV., adreça, e-mail, etc.)? No 	<p>1</p> <hr/> <p>1</p> <hr/> <p>0</p> <hr/> <p>0</p> <hr/> <p>0</p>	63%	3
Actualització (2,3)	<ul style="list-style-type: none"> - Nivell i/o polítiques d'actualització de la col·lecció. Són altes, formen part de la seva raó de ser (vegeu capítol Evolució i IQ). IQ=5 - Hi ha una declaració sobre la freqüència d'actualització?Si - Si es tracta d'un recurs estàtic (no actualitzat) és la informació útil per a l'usuari/audiència?Si - Com sensible al temps és la informació continguda, i com es relaciona amb la freqüència d'actualització? En general la informació no es sensible amb el temps i en la majoria dels casos no afecta l'actualització (només per a la revisió de enllaços trencats i aquests són actualitzats) 	<p>1</p> <hr/> <p>1</p> <hr/> <p>1</p> <hr/> <p>1</p>	100%	5
Completesa (2,3), [Cobertura, Abast,	Quocient del nombre d'elements inclosos en la col·lecció i el nombre d'elements del món real. No es pot respondre la equació de forma objectiva.	<p>0</p> <hr/> <p>1</p>	90%	5

<p>Granularitat, Densitat i Extensió], aquí s'inclou també l'ús de Metadades (ús 15 elements Dublin Core) o altres normes de descripció</p>	- Explicació de criteris de selecció i incorporació de noves fonts/documents? Si	1		
	- Nivell de descripció dels recursos (especificitat dels recursos, etc.): alt	1		
	- Ús de metadades o normatives estàndards per a la descripció dels recursos? Si	1		
	- Quina és la quantitat mínima d'informació necessària per crear una descripció dels recursos al catàleg? O quina és la informació bàsica que ha de contenir un recurs per a ser seleccionat? (ús de normatives, (AACR2 ⁸² , ISO 690-2 ⁸³ , Dublin Core, etc.) Nivell mínim en tots els casos, nivell complet en alguns (AACR2). IQ=3	1		
	- A quin nivell es poden seleccionar els recursos catalogats? A tots	1		
	- Ús de descriptors/termes específics per a la indexació/Navegació? Si (ús de BT ⁸⁴ i NT)	1		
	- Cobertura documental: quin tipus de fonts es recullen? Tot tipus	1		
	- Cobertura temàtica: apareixen els descriptors indexats en el document de referència de la col·lecció? Si	1		
	- Cobertura temporal: apareix la data d'inici de la referència?	1		
- Quin és el valor relatiu o el recurs en comparació amb altres disponibles en aquest tema? Alt (polítiques fortes de	1			

⁸² AACR2: Anglo-American Cataloguing Rules, 2nd ed. Url: <http://www.aacr2.org/>

⁸³ ISO 690-2: Referencias bibliográficas: documentos electrónicos y sus partes. Url: <http://biblioteca.ucv.cl/site/servicios/documentos/ISO690-2.pdf?cp=5>

⁸⁴ BT= Broader term o terme genèric (llenguatge emprat en vocabularis controlats); NT= Narrower Term o terme específic d'un terme BT.

	selecció, tria de documents)			
Disponibilitat/ Accessibilitat (1,2,3)	-Què hi ha disponible fora de la col·lecció? No es pot respondre objectivament	0	75%	5
[Nivell de fiabilitat]	- Inclusió del text complet? Si (si està disponible en la font d'origen)	1		
	- Disponibilitat d'enllaços/referències al text complet? Si	1		
	- Integració de diferents formats en la col·lecció? Si (.txt,.html,.pdf., etc.)	1		
		1		
Precisió (1,2, 3)	Quocient entre el nombre de valors correctes recuperats pel sistema i el nombre total de valors continguts en el sistema. (Valor no determinant, perquè no apareixen dades del total de documents recollits en la col·lecció, ni subcol·leccions temàtiques). Nivell IQ baix = 1	0	80%	5
	-La informació es pot verificar amb exactitud? Si	1		
	-El recurs, mostra cites bibliogràfiques o referències per confirmar la informació? Si (s'inclouen referències)	1		
	-La presentació de la informació en la RI és correcta (sense faltes ortogràfiques/gramaticals)? Si	1		
	-Hi ha prevalença d'errors ortogràfics? No	1		
		1		

Rellevància (1, 2,3) [precisió, redundància, aplicabilitat i utilitat]	Atribut subjectiu, no valorable col·lectivament perquè és l'usuari final qui valora la col·lecció segons les seves necessitats. A nivell de precisió, aplicabilitat i utilitat, si es pot dir que els recursos disponibles són científics i tècnics, per tant, corresponen a l'audiència, i a més a més són avaluats sota criteris d'IQ abans de ser indexats.	1	71%	5
	- Hi ha accés de navegació per matèries (tesaurus, etc.)? Si	0		
	- Hi ha possibilitat de cercar termes genèrics i específics (navegabilitat per estructura jeràrquica)? No (només genèrics)	1		
	- Fiabilitat de la cerca per classificació/descriptors de termes específics? Si	1		
	- Consistència dels termes d'indexació (un concepte/tema ha d'aparèixer d'una forma única)? Si	0		
	- Suggestions de cerca per termes/conceptes alternatius? No	1		
	- Capacitat per discriminació de termes? Si (ús operador NOT)			
Valoració Màxima (30)	Puntuació total paràmetres col·lecció	28 (93% IQ) = 5		
Total (90)	Valoració total	81 (91% IQ) = 5		

2.9.2. Qüestionari Vascoda (2003-2011)

Aquesta passarel·la tanca al 2011, i la darrera informació actualitzada en aquesta avaluació és en data de gener 2011. No obstant això, aquest fet no afecta a l'estudi de comparació i avaluació de sistemes d'informació distribuïts de qualitat, ja que els sistemes s'avaluen des del punt de vista de la seva arquitectura de la informació i els seus paràmetres d'informació de qualitat, i no pas des de la seva col·lecció, que com ja s'ha explicat anteriorment, la col·lecció no potser comparable plenament entre els sistemes ja que la cobertura temàtica dels sistemes és molt diferent.

Atributs ^{85,86}	Escala de valor d'IQ			
	Avaluació Qualitativa/Quantitativa	Puntuació		Total
1. Dimensió Sistema (Interrogació)	Descripció/Qüestions⁸⁷	Si=1; No=0	Nivell IQ= 1,3,5]	
Autoria (1,2,3)	Es coneix l'autoria i l'origen del recurs? Si Altres elements addicionals: - Apareixen dades de contacte (adreça, telèfon, etc.)? No - Hi ha opció a suggeriments, comentaris, valoració, etc.? No - Apareixen cites/referències de l'autor? No (però no es rellevant) - Altres? No	1 0 0 0	25 %	1
Comprensibilitat (1)	Grau en que la informació és compresa per l'usuari (per exemple, multilingüisme): major cobertura amb alemany i no amb anglès i és un recurs global - La informació és complerta (en	0	25 %	1

⁸⁵ Naumann, F., Rolker, C., 2000. Assessment methods for information quality criteria, IQ. Citeseer, pp. 148-162.

⁸⁶ 1999. A list of quality selection criteria: a reference tool for Internet subject gateways, Selection criteria for quality controlled information gateways UKOLN Metadata Group. Url: <http://www.ukoln.ac.uk/metadata/desire/quality/report-2.html>

⁸⁷ Nota: algunes respostes poden ser respostes amb si/no i d'altres amb la descriptiva qualitativa (DQ). Les preguntes específiques són adaptades del model DESIRE.

	<p>algunes cerques no es troba informació que si s'ha trobat amb Intute per la mateixa cerca, la cobertura de la col·lecció és inferior)?No</p> <p>- El títol és informatiu? Si</p> <p>- Es dona un resum? Si, però no en la primera pàgina del registre recuperat, cal seleccionar el recurs per veure el resum.</p>	0		
		1		
		0		
Estructura / Consistència en la representació de la informació (1)	<p>Grau de representació consistent en què l'estructura de la informació s'ajusta a la d'altres fonts, per exemple nivell d'integritat, homogeneïtat, consistència semàntica, compatibilitat, etc. (IQ)</p> <p>-Visualització immediata dels resultats sense passos intermedis?Si</p> <p>-Qualitat de la referència bibliogràfica (ús d'estàndards)?Si</p> <p>- Possibilitat de selecció del criteri d'ordenació dels resultats?Si (rellevància i any de publicació)</p> <p>- Possibilitat de selecció dels registres de sortida (impressió, gravació o enviament via e-mail)?Si (versió subscripció)</p> <p>- Possibilitat de limitar o augmentar el nombre de resultats de la cerca en cada pantalla?Si (permet limitar)</p> <p>- Dona com a resultat els termes de cerca? Si però no s'assenyalen)</p> <p>- Opcions extra de visualització dels resultats (per exemple: agrupacions o índex KWIC)?No</p>	1	71%	5
		1		
		1		
		1		
		0		
		1		
		0		
Integritat (1):	- La informació proporcionada pel			

integritat de la informació, del lloc web i del sistema	sistema és actual i està actualitzada?Si	1	72%	5
	- Existeix un manteniment adequat dels continguts informatius?Si	1		
	S'amplia informació sobre el recurs (per exemple, recursos/enllaços relacionats, biblioteca o base de dades dipositària, etc.)?Si però mínim, enllaç a al font original)	1		
	- Hi ha enllaços morts?Si	0		
	- Hi ha enllaços a llocs on s'ha mogut o hi ha redireccionament a un nou lloc web? Si	1		
	- Les pàgines web del sistema tenen la data amb la última data de revisió? Si	1		
	- Apareix el número de versió dels recursos?No (tot i que no és un element negatiu)	0		
	- Apareix la data de la darrera actualització del recurs? Si	0		
	- Hi ha una descripció de les freqüències de l'actualització dels recursos?No	1		
	- L'organització/persona que allotja el recurs sembla tenir el compromís amb el manteniment i l'estabilitat dels recursos?Si	1		
- El lloc s'actualitza/manté amb freqüència / manteniment?Si	1			
-Fiabilitat del sistema: Si	1			
Disponibilitat / Accessibilitat (1,2,3)	-Ofereix serveis per facilitar l'accés a informació disponible en diferents formats, suports, etc.?Si	1	83%	5
	-Té serveis que faciliten la			

[nivell d'accessibilitat i recuperabilitat]	conversió/extracció de la informació amb formats diferents? No	0		
	-Ofereix serveis per a que un usuari pugui accedir a la informació de forma indirecta? Si (serveis d'alerta, butlletins, via e-mail, etc.)	1		
	- Informació sobre existència de limitacions o condicions d'accés (per exemple, data límit d'accés, si es requereix quota, etc.? Si	1		
	- Ofereixen servei d'Obtenció de Documents?Si	1		
	- Existència de missatges clars quan la connexió no es possible?Si	1		
Seguretat (1) [Privacitat i seguretat d'accés]	Grau en què la informació es transmet de forma privada a l'usuari. IQ= 5 (opció de personalització, cerca i emmagatzemament privat de l'usuari, amb MYVascosa) Hi ha polítiques de privacitat, protocols d'intercanvi d'informació, etc.)? Si	1	100%	5
		1		
Valoració Màxima (30)	Puntuació total paràmetres de RI	22 (73% IQ) = 5		
2. Dimensió Serveis (Tipologia)	Descripció/Qüestions	[Si=1; No=0]; [Nivell d'IQ=1,3,5]		
Actualització (2,3)	- Serveis actualitzats? Si	1	100%	5
	- Apareix data d'actualització?Si	1		
	- Informació sobre noves versions/serveis/aplicacions?Si	1		
	- Tutorials actualitzats?Si	1		

	- etc.			
<p>Autoria (2,3)</p> <p>[informació referent als serveis i la informació que aquests proporcionen en cada servei]</p>	- És la font atribuïble a un autor o una organització de bona reputació? Si	1	75%	5
	- Es coneix l'origen del recurs? Si			
	- Apareix el nom complet de l'autor? Si	1		
	- Es coneix l'autor/institució o editor del recurs? Si (normalment si, però des del sistema, no l'usuari extern)	1		
	- Hi ha una URL a un servidor conegut (universitat, centre de recerca, etc.) Si	1		
	- L'autor ja ha estat citat abans? No es coneixen dades, des del sistema	1		
	-El recurs proporciona altres cites/referències de l'autor? Si	0		
	-Hi ha disponible informació sobre l'autor (CV., adreça, e-mail, etc.)? No (però no es rellevant)	1		
		0		
<p>Completesa (2,3), [Cobertura, Abast, Granularitat, Densitat i Extensió]</p>	-Quocient del nombre de serveis inclosos i el nombre de serveis bàsics existents (cerques, alertes, butlletins, etc.) ⁸⁸ : 99% són serveis bàsics, n'hi ha pocs d'addicionals (en desenvolupament, IQ=3)	0	75%	5
	-Nivell de descripció dels recursos inclosos en els serveis i dels propis serveis? Si	1		
	- Quina és la informació bàsica que ha			

⁸⁸ Serveis bàsics o mínims: s'interpreten serveis mínims els motors de cerca i navegació, serveis de notícies i arxiu (sigui quin sigui el seu mitjà o suport), serveis que permetin guardar continguts/cerques, exportar resultats, etc., i servei d'alertes sobre el contingut (independentment del mitjà i suport: e-mail, RSS, etc.)

	de contenir el servei (metadades)? Alt, inclou metadades Dublin Core (5)	1		
	-Nivell d'especificitat, els serveis s'adeqüen a l'audiència, etc.?	1		
Disponibilitat / Accessibilitat (1,2,3) [Accessibilitat, recuperabilitat, etc.]	El lloc conté informació que està fàcilment disponible? Si	1	100%	5
	Està disponible a través d'altres fonts d'informació? Si	1		
	El recurs facilitar l'accés a informació de diferents formats? Si, i aquests són accessibles per tots els usuaris? Si (per exemple cas RSS, incloure aplicació per instal·lar lector de RSS).	1		
	El recurs de proporcionar accés a la informació que els usuaris poden no ser capaços de trobar / accedir a les biblioteques? No (positiu)	1		
		1		
Rellevància (1, 2, 3) [precisió, redundància, aplicabilitat i utilitat]	Grau en què la informació recuperada compleix les necessitats dels usuaris. Quant a serveis es refereix, s'aplica a la perspectiva de l'usuari, sobretot referent al serveis d'informació que puguin proporcionar els sistemes (serveis d'alerta, serveis de cerca, butlletins, guies temàtiques, etc.)? Si Aquest atribut potser difícil d'avaluar, perquè alguns serveis són privats o de pagament. En aquest cas d'estudi la majoria de serveis són bàsics, alguns serveis són gratuïts (sistemes de cerca, canals RSS, etc.), però d'altres són privats i no es poden avaluar). Els resultats només es poden exportar via e-mail, i manca d'un servei d'alt valor afegit per al món de la recerca com és l'exportació dels resultats en formats bibliogràfics	1	50%	3

	(.enl, .ris, etc.), etc.	0		
Precisió (1,2, 3)	<p>Aquest atribut és limita als serveis d'informació/documentació (obtenció de documents, serveis d'alertes, butlletins de notícies, guies temàtiques, etc.), cerca i navegació. Es poden incloure les mateixes interrogacions que en la dimensió del sistema? No (són serveis més simples)</p> <p>- Informació obtinguda a través del servei és precisa i adequada? Si</p> <p>- La informació es pot verificar amb exactitud? Si</p> <p>- Correctesa de la informació rebuda (sense faltes ortogràfiques, gramaticals o tipogràfiques), etc.? Si</p>	0	75%	5
		1		
		1		
		1		
Valoració Màxima (30)	Puntuació total paràmetres serveis	28 (93% IQ) = 5		
3. Dimensió Col·lecció	Descripció/Qüestions	[Si=1; No=0]; [Nivell d'IQ=1,3,5]		
Autoria (2,3)	<p>- És la font/recurs atribuïble a un autor o una organització de bona reputació? Si</p> <p>- Es coneix l'origen del recurs? Si</p> <p>- Apareix el nom complet de l'autor? Si</p> <p>- Es coneix l'autor/institució o editor del recurs? Si (normalment)</p> <p>- Hi ha una URL a un servidor conegut (universitat, centre de recerca, etc.) Si</p>	1	63%	3
		1		
		1		
		1		
		1		
		0		

	- L'autor ja ha estat citat abans?Es desconeix	0		
	- El recurs proporciona altres cites/referències de l'autor?No			
	-Hi ha disponible informació sobre l'autor (CV., adreça, e-mail, etc.)? No	0		
Actualització (2,3)	- Nivell i/o polítiques d'actualització de la col·lecció. Són altes, formen part de la seva raó de ser (vegeu capítol Evolució i IQ). IQ=5	1	75%	5
	- Hi ha una declaració sobre la freqüència d'actualització?Si	0		
	- Si es tracta d'un recurs estàtic (no actualitzat) és la informació útil per a l'usuari/audiència?Si	1		
	- Com sensible al temps és la informació continguda, i com es relaciona amb la freqüència d'actualització? En general la informació no es sensible amb el temps i en la majoria dels casos no afecta l'actualització (només per a la revisió de enllaços trencats i aquests són actualitzats)	1		
Completesa (2,3), [Cobertura, Abast, Granularitat, Densitat i Extensió], aquí s'inclou també l'ús de Metadades (ús 15 elements Dublin Core) o	-Quocient del nombre d'elements inclosos en la col·lecció i el nombre d'elements del món real. No es pot respondre la equació de forma objectiva, però es detecta poca cobertura en alguns temes (ciències marines)	0	72%	5
	- Explicació de criteris de selecció i incorporació de noves fonts/documentos?No	0		
	- Nivell de descripció dels recursos (especificitat dels recursos, etc.):			

altres normes de descripció	complet			
	- Ús de metadades o normatives estàndards per a la descripció dels recursos? Si	1		
	-Quina és la quantitat mínima d'informació necessària per crear una descripció dels recursos al catàleg? O quina és la informació bàsica que ha de contenir un recurs per a ser seleccionat? (ús de normatives, (AACR2 ⁸⁹ , ISO 690-2 ⁹⁰ , Dublin Core, etc.) Nivell complet en la majoria dels casos. IQ=5	1		
	-A quin nivell es poden seleccionar els recursos catalogats? A nivell de camps de descripció DC	1		
	-Ús de descriptors/termes específics per a la indexació/Navegació? Si (2 nivells)	1		
	- Cobertura documental: quin tipus de fonts es recullen? Tot tipus	1		
	- Cobertura temàtica: apareixen els descriptors indexats en el document de referència de la col·lecció? No (en la pantalla principal d'informació)	0		
	- Cobertura temporal: apareix la data d'inici de la referència?	1		
	- Quin és el valor relatiu o el recurs en comparació amb altres disponibles en aquest tema? Alt (polítiques fortes de selecció, tria de documents)	1		
Disponibilitat/ Accessibilitat	-Què hi ha disponible fora de la col·lecció? No es pot respondre	0	75%	

⁸⁹ AACR2: Anglo-American Cataloguing Rules, 2nd ed. Url: <http://www.aacr2.org/>

⁹⁰ ISO 690-2: Referencias bibliográficas: documentos electrónicos y sus partes. Url: <http://biblioteca.ucv.cl/site/servicios/documentos/ISO690-2.pdf?cp=5>

<p>(1,2,3)</p> <p>[Nivell de fiabilitat]</p>	<p>objectivament</p> <ul style="list-style-type: none"> - Inclusió del text complert? Si (si està disponible en la font d'origen) - Disponibilitat d'enllaços/referències al text complert? Si - Integració de diferents formats en la col·lecció? Si (.txt, .html, .pdf., etc.) 	<p>1</p> <p>1</p> <p>1</p>		<p>5</p>
<p>Precisió (2, 3)</p>	<p>Quocient entre el nombre de valors correctes recuperats pel sistema i el nombre total de valors continguts en el sistema. (No hi ha dades per avaluar-lo, s'assigna un valor negatiu per representar l'absència de l'element.</p> <ul style="list-style-type: none"> -La informació es pot verificar amb exactitud? Si -El recurs, mostra cites bibliogràfiques o referències per confirmar la informació? Si (s'inclouen referències) -La presentació de la informació en la RI és correcta (sense faltes ortogràfiques/gramaticals)? Si -Hi ha prevalença d'errors ortogràfics? No 	<p>0</p> <p>1</p> <p>1</p> <p>1</p> <p>1</p>	<p>80%</p>	<p>5</p>
<p>Rellevància (1, 2,3) [precisió, redundància, aplicabilitat i utilitat]</p>	<p>Atribut subjectiu, és l'usuari final qui valora la col·lecció segons les seves necessitats d'informació, i l'especificitat de la col·lecció. En l'exemple, aplicat, els resultats són nuls, per això la valoració és negativa. A nivell de precisió, aplicabilitat i utilitat dels recursos, la IQ és elevada, els recursos disponibles són científics i tècnics, per tant corresponen a l'audiència, i a més a més són avaluats sota criteris d'IQ abans de ser indexats.</p>	<p>1</p> <p>1</p>	<p>71%</p>	<p>5</p>

	<ul style="list-style-type: none"> - Hi ha accés de navegació per matèries (tesaurus, etc.)?Si - Hi ha possibilitat de cercar termes genèrics i específics (navegabilitat per estructura jeràrquica)?Si (2 nivells) - Fiabilitat de la cerca per classificació/descriptors de termes específics?No (en desenvolupament) - Consistència dels termes d'indexació (un concepte/tema ha d'aparèixer d'una forma única)?Si - Suggestions de cerca per termes/conceptes alternatius?No - Capacitat per discriminació de termes?Si (camps limitadors i possibilitat d'ús d'operador booleà NOT) 	1		
		0		
		1		
		0		
		1		
Valoració Màxima (30)	Puntuació total paràmetres col·lecció	28 (93 % IQ) = 5		
Total (90)	Valoració total	78 (87% IQ) = 5		

2.9.3. Qüestionari AustLII (1994 -)

Atributs ^{91,92}	Escala de valor d'IQ
----------------------------------	-----------------------------

⁹¹ Naumann, F., Rolker, C., 2000. Assessment methods for information quality criteria, IQ. Citeseer, pp. 148-162.

⁹² 1999. A list of quality selection criteria: a reference tool for Internet subject gateways, Selection criteria for quality controlled information gateways UKOLN Metadata Group. Url:

	Avaluació Qualitativa/Quantitativa	Puntuació		Total
1. Dimensió Sistema (Interrogació)	Descripció/Qüestions⁹³	Si=1; No=0	Nivell IQ= 1,3,5]	
Autoria (1,2,3)	<p>Es coneix l'autoria i l'origen del recurs? Si</p> <p>Altres elements addicionals:</p> <ul style="list-style-type: none"> - Apareixen dades de contacte (adreça, telèfon, etc.)? No de forma complerta, només hi ha un link a la web, però si es donen dades addicionals referent a la font principal d'informació que normalment és l'autoria (cobertura, numero de registres dipositat a la base de dades, copyright, etc.) - Hi ha opció a suggeriments, comentaris, valoració, etc.? No (s'hi està treballant). - Apareixen cites/referències de l'autor? Si (més que a l'autor, a la llei corresponent, apareix el nombre de cites que ha rebut cada publicació i s'ordenen els resultats per freqüència de citació, també hi ha un servei per cercar altres documents relacionats "Noteup" i notes i esmenes associades a la llei, opció "Note", i també i també informació disponible sobre la font que acull el registre que sol ser l'autoria: web, registres continguts en la base de dades, data de la darrera d'actualització de la base de dades, cobertura, copyright, etc.) 	1	50%	3
		0		
		0		
		1		
Comprensibilitat (1)	<p>Grau en que la informació és compresa per l'usuari.</p> <ul style="list-style-type: none"> - La informació és complerta (de la 	1	50%	3
0				

<http://www.ukoln.ac.uk/metadata/desire/quality/report-2.html>

⁹³ Nota: algunes respostes poden ser respostes amb si/no i d'altres amb la descriptiva qualitativa (IQ). Les preguntes específiques són adaptades del model DESIRE.

	informació recuperada)?No	1		
	- El títol és informatiu? Si	0		
	- Es dona un resum? Si	1		
Estructura / Consistència en la representació de la informació (1)	<p>Grau de representació consistent en què l'estructura de la informació s'ajusta a la d'altres fonts: nivell d'integritat, homogeneïtat, consistència semàntica, etc.</p> <p>-Visualització dels resultats sense passos intermedis?Si</p> <p>-Qualitat de la referència bibliogràfica (nivell mínim de metadades)? No</p> <p>- Possibilitat de selecció del criteri d'ordenació dels resultats?Si (rellevància, data, títol, base de dades i freqüència de citació)</p> <p>- Possibilitat de selecció dels registres de sortida (impressió, gravació o enviament via e-mail)? Si (només selecció manual, es valora negativament)</p> <p>- Possibilitat de limitar o augmentar el nombre de resultats de la cerca en cada pantalla?Si (grups de 10,20, 50 i 100 resultats)</p> <p>- Dona com a resultat els termes de cerca? No</p> <p>- Opcions extra de visualització dels resultats? Si (Veure tots els camps o només els camps rellevants, "Collapse multi-section": nom de la Llei/Reglament i el nom de la secció/clàusula rellevant).</p>	1	57%	3
		0		
		1		
		0		
		1		
		0		
		1		
		0		
		1		
Integritat (1):	- La informació proporcionada pel			

integritat de la informació, del lloc web i del sistema	sistema és actual i està actualitzada?Si	1	75%	5
	- Existeix un manteniment adequat dels continguts informatius?Si	1		
	-S'amplia informació sobre el recurs (per exemple, recursos/enllaços relacionats, biblioteca o base de dades dipositària, etc.)?Si (enllaç a la font original, i dades sobre aquesta: nº de registres, darrera actualització, i opció Noteup, cerca de materials associats, etc.)	1		
		0		
	- Hi ha enllaços morts?Si (però pocs)	0		
	- Hi ha enllaços a llocs on s'ha mogut o hi ha redireccionament a un nou lloc web?No	1		
	- Les pàgines web del sistema tenen la data amb la última data de revisió? Si	1		
	- Apareix el número de versió dels recursos? Si (element significatiu)	1		
	- Apareix la data de la darrera actualització del recurs? Si ⁹⁴	0		
	- Hi ha una descripció de les freqüències de l'actualització dels recursos?No	1		
	- L'organització/persona que allotja el recurs sembla tenir el compromís amb el manteniment i l'estabilitat dels recursos?Si	1		
	- El lloc s'actualitza/manté amb freqüència / manteniment?Si	1		
	-Fiabilitat del sistema: Si			

⁹⁴ Però no s'indica si aquesta data és de la legislació vigent o no.

<p>Disponibilitat / Accessibilitat (1,2,3)</p> <p>[nivell d'accessibilitat i recuperabilitat]</p>	<p>-Ofereix serveis per facilitar l'accés a informació disponible en diferents formats, suports, etc.? Si (però només a 1 format, RTF i ASCII, estàndards que utilitzen per la extracció dels resultats)</p> <p>-Té serveis que faciliten la conversió/extracció de la informació amb formats diferents? Si (opció "Download" dintre de cada registre recuperat en html, els registres en PDF, la seva descarrega és automàtica)</p> <p>-Ofereix serveis per a que un usuari pugui accedir a la informació de forma indirecta? (serveis d'alerta, butlletins, guies temàtiques, etc.) Si</p> <p>- Informació sobre existència de limitacions o condicions d'accés (per exemple, data límit d'accés, si es requereix quota, etc.? No (tota la informació està disponible, és informació pública/governamental)</p> <p>- Ofereixen servei d'Obtenció de Documents?No (no aplica, tota la informació és accessible i gratuïta)</p> <p>- Existència de missatges clars quan la connexió no es possible?Si</p>	<p>1</p> <p>1</p> <p>1</p> <p>1</p> <p>0</p> <p>1</p>	<p>83%</p>	<p>5</p>
<p>Seguretat (1)</p> <p>[Privacitat i seguretat d'accés]</p>	<p>-Grau en què la informació es transmet de forma privada a l'usuari. IQ= 3 (no hi ha opcions de personalització ni emmagatzemament privat per a l'usuari)</p>	<p>0</p>	<p>50%</p>	<p>3</p>

	-Hi ha polítiques de privacitat ⁹⁵ , protocols d'intercanvi d'informació, etc.)? Si	1		
Valoració Màxima (30)	Puntuació total paràmetres de RI	22 (73% IQ) = 5		
2. Dimensió Serveis (Tipologia)	Descripció/Qüestions	[Si=1; No=0]; [Nivell d'IQ=1,3,5]		
Actualització (2,3)	<ul style="list-style-type: none"> - Serveis actualitzats? No tots, però si la majoria - Apareix data d'actualització? Si - Informació sobre noves versions/serveis/aplicacions? Si - Tutorials actualitzats? Si - Bases de dades actualitzades? Si (sempre i quant no hagin tancat) 	0 1 1 1 1	80%	5
Autoria (1,2,3) [informació referent als serveis i la informació que aquests proporcionen en cada servei]	<ul style="list-style-type: none"> - Es coneix l'origen del servei? Si - Apareix el nom complet de l'autor? Si - És el servei/contingut del servei atribuïble a un autor o una organització de bona reputació? Si - Es coneix l'autor/institució o editor del contingut del servei? Si (la major part són fonts governamentals) - Hi ha una URL a un servidor conegut (universitat, centre de recerca, etc.) Si - El servei ja estat citat abans i/o té documentació externa? Si quan és 	1 1 1 1 1	100%	5

⁹⁵ Privacy Policy. AustLII, July 2003. Url: <http://www.AustLII.edu.au/AustLII/privacy.html>; entre d'altres polítiques (copyright, enllaços, etc.), url: <http://www.AustLII.edu.au/AustLII/#policies>

	aplicable (per exemple, motor de cerca SINO, hi ha literatura científica al respecte, o veure secció Tools → AustLII Apps)	1		
Completesa (2,3), [Cobertura, Abast, Granularitat, Densitat i Extensió]	-Quocient del nombre de serveis inclosos i el nombre de serveis bàsics existents (cerques, alertes, butlletins, etc.) ⁹⁶ : 54% són serveis bàsics (no obstant, manca d'un servei fonamental, opcions per guardar cerques i reutilització d'aquestes). -Nivell de descripció dels recursos inclosos en els serveis i/o dels propis serveis? Si - Quina és la informació bàsica que ha de contenir el servei (metadades)? dades mínimes i insuficients (títol, data + font) ⁹⁷ -Nivell d'especificitat, els serveis s'adeqüen a l'audiència, etc.?	0	50%	3
		1		
		0		
		1		
Disponibilitat / Accessibilitat (1,2,3) [Accessibilitat, recuperabilitat, etc.]	El lloc conté informació que està fàcilment disponible? Si Està disponible a través d'altres fonts d'informació? Si (quan aplica, veure Apps) El recurs facilitar l'accés a informació de diferents formats? Si, però limitada (en el cas dels serveis, l'accés de la informació es limita a HTML, PDF i RTF/ASCII, no inclou serveis RSS, etc.). Els serveis proporcionen accés a la informació que els usuaris poden no	1	80%	5
		1		
		0		
		1		
		1		

⁹⁶ Serveis mínims: s'interpreten serveis mínims els motors de cerca i navegació, serveis de notícies i arxiu (sigui quin sigui el seu mitjà o suport), serveis que permetin guardar continguts/cerques, exportar resultats, etc., servei d'alertes sobre el contingut (independentment del mitjà i suport: e-mail, RSS, etc.), i guies/tutorials del sistema per als usuaris.

⁹⁷ Les dades de citació mínimes recomanades per a la citació de materials per part dels proveïdors de dades dintre d'AustLII són: any de publicació, designador/autoria, i numero de seqüència. Més informació a: *Vendor and Medium Neutral Citations*. Url: <http://www.AustLII.edu.au/techlib/standards/mnc.html>

	ser capaços de trobar / accedir a les biblioteques? Si (per exemple, els serveis d'actualització de Legislació, Revistes, i actualització de l'estat de casos de lleis (vegeu secció "User Tools" ⁹⁸))			
Rellevància (1, 2, 3) [precisió, redundància, aplicabilitat i utilitat]	Grau en què la informació recuperada compleix les necessitats dels usuaris. Quant a serveis es refereix, s'aplica a la perspectiva de l'usuari, sobretot referent al serveis d'informació que puguin proporcionar els sistemes (serveis d'alerta, serveis de cerca, butlletins, guies temàtiques, etc.)? Si Aquest atribut potser difícil d'avaluar, pel nivell de disponibilitat dels serveis, però en aquest cas tots els serveis són gratuïts. Mancança però de servei d'alt valor afegit com per exemple, exportació d'un conjunt de resultats rellevants (la descàrrega és registre a registre), exportació de resultats amb d'altres formats (només exportació formats textuais), i no hi ha servei d'exportació de cites ni de emmagatzemament de cerques. L'exportació dels resultats en formats bibliogràfics (.enl, .ris, etc.)	1	50%	3
		0		
Precisió (1,2, 3)	Aquest atribut és limita als serveis d'informació/documentació (obtenció de documents, serveis d'alertes, butlletins de notícies, serveis d'actualització, etc.), cerca i	0	75%	5
		1		

⁹⁸ User Tools. AustLII. Url: <http://www.AustLII.edu.au/techlib/tools/> (Darrer accés:02/09/2013)

	<p>navegació. Es poden incloure les mateixes interrogacions que en la dimensió del sistema per tal de crear filtres, etc.? No (informació predefinida pel sistema, exemple, butlletins via e-mail)</p> <p>- Informació obtinguda a través del servei és precisa i adequada? Si (element subjectiu)</p> <p>- La informació es pot verificar amb exactitud? Si</p> <p>- Correctesa de la informació rebuda (sense faltes ortogràfiques, gramaticals o tipogràfiques), etc.? Si</p>	1		
		1		
Valoració Màxima (30)	Puntuació total paràmetres serveis	26 (86% IQ) = 5		
3. Dimensió	Descripció/Qüestions	[Si=1; No=0];		
		[Nivell d'IQ=1,3,5]		
Autoria (2,3)	<p>- És la font/recurs atribuïble a un autor o una organització de bona reputació? Si</p> <p>- Es coneix l'origen del recurs? Si</p> <p>- Apareix el nom complet de l'autor? Si</p> <p>- Es coneix l'autor/institució o editor del recurs? Si</p> <p>- Hi ha una URL a un servidor conegut (universitat, centre de recerca, etc.) Si</p> <p>- L'autor ja ha estat citat abans? Si (els autors són institucions, que integren les seves bases de dades al sistema, etc.)</p>	1	100%	5
		1		
		1		
		1		
		1		

	- El recurs proporciona altres cites/referències de l'autor? Si (seleccionant la institució autora del registre apareix informació i accés de les seves publicacions)	1		
	-Hi ha disponible informació sobre l'autor? Si	1		
Actualització (2,3)	- Nivell i/o polítiques d'actualització de la col·lecció? No apareixen	0	75%	5
	- Hi ha una declaració sobre la freqüència d'actualització? Si	1		
	- Si es tracta d'un recurs estàtic (no actualitzat) és la informació útil per a l'usuari/audiència? Si	1		
	- Com sensible al temps és la informació continguda, i com es relaciona amb la freqüència d'actualització? En general la informació legal, és susceptible en el temps, per això, i tot i que l'actualització en el sistema està present a tots els nivells, AustLII ofereix 3 serveix extra d'actualització de continguts (legislació, revistes i casos legals)	1		
Completesa (2,3), [Cobertura, Abast, Granularitat, Densitat i Extensió], aquí s'inclou també	-Quocient del nombre d'elements inclosos en la col·lecció i el nombre d'elements del món real? No es pot respondre de forma objectiva, però el grau de cobertura és molt alt, perquè la informació que forma la col·lecció prové de totes les fonts legals possibles a nivell d'Australasia, i directament dels seus creadors ⁹⁹ .	1	54%	3
	- Explicació de criteris de selecció i			

⁹⁹ AustLII dona accés a 554 bases de dades de lleis i més de 500.000 registres

l'ús de Metadades (ús 15 elements Dublin Core) o altres normes de descripció	incorporació de noves fonts/documents?No	0	75%	5
	- Nivell de descripció dels recursos (especificitat dels recursos, ús de metadades, etc.)?Baix, no ús de metadades estàndards)	0		
	-Quina és la quantitat mínima d'informació necessària per crear una descripció dels recursos al catàleg? Nivell mínim, citat anteriorment, no usen DublinCore, MARC, AACR2, etc. per a la descripció de recursos (vegeu secció estàndards propis ¹⁰⁰)	0		
	-A quin nivell es poden seleccionar els recursos catalogats? Nivell mínim (títol, any, referència llei, autoria, etc.)	1		
	-Ús de descriptors/termes específics per a la indexació/Navegació?Si	1		
	- Cobertura documental: quin tipus de fonts es recullen? Tot tipus dintre del món legal	1		
	- Cobertura temàtica: apareixen els descriptors indexats en el document de referència de la col·lecció?Si	1		
	- Cobertura temporal: apareix la data d'inici de la referència?	1		
	- Quin és el valor relatiu o el recurs en comparació amb altres disponibles en aquest tema? Alt (fonts d'informació d'alta reputació)	1		
	Disponibilitat/ Accessibilitat (1,2,3)	-Què hi ha disponible fora de la col·lecció?No es pot respondre objectivament		
	- Inclusió del text complet?Si (si està			

¹⁰⁰ Standards, Recommendations, and Guidelines. AustLII. Url: <http://www.AustLII.edu.au/techlib/standards/>

[Nivell de fiabilitat]	disponible en la font d'origen)	1		
	- Disponibilitat d'enllaços/referències al text complet? Si	1		
	- Integració de diferents formats en la col·lecció? Si (.txt,.html,.pdf., etc.)	1		
Precisió (1,2, 3)	Quocient entre el nombre de valors correctes recuperats pel sistema i el nombre total de valors continguts en el sistema. (Valor no determinant, perquè no apareixen dades del total de documents recollits en la col·lecció, ni subcol·leccions temàtiques). Nivell IQ baix = 1	0	80%	5
	-La informació es pot verificar amb exactitud? Si	1		
	-El recurs, mostra cites bibliogràfiques o referències per confirmar la informació? Si ¹⁰¹	1		
	- La presentació de la informació en la RI és correcta (sense faltes ortogràfiques/gramaticals)? Si	1		
	-Hi ha prevalença d'errors ortogràfics? No	1		
Rellevància (1, 2,3) [precisió, redundància, aplicabilitat i utilitat]	Atribut subjectiu, no valorable col·lectivament perquè és l'usuari final qui valora la col·lecció segons les seves necessitats d'informació, i l'especificitat de la col·lecció. Precisió del domini, aplicabilitat, utilitat, etc.? Si (recursos disponibles són científics		43%	3

¹⁰¹ En la interfície individual dels registres recuperats, es poden cercar tots els materials que hi al sistema referents a un registre, opció "Noteup" i/o escollir l'opció de "Notes" que cerca notes o esmenes associades al recurs.

	i tècnics, i corresponen a l'audiència)	1		
	- Hi ha accés de navegació per matèries? Si (però només secció catàleg)			
	- Hi ha possibilitat de cercar termes genèrics i específics (navegabilitat per estructura jeràrquica)? Si (fins a 2 nivells, 1r temàtic, 2n per institució/tipus documental, especificitat temàtica baixa)	0		
	- Fiabilitat de la cerca per classificació/descriptors de termes específics? No (no hi ha NT ¹⁰²)	0		
	- Consistència dels termes d'indexació (un concepte/tema ha d'aparèixer d'una forma única)? Si	0		
	- Sugerències de cerca per termes/conceptes alternatius? No	1		
	- Capacitat per discriminació de termes? Si (possibilitat d'ús d'operador booleà NOT)	0		
		1		
Valoració Màxima (30)	Puntuació total paràmetres col·lecció	26 (87% IQ) = 5		
Total (90)	Valoració total	74 (82% IQ) = 5		

2.9.4. Qüestionari LIBRIS (2003 -)

Atributs ^{103,104}	Escala de valor d'IQ
------------------------------------	-----------------------------

¹⁰² NT= Narrower Term, els termes específics no estan visibles en la jerarquia dels encapçalaments de matèries, només apareixen termes relacionats (RT).

	Avaluació Qualitativa/Quantitativa	Puntuació		Total
Dimensió Sistema (Interrogació)	Descripció/Questions ¹⁰⁵	Si=1; No=0	Nivell IQ= 1,3,5]	
Autoria (1,2,3)	Es coneix l'autoria i l'origen del recurs? Si Altres elements addicionals: - Apareixen dades de contacte (adreça, telèfon, etc.)? No - Hi ha opció a suggeriments, comentaris, valoració, etc.? Si - Apareixen cites/referències de l'autor? Si - Altres?	1 0 0 1	50%	3
Comprensibilitat (1)	Grau en que la informació és compresa per l'usuari (per exemple, multilingüisme): alt - La informació és complerta (de la informació recuperada)? Si - El títol és informatiu? Si - Es dona un resum o dades del contingut? Si (sumari o "table of contents")	1 1 1 1	100%	5
Estructura / Consistència en la representació de la informació (1)	Grau de representació consistent en què l'estructura de la informació s'ajusta a la d'altres fonts, per exemple nivell d'integritat, consistència homogeneïtat, consistència semàntica, compatibilitat, etc. (IQ)	1 1	100%	5

¹⁰³ Naumann, F., Rolker, C., 2000. Assessment methods for information quality criteria, IQ. Citeseer, pp. 148-162.

¹⁰⁴ 1999. A list of quality selection criteria: a reference tool for Internet subject gateways, Selection criteria for quality controlled information gateways UKOLN Metadata Group. Url: <http://www.ukoln.ac.uk/metadata/desire/quality/report-2.html>

¹⁰⁵ Nota: algunes respostes poden ser respostes amb si/no i d'altres amb la descriptiva qualitativa (DQ). Les preguntes específiques són adaptades del model DESIRE.

	-Visualització immediata dels resultats sense passos intermedis?Si	1		
	-Qualitat de la referència bibliogràfica (ús d'estàndards)?Si	1		
	- Possibilitat de selecció del criteri d'ordenació dels resultats?Si (rellevància, alfabètic A-Z i Z-A, novetat/antiguitat o nº biblioteques que el contenen)	1		
	- Possibilitat de selecció dels registres de sortida (impressió, gravació o enviament via e-mail)?Si (sota cada registre opció de guardar, enviar via e-mail, citar o crear RSS de la cerca)	1		
	- Possibilitat de limitar o augmentar el nombre de resultats de la cerca en cada pantalla?Si (grups de 10, 25 i 50)	1		
	- Dona com a resultat els termes de cerca? Si			
	- Opcions extra de visualització dels resultats?Si (llista curta o extensa)			
Integritat (1): integritat de la informació, del lloc web i del sistema	- La informació proporcionada pel sistema és actual i està actualitzada?Si	1		
		1		
	- Existeix un manteniment adequat dels continguts informatius?Si	1		
	S'amplia informació sobre el recurs?Si (enllaç a al font original, enllaços a la biblioteca/ques dipositàries, i enllaç a altres documents del mateix autor)	1		
		1		
	- Hi ha enllaços morts?No (positiu)	1		
	- Hi ha enllaços a llocs on s'ha mogut o hi ha redireccionament a un nou lloc web? No aplica (es comptabilitza	1		
		1	92 %	5

	positiu) <ul style="list-style-type: none"> - Les pàgines web del sistema tenen la data amb la última data de revisió? Si - Apareix el número de versió dels recursos? Si (forma part de les regles de catalogació, ús de MARC) - Apareix la data de la darrera actualització del recurs? Si (vegeu opció "details" dintre d'un registre) - Hi ha una descripció de les freqüències de l'actualització dels recursos?No - L'organització/persona que allotja el recurs sembla tenir el compromís amb el manteniment i l'estabilitat dels recursos?Si - El lloc s'actualitza/ amb freqüència / manteniment?Si -Fiabilitat del sistema: Si 	0 1 1 1		
Disponibilitat / Accessibilitat (1,2,3) [nivell d'accessibilitat i recuperabilitat]	-Ofereix serveis per facilitar l'accés a informació disponible en diferents formats, suports, etc.?Si (cas de rebre informació via RSS, e-mail, etc.) -Té serveis que faciliten la conversió/extracció de la informació amb formats diferents? Si (sota de cada registre hi ha l'opció de "cites" que és un servei per a l'exportació de les cites en diferents formats: MARC-XML, bibliogràfics, text pla, etc.) -Ofereix serveis per a que un usuari pugui accedir a la informació de forma indirecta? (serveis d'alerta, butlletins, guies temàtiques, etc.) Si (però	1 1 0 1 1	83%	5

	<p>mínims)</p> <ul style="list-style-type: none"> - Informació sobre existència de limitacions o condicions d'accés (per exemple, data límit d'accés, si es requereix quota, etc.?Si - Ofereixen servei d'Obtenció de Documents?Si (loan/interlibrary loan, opcions disponibles a la part de sota del registre seleccionat) - Existència de missatges clars quan la connexió no es possible?Si 	1		
Seguretat (1) [Privacitat i seguretat d'accés]	<p>Grau en què la informació es transmet de forma privada a l'usuari. IQ= 5 (opció de personalització, cerca i emmagatzemament privat de l'usuari, amb MYIntute)</p> <p>Hi ha polítiques de privacitat, protocols d'intercanvi d'informació, etc.)? Si</p>	1		
		1	100%	5
Valoració Màxima (30)	Puntuació total paràmetres de RI	28 (93% IQ) = 5		
2. Dimensió Serveis (Tipologia)	Descripció/Qüestions	[Si=1; No=0];		
		[Nivell d'IQ=1,3,5]		
Actualització (2,3)	<ul style="list-style-type: none"> - Serveis actualitzats? Si - Apareix data d'actualització?Si - Informació sobre noves versions/serveis/aplicacions?Si - Tutorials actualitzats?Si - etc. 	1		
		1		
		1	100%	5
		1		
Autoria (2,3)	- És la font atribuïble a un autor o una organització de bona reputació?Si	1		
		1		

[informació referent als serveis i la informació que aquests proporcionen en cada servei]	- Es coneix l'origen del recurs? Si	1	71%	5
	- Apareix el nom complet de l'autor? Si	1		
	- Es coneix l'autor/institució o editor del recurs? Si (dades catalogràfiques)	0		
	- Hi ha una URL a un servidor conegut (universitat, centre de recerca, etc.) Si	1		
	- L'autor ja ha estat citat abans? No es coneixen dades (si apareixen el nombre de biblioteques que tenen l'element, seria interessant que aparegués el nombre de vegades que ha estat prestat)	0		
El recurs proporciona altres cites/referències de l'autor? Si				
-Hi ha disponible informació sobre l'autor (CV., adreça, e-mail, etc.)? No, a excepció de les publicacions i enllaços				
Completesa (2,3), [Cobertura, Abast, Granularitat, Densitat i Extensió]	-Quocient del nombre de serveis inclosos i el nombre de serveis bàsics existents (cerques, alertes, butlletins, etc.) ¹⁰⁶ : 46% són serveis bàsics	0	75%	5
	-Nivell de descripció dels recursos inclosos en els serveis i dels propis serveis? Si	1		
	- Quina és la informació bàsica que ha de contenir el servei? Alt (quan es tracta del contingut, aporta metadades MARC complertes i/o cites bibliogràfiques complertes, quan es tracta d'altres serveis: cerca, ajuda, etc. Inclou tota la informació possible)	1		
		1		
	-Nivell d'especificitat, els serveis			

¹⁰⁶ Serveis mínims: s'interpreten serveis mínims els motors de cerca i navegació, serveis de notícies i arxiu (sigui quin sigui el seu mitjà o suport), serveis que permetin guardar continguts/cerques, exportar resultats, etc., i servei d'alertes sobre el contingut (independentment del mitjà i suport: e-mail, RSS, etc.)

	s'adeqüen a l'audiència, etc.? Si			
Disponibilitat / Accessibilitat (1,2,3) [Accessibilitat, recuperabilitat, etc.]	-El lloc conté informació que està fàcilment disponible? Si	1	100%	5
	-Està disponible a través d'altres fonts d'informació?Si	1		
	- El recurs facilita l'accés a informació de diferents formats? I aquests són accessibles per tots els usuaris? Si (per exemple, serveis per descàrrega de contingut amb accés a diferents formats).	1		
		1		
	-El recurs de proporcionar accés a la informació que els usuaris poden no ser capaços de trobar / accedir a les biblioteques?Si (sistemes de cerca/navegació alternatius que faciliten l'accés a informació de recursos interns i externs, per exemple "search facilities")	1		
Rellevància (1, 2, 3) [precisió, redundància, aplicabilitat i utilitat]	- Grau en què la informació recuperada compleix les necessitats dels usuaris. Quant a serveis es refereix, s'aplica a la perspectiva de l'usuari, sobretot referent al serveis d'informació que puguin proporcionar els sistemes? Si, però són insuficients (no hi han butlletins, guies temàtiques o de selecció de llibres, etc. que puguis consultar-ho o subscriure't)	0	50%	3
	- L'adequació i utilitat dels serveis de cara a l'audiència és alt, i hi ha molts serveis addicionals interessants (54%), com per exemple realització de cites bibliogràfiques, sistemes de cerca/fonts d'informació externes, etc. Manca d'app per a mòbil i punts d'accés a xarxes socials.	1		

Precisió (1,2, 3)	<p>Es limita als serveis d'informació/ documentació (SOD, serveis d'alertes, etc.), i cerca/navegació:</p> <ul style="list-style-type: none"> - Es poden incloure les mateixes interrogacions que en la dimensió del sistema? Si (quan aplica, per exemple, serveis d'alerta¹⁰⁷) - Informació obtinguda a través del servei és precisa i adequada? Si - La informació es pot verificar amb exactitud? Si (enllaços externs, catalogació complerta, permanent link, etc.) - Correctesa de la informació rebuda (sense faltes ortogràfiques, gramaticals, etc.)? Si 	1	100%	5
		1		
		1		
		1		
Valoració Màxima (30)	Puntuació total paràmetres serveis	28 (83% IQ) = 5		
3. Dimensió Col·lecció	Descripció/Qüestions	[Si=1; No=0]; [Nivell d'IQ=1,3,5]		
Autoria (2,3)	<ul style="list-style-type: none"> - És la font/recurs atribuïble a un autor o una organització de bona reputació? Si - Es coneix l'origen del recurs? Si - Apareix el nom complet de l'autor? Si - Es coneix l'autor/institució o editor del recurs? Si (normalment) - Hi ha una URL a un servidor conegut (universitat, centre de recerca, etc.) Si - L'autor ja ha estat citat abans? Si, 	1	88%	5
		1		
		1		
		1		
		1		
		1		
		0		

¹⁰⁷ Per exemple, l'ús del RSS (servei d'alerta) s'aplica sobre la equació de cerca realitzada per l'usuari.

	<p>apareixen enllaços a altres publicacions</p> <p>- El recurs proporciona altres cites/referències de l'autor?No</p> <p>-Hi ha disponible informació sobre l'autor (CV., adreça, e-mail, etc.)? No</p>			
Actualització (2,3)	<p>- Nivell i/o polítiques d'actualització de la col·lecció: altes (no s'han trobat errors, i apareixen tots els detalls possibles al respecte, veure secció "details" en qualsevol registre bibliogràfic.</p> <p>- Hi ha una declaració sobre la freqüència d'actualització?No</p> <p>- Si es tracta d'un recurs estàtic (no actualitzat) és la informació útil per a l'usuari/audiència?Si</p> <p>- Com sensible al temps és la informació continguda, i com es relaciona amb la freqüència d'actualització? En general la informació no es sensible amb el temps i en la majoria dels casos no afecta a actualització, hi ha accés a documents electrònics però s'informa de si hi ha accés a altres edicions, etc.</p>	<p>1</p> <hr/> <p>0</p> <hr/> <p>1</p> <hr/> <p>1</p>	75%	5
<p>Completesa (2,3),</p> <p>[Cobertura,</p> <p>Abast,</p> <p>Granularitat,</p> <p>Densitat i],</p> <p>aquí s'inclou també l'ús de Metadades (ús 15 elements</p>	<p>Quocient del nombre d'elements inclosos en la col·lecció i el nombre d'elements del món real. No es pot respondre la equació de forma objectiva, dependrà de l'usuari.</p> <p>- Explicació de criteris de selecció i incorporació de noves fonts/documents?Si</p> <p>- Nivell de descripció dels recursos (especificitat dels recursos, etc.): molt</p>	<p>0</p> <hr/> <p>1</p> <hr/> <p>1</p> <hr/> <p>1</p>	90%	5

Dublin Core) o altres normes de descripció	alt			
	- Ús de metadades o normatives estàndards per a la descripció dels recursos? Si (MARC, ISO 690-2, etc.)	1		
	-Quina és la quantitat mínima d'informació necessària per crear una descripció dels recursos al catàleg? O quina és la informació bàsica que ha de contenir un recurs per a ser seleccionat? Alt, dades complertes amb tots els registres segons les normatives aplicades.	1		
	-A quin nivell es poden seleccionar els recursos catalogats? A nivell individual i col·lectiu.	1		
	-Ús de descriptors/termes específics per a la indexació/Navegació? Si (ús de diferents vocabularis controlats propis i externs)	1		
	- Cobertura documental: quin tipus de fonts es ? Tot tipus			
	- Cobertura temàtica: apareixen els descriptors indexats en el document de referència de la col·lecció? Si (i són navegables)	1		
- Cobertura temporal: apareix la data d'inici de la referència?	1			
- Quin és el valor relatiu o el recurs en comparació amb altres disponibles en aquest tema? Alt (ús de <i>Linked Data</i> , la majoria de catàlegs col·lectius no utilitzen aquesta tecnologia)				
Disponibilitat/ Accessibilitat (1,2,3)	-Què hi ha disponible fora de la col·lecció? No es pot respondre objectivament, és una font multidisciplinar, és una font d'àmbit	0	75%	5

[Nivell de fiabilitat]	nacional, i és una font única i d'alt valor	1		
	- Inclusió del text complert? Si (si està disponible en la font d'origen, sinó hi ha servei de préstec local o interbibliotecari)	1		
	- Disponibilitat d'enllaços/referències al text complert? Si - Integració de diferents formats en la col·lecció? Si (.txt,.html,.pdf., etc.)	1		
Precisió (1,2, 3)	Quocient entre el nombre de valors correctes recuperats pel sistema i el nombre total de valors continguts en el sistema. (Valor no determinant, perquè no apareixen dades del total de documents recollits en la col·lecció, i/o el de les biblioteques integrades en el sistema). Nivell IQ baix = 1	0	60%	3
	-La informació es pot verificar amb exactitud? Si (enllaços externs, catalogació de nivell complert, accés al text complert, etc.).	1		
	-El recurs, mostra cites bibliogràfiques o referències per confirmar la informació? No (referències únicament a enllaços externs)	0		
	-La presentació de la informació en la RI és correcta (sense faltes ortogràfiques/gramaticals)? Si	1		
	-Hi ha prevalença d'errors ortogràfics? No	1		
Rellevància (1, 2,3) [precisió, redundància, aplicabilitat i	Atribut subjectiu, no valorable col·lectivament perquè és l'usuari final qui valora la col·lecció segons les seves necessitats. A nivell de precisió,	1	86%	5

utilitat]	aplicabilitat i utilitat, si es pot dir que els recursos disponibles corresponen a l'audiència tant per la seva tipologia com per la seva especialització ¹⁰⁸ .	1		
	- Hi ha accés de navegació per matèries (tesaurus, classificacions, etc.)? Si	1		
	- Hi ha possibilitat de cercar termes genèrics i específics (navegabilitat per estructura jeràrquica)? Si (només en suec)	1		
	- Fiabilitat de la cerca per classificació/descriptors de termes específics? Si	1		
	- Consistència dels termes d'indexació (un concepte/tema ha d'aparèixer d'una forma única)? Si	0		
	- Suggestions de cerca per termes/conceptes alternatius? No	1		
	- Capacitat per discriminació de termes? Si (ús operador NOT)			
Valoració Màxima (30)	Puntuació total paràmetres col·lecció	28 (93% IQ) = 5		
Total (90)	Valoració total	84 (95% IQ) = 5		

2.9.5 Qüestionari Europea (novembre 2008 -)

Atributs ^{109,110}	Escala de valor d'IQ
------------------------------------	-----------------------------

¹⁰⁸ LIBRIS: proporciona referències a més de 7 milions de llibres i revistes que formen part de 200 biblioteques de recerca de Suècia. S'inclouen pràcticament tots els títols suecs des de 1866. LIBRIS també conté referències a articles de revistes, notes, mapes i documents electrònics.

¹⁰⁹ Naumann, F., Rolker, C., 2000. Assessment methods for information quality criteria, IQ. Citeseer, pp. 148-162.

¹¹⁰ 1999. A list of quality selection criteria: a reference tool for Internet subject gateways,

	Avaluació Qualitativa/Quantitativa	Puntuació		Total
1. Dimensió Sistema (Interrogació)	Descripció/Qüestions ¹¹¹	Si=1; No=0	Nivell IQ= 1,3,5]	
Autoria (1,2,3)	<p>Es coneix l'autoria i l'origen del recurs? Si</p> <p>Altres elements addicionals:</p> <ul style="list-style-type: none"> - Apareixen dades de contacte (adreça, telèfon, etc.)? No - Hi ha opció a suggeriments, comentaris, valoració, etc.? Si (però de forma externa, hi ha opció de compartir el registre en aplicacions 2.0, o bé citar el document a Wikipedia, a part d'escriure comentaris al propi blog, facebook o twitter d'Europeana entre d'altres aplicacions 2.0). - Apareixen cites/referències de l'autor? Si (el nom de l'autor es navegable i per tant, apareixen la resta d'obres contingudes en el sistema del mateix autor) 	1 0 1 1	75%	3
Comprensibilitat (1)	<p>Grau en que la informació és compresa per l'usuari: alt (ús de sistemes de traducció automàtica dels registres en tots els idiomes EU, i possibilitat de canviar l'idioma de tot el sistema)</p> <ul style="list-style-type: none"> - La informació és completa (de la informació recuperada)? Si - El títol és informatiu? Si - Es dona un resum? Si (un cop s'entra al registre) 	1 1 1 1	100%	5

Selection criteria for quality controlled information gateways UKOLN Metadata Group. Url: <http://www.ukoln.ac.uk/metadata/desire/quality/report-2.html>

¹¹¹ Nota: algunes respostes poden ser respostes amb si/no i d'altres amb la descriptiva qualitativa (DQ). Les preguntes específiques són adaptades del model DESIRE.

Estructura / Consistència en la representació de la informació (1)	Grau de representació consistent en què l'estructura de la informació s'ajusta a la d'altres fonts, per exemple nivell d'integritat, homogeneïtat, consistència semàntica, compatibilitat, etc. (IQ) -Visualització immediata dels resultats sense passos intermedis?Si -Qualitat de la referència bibliogràfica (ús d'estàndards)?Si (DC i EDM ¹¹² , etc.) - Possibilitat de selecció del criteri d'ordenació dels resultats?No - Possibilitat de selecció dels registres de sortida (impressió, gravació o enviament via e-mail)?No (es poden realitzar aquestes opcions i d'altres però només de la cerca o de forma individual de cada registre). - Possibilitat de limitar o augmentar el nombre de resultats de la cerca en cada pantalla?Si - Dona com a resultat els termes de cerca? Si - Opcions extra de visualització dels resultats (per exemple: agrupacions o índex KWIC)?No	1	85%	5
		1		
		1		
		0		
		0		
		1		
		0		
Integritat (1): integritat de la informació, del lloc web i del sistema	- La informació proporcionada pel sistema és actual i està actualitzada?Si - Existeix un manteniment adequat dels continguts informatius?Si S'amplia informació sobre el recurs (per exemple, recursos/enllaços relacionats, biblioteca o base de	1	57%	3
1				
1				
1				
1				

¹¹² Europeana Data Model (EDM)

	<p>dades dipositària, etc.)?Si (enllaç a al font original)</p> <p>- Hi ha enllaços morts?No</p> <p>- Hi ha enllaços a llocs on s'ha mogut o hi ha redireccionament a un nou lloc web? Si</p> <p>- Les pàgines web del sistema tenen la data amb la última data de revisió? Si</p> <p>- Apareix el número de versió dels recursos?No (tot i que no és un element negatiu)</p> <p>- Apareix la data de la darrera actualització del recurs? Si</p> <p>- Hi ha una descripció de les freqüències de l'actualització dels recursos?Si/No</p> <p>- L'organització/persona que allotja el recurs sembla tenir el compromís amb el manteniment i l'estabilitat dels recursos?Si/No</p> <p>- El lloc s'actualitza/manté amb freqüència / manteniment?Si</p> <p>-Fiabilitat del sistema: Si</p>	<p>1</p> <p>0</p> <p>1</p> <p>0</p> <p>1</p> <p>1</p> <p>1</p>		
<p>Disponibilitat / Accessibilitat (1,2,3)</p> <p>[nivell d'accessibilitat i recuperabilitat]</p>	<p>-Ofereix serveis per facilitar l'accés a informació disponible en diferents formats, suports, etc.?Si</p> <p>-Té serveis que faciliten la conversió/extracció de la informació amb formats diferents? No</p> <p>-Ofereix serveis per a que un usuari pugui accedir a la informació de forma indirecta? (serveis d'alerta, butlletins,</p>	<p>1</p> <p>0</p> <p>1</p>	57%	3

	<p>guias temàtiques, etc.) Si</p> <p>- Informació sobre existència de limitacions o condicions d'accés (per exemple, data límit d'accés, si es requereix quota, etc.? Si</p> <p>- Ofereixen servei d'Obtenció de Documents? No</p> <p>- Existència de missatges clars quan la connexió no es possible? Si</p>	1		
		0		
		1		
Seguretat (1) [Privacitat i seguretat d'accés]	<p>Grau en què la informació es transmet de forma privada a l'usuari. IQ= 5 (opció de registrar-se, etc. Veure polítiques¹¹³)</p> <p>Hi ha polítiques de privacitat, protocols d'intercanvi d'informació, etc.)? Si</p>	1	100%	5
		1		
Valoració Màxima (30)	Puntuació total paràmetres de RI	24 (80% IQ) = 5		
2. Dimensió Serveis (Tipologia)	Descripció/Qüestions	[Si=1; No=0];		
		[Nivell d'IQ=1,3,5]		
Actualització (2,3)	<p>- Serveis actualitzats? Si (n'hi ha però, en desenvolupament)</p> <p>- Apareix data d'actualització? Si</p> <p>- Informació sobre noves versions/serveis/aplicacions? Si</p> <p>- Tutorials actualitzats? Si</p> <p>- etc.</p>	1	100%	5
		1		
		1		
		1		
Autoria (2,3)	- És la font atribuïble a un autor o una organització de bona reputació? Si	1	71%	5
		1		

¹¹³ Policy. Europeana.eu, 2008. Url: <http://www.europeana.eu/portal/rights/privacy.html>

[informació referent als serveis i la informació que aquests proporcionen en cada servei]	- Es coneix l'origen del recurs? Si	1		
	- Apareix el nom complet de l'autor? Si	1		
	- Es coneix l'autor/institució o editor del recurs? Si (enllaços a les seves webs, etc.)	0		
	- Hi ha una URL a un servidor conegut (universitat, centre de recerca, etc.) Si	1		
	- L'autor ja ha estat citat abans? No es coneixen dades	0		
	El recurs proporciona altres cites/referències de l'autor? Si			
-Hi ha disponible informació sobre l'autor (CV., adreça, e-mail, etc.)? No				
Completesa (2,3), [Cobertura, Abast, Granularitat, Densitat i]	-Quocient del nombre de serveis inclosos i el nombre de serveis bàsics existents (cerques, alertes, butlletins, etc.) ¹¹⁴ : mitjà-alt, 67% es cobreixen els serveis bàsics i els serveis addicionals són innovadors (per exemple, servei de recopilació d'història oral audiovisual d'usuaris externs sobre la 1 ^a Guerra Mundial)	1	75%	5
	-Nivell de descripció dels recursos inclosos en els serveis i dels propis serveis? Alt (sempre que aplica ús sempre de EDM)	1		
	- Quina és la informació bàsica que ha de contenir el servei (metadades)? Si inclou sempre informació bàsica i quan és possible d'addicional	1		
	Nivell d'especificitat, els serveis s'adeqüen a l'audiència, etc.? Si	1		

¹¹⁴ Serveis mínims: s'interpreten serveis mínims els motors de cerca i navegació, serveis de notícies i arxiu (sigui quin sigui el seu mitjà o suport), serveis que permetin guardar continguts/cerques, exportar resultats, etc., i servei d'alertes sobre el contingut (independentment del mitjà i suport: e-mail, RSS, etc.), i serveis com "About us" o "Help".

Disponibilitat / Accessibilitat (1,2,3) [Accessibilitat, recuperabilitat, etc.]	El lloc conté informació que està fàcilment disponible? Si	1	100%	5
	Està disponible a través d'altres fonts d'informació? Si	1		
	El recurs facilitar l'accés a informació de diferents formats? i aquests són accessibles per tots els usuaris? Si (textuals, àudio, so, hipertextuals, etc.)	1		
		1		
	El recurs proporciona accés a la informació que els usuaris poden no ser capaços de trobar / accedir a les biblioteques? Si, perquè els cercadors d'Europeana tenen la funció d'autocompletar, detecten l'arrel de la paraula i et proposen termes, a més a més d'indicar-te si el terme està indexat com a matèria, títol, etc.	1		
Rellevància (1, 2, 3) [precisió, redundància, aplicabilitat i utilitat]	Grau en què la informació recuperada compleix les necessitats dels usuaris. Quant a serveis es refereix, s'aplica a la perspectiva de l'usuari, sobretot referent al serveis d'informació que puguin proporcionar els sistemes (serveis d'alerta, serveis de cerca, butlletins, guies temàtiques, etc.)? Si	1	50%	3
	Aquest atribut és difícil d'avaluar per la seva subjectivitat, però en general els serveis són gratuïts, s'adeqüen als usuaris i el nivell d'utilitat dels serveis a l'audiència és alt, però encara falta molt de desenvolupament orientat a l'usuari i millora dels sistemes de cerca/navegació, i selecció i gestió de recursos.	0		
Precisió (1,2, 3)	Aquest atribut és limita als serveis d'informació/documentació (obtenció de documents, serveis d'alertes, butlletins de notícies, guies temàtiques, etc.), cerca i navegació.	0	80%	5

	Es poden incloure les mateixes interrogacions que en la dimensió del sistema? No hi ha serveis d'informació o de préstec, si hi ha serveis de d'informació/comunicació, però no pots realitzar cerques.	1		
		1		
	- Informació obtinguda a través del servei és precisa i adequada? Si	1		
	- La informació es pot verificar amb exactitud? Si			
	- Correctesa de la informació rebuda (sense faltes ortogràfiques, gramaticals o tipogràfiques), etc.?			
Valoració Màxima (30)	Puntuació total paràmetres serveis	28 (93% IQ) = 5		
3. Dimensió Col·lecció	Descripció/Qüestions	[Si=1; No=0];		
		[Nivell d'IQ=1,3,5]		
Autoria (2,3)	- És la font/recurs atribuïble a un autor o una organització de bona reputació? Si	1		
		1		
	- Es coneix l'origen del recurs? Si	1		
	- Apareix el nom complet de l'autor? Si	1		
		1		
	- Es coneix l'autor/institució o editor del recurs? Si (normalment)	0	63%	3
		0		
	- Hi ha una URL a un servidor conegut (universitat, centre de recerca, etc.) Si	0		
		0		
	- L'autor ja ha estat citat abans? Es desconeix			
	- El recurs proporciona altres cites/referències de l'autor? No			
	- Hi ha disponible informació sobre			

	l'autor (CV., adreça, e-mail, etc.)? No			
Actualització (2,3)	- Nivell i/o polítiques d'actualització de la col·lecció. Són altes, formen part de la seva raó de ser (vegeu capítol Evolució i IQ). IQ=5	1	100%	5
	- Hi ha una declaració sobre la freqüència d'actualització? Si			
	- Si es tracta d'un recurs estàtic (no actualitzat) és la informació útil per a l'usuari/audiència? Si	1		
	- Com sensible al temps és la informació continguda, i com es relaciona amb la freqüència d'actualització? En general la informació no es sensible amb el temps i en la majoria dels casos no afecta l'actualització (només per a la revisió de enllaços trencats i aquests són actualitzats)	1		
Completesa (2,3), [Cobertura, Abast, Granularitat, Densitat i Extensió], aquí s'inclou també l'ús de Metadades (ús 15 elements Dublin Core) o altres normes de descripció	Quocient del nombre d'elements inclosos en la col·lecció i el nombre d'elements del món real. No es pot respondre la equació de forma objectiva.	0	81%	5
	- Explicació de criteris de selecció i incorporació de noves fonts/documentos? Si	1		
	- Nivell de descripció dels recursos (especificitat dels recursos, etc.): alt	1		
	- Ús de metadades o normatives estàndards per a la descripció dels recursos? Si	1		
	- Quina és la quantitat mínima d'informació necessària per crear una descripció dels recursos al catàleg? O quina és la informació bàsica que ha de contenir un recurs per a ser	1		
		1		

	<p>seleccionat? (ús de normatives, (AACR2¹¹⁵, ISO 690-2¹¹⁶, Dublin Core, EDM, etc.). Nivell mínim en tots els casos, nivell complert en alguns (AACR2). IQ=3</p> <p>-A quin nivell es poden seleccionar els recursos catalogats? A tots</p> <p>-Ús de descriptors/termes específics per a la indexació/Navegació?No (ús de BT¹¹⁷)</p> <p>- Cobertura documental: quin tipus de fonts es recullen? Tot tipus</p> <p>- Cobertura temàtica: apareixen els descriptors indexats en el document de referència de la col·lecció?Si</p> <p>- Cobertura temporal: apareix la data d'inici de la referència?</p> <p>- Quin és el valor relatiu o el recurs en comparació amb altres disponibles en aquest tema? Alt (polítiques fortes de selecció, tria de documents)</p>	0		
		1		
		1		
		1		
		1		
		1		
Disponibilitat/ Accessibilitat (1,2,3)	-Què hi ha disponible fora de la col·lecció? No es pot respondre objectivament, però encara falten moltes fons possibles i institucions de l'àmbit patrimonial per incorporar-se per tal de tenir una major cobertura.	0	75%	5
[Nivell de fiabilitat]	- Inclusió del text complert?Si (si està disponible en la font d'origen i si el copyright ho permet)	1		
	- Disponibilitat d'enllaços/referències al text complert?Si	1		
	- Integració de diferents formats en la	1		

¹¹⁵ AACR2: Anglo-American Cataloguing Rules, 2nd ed. Url: <http://www.aacr2.org/>

¹¹⁶ ISO 690-2: Referencias bibliográficas: documentos electrónicos y sus partes. Url: <http://biblioteca.ucv.cl/site/servicios/documentos/ISO690-2.pdf?cp=5>

¹¹⁷ BT= Broader term o terme genèric (llenguatge emprat en vocabularis controlats)

	col·lecció? Si (.txt,.html,.pdf., etc.)			
Precisió (1,2, 3)	Quocient entre el nombre de valors correctes recuperats pel sistema i el nombre total de valors continguts en el sistema. (Valor no determinant, perquè no apareixen dades del total de documents recollits en la col·lecció, ni subcol·leccions temàtiques). Nivell IQ=1 (per no ser puntuable)	0	80%	5
	-La informació es pot verificar amb exactitud?Si	1		
	-El recurs, mostra cites bibliogràfiques o referències per confirmar la informació? Si (quan aplica: notes i referències)	1		
	-La presentació de la informació en la RI és correcta (sense faltes ortogràfiques/gramaticals)?Si	1		
	-Hi ha prevalença d'errors ortogràfics? No	1		
Rellevància (1, 2,3) [precisió, redundància, aplicabilitat i utilitat]	Atribut subjectiu, cada usuari valora la col·lecció segons les seves necessitats. A nivell de precisió, aplicabilitat i utilitat, es pot dir que els recursos disponibles són de l'àmbit patrimonial i per tant, corresponen a l'audiència a la que es dirigeixen.	1	83%	5
	- Hi ha accés de navegació per matèries (tesaurus, etc.)? Si (cercador semàntic, i en el principal, cada matèria indexada al registre és un enllaç a la consulta de vocabularis d'on s'ha extret, etc.)	1		
	- Hi ha possibilitat de cercar termes genèrics i específics (navegabilitat per	0		

	<p>estructura jeràrquica)?No (tampoc en la cerca semàntica)</p> <p>- Fiabilitat de la cerca per classificació/descriptors de termes específics?Si</p> <p>- Consistència dels termes d'indexació (un concepte/tema ha d'aparèixer d'una forma única)? Si (ús d'operador booleà NOT, o a posteriori de la cerca afegir descriptors addicionals, opció "Add" de refinament¹¹⁸)</p> <p>- Suggestions de cerca per termes/conceptes alternatius?Si</p> <p>- Capacitat per discriminar termes?Si</p>	1		
		1		
		1		
		1		
Valoració Màxima (30)	Puntuació total paràmetres col·lecció		28 (93% IQ) = 5	
Total (90)	Valoració total		80 (89% IQ) = 5	

¹¹⁸ Opcions de refinament: per tipus de mitjà emprat pel recurs (text, vídeo, so, imatge, etc.), la llengua de descripció del recurs, any, país proveïdor i copyright. Addicionalment es pot incloure o excloure el contingut amb contribucions per part de l'usuari.

Annexes Capítol 3. Indexació i navegació creuades

Annex 3.1. Navegació creuada

a. Mapeig de relacions i sintaxi en la navegació creuada:

Mapeig de relacions establertes per Renardus:

- Plena equivalència (=)
- Equivalència entre termes “NO ACCEPTATS” (>)
- Equivalència entre termes “ACCEPTATS” (<)
- Major overlap (MJ)
- Minor overlap (MN)

Mapeig de Sintaxi:

- LN (Local classification notation)
- LC (Local classification caption)
- RL (Codis de relacions, p.ex.: =, >, <, MJ, MN)
- DN (DDC notation)
- DC (DDC caption)
- UR (Local browsing URL for the class)
- NT (notes)
- :RE: (Record end)

b. Eina de mapeig Renardus

Ús de la classificació universal DDC per al mapeig dels vocabularis controlats

The screenshot displays a web interface for classifying mining-related terms. It features two hierarchical trees at the top and a mapping table below.

Left Tree (Local GOK Classification):

- ▲ ZU+000 Mining: General
 - ddc: [622](#)
 - ZU+060 Mining Districts
 - ddc: [307.766](#)
 - ZU+150 Mining Law
 - ZU+180 Mineral Economics
 - ddc: [338.2](#) [338.85](#)
 - ZU+200 Education, Jobs, Future
 - ZU+400 Mining History
 - ddc: [622](#)
 - ZUB+000 Mine Surveying, Prospecting
 - ZUE+000 Geomechanics, Rock Mechanics
 - ▼ ZUG+000 Mining Technology
 - ddc: [622.2](#)
 - ▼ ZUS+000 Mining Branches
 - ddc: [622.3](#)

Right Tree (Main Classes):

- ▲ --- Main Classes
 - ▲ 6-- Technology
 - GOK: [XX+300](#)
 - ▲ 62- Engineering
 - ▼ 620 Engineering and allied operations
 - ▼ 621 Applied physics
 - ▼ 622 Mining and related operations
 - GOK: [ZU+000](#) [ZU+400](#)
 - ▼ 623 Military and nautical engineering
 - ▼ 624 Civil engineering
 - ▼ 625 Engineering of railroads and roads
 - 626
 - ▼ 627 Hydraulic engineering
 - ▼ 628 Sanitary and municipal engineering
 - ▼ 629 Other branches of engineering

Mapping Table (existing relation (one-sided →)):

GOK	relation	ddc	relevance	remark
ZU+000	LNE >	622		
cap 1	Mining: General			
cap 2	Mining and related operations			
url	http://www.Geo-Guide.de/cgi-bin/ssgfi/anzeige.pl?db=geo			

Buttons: modify delete

Com s'observa en la imatge superior, la columna de la dreta és la classificació universal DDC, i la columna de l'esquerra equival a un esquema local GOK (*Göttinger Online-Klassifikation*). A sota de les dues classificacions s'observa el sistema de mapeig, on d'està mapejant la classificació GOK (primera columna) en relació a la classificació DCC (tercera Columna del sistema de mapeig).

Com s'observa en la part inferior de la imatge, apareix un exemple de mapeig, on es mapeja la classificació local (GOK) amb la classificació DDC (622). La classificació local, ZU+000, correspon al terme "Mining and related operations", mentre que la classificació 622 (DDC) significa "Mining General". El mateix sistema de mapeig, a més a més, permet fer una evaluació del mapeig segons la seva relevància (*relevance*).

Annex 3.2. Exemples de navegació creuades

a) Exemple 1. Navegació per DDC

Renardus WELCOME TO RENARDUS
Search and browse quality Internet resources

Home Advanced search Browse by subject About Us Help

Search help

1. Search logic:
 All terms must occur in the results
 Any (at least one) term must occur in the results

2. Search terms: (Enter only one search term per box. Use * for truncation.)

Title/Subject/Description

Creator

DDC Classification

(When using Title, DDC Classification, Subject and Type a term selection window will appear for your assistance.)

3. Options - see below

4. Execute: Search

Gateway Selection:

- Anglistik Guide: Virtual Library of Anglo-American Culture - Literature at SUB Göttingen (GoeAng) ([about](#))
- DAINet - German Agricultural information Network, Germany (DAINet) ([about](#))
- Danmarks Elektroniske Forskningsbibliotek (deff) ([about](#))
- DutchESS - Dutch Electronic Subject Service (DutchESS) ([about](#))
- Finnish Virtual Library (FVL) ([about](#))
- Geo-Guide: Virtual Library of Earth Sciences, Geography, Thematic Maps, and Mining at SUB Göttingen (GoeGeo) ([about](#))
- History Guide: Virtual Library of Anglo-American Culture - History at SUB Göttingen (GoeHist) ([about](#))
- MathGuide: Part of the Virtual Library of Mathematics at SUB Göttingen (GoeMath) ([about](#))
- NOVAGate - Nordic Gateway to Information in Forestry, Veterinary and Agricultural Sciences (NOVAGate) ([about](#))
- Online theses, Germany (DNB-Theses) ([about](#))
- Resource Discovery Network (RDN) ([about](#))

Home Advanced search Browse by subject About Us Help

b) Exemple 2. Navegació per l'esquema de matèries que apareix en la pàgina principal:

Renardus WELCOME TO RENARDUS
Search and browse quality Internet resources

Home Advanced search Browse by subject About Us Help

English Dutch French Finnish German

Browse help

Browse by Subject


- Computers, information & general reference
- Philosophy & psychology
- Religion
- Social sciences
- Language
- Science
- Technology
- Arts & recreation
- Literature
- History & geography

Search help

Search

c) **Exemple 2.1. Nivell de jerarquies:** Un cop seleccionada una matèria des de l' "Esquema de matèries" de Renardus (*Arts & recreation*) passem a un segon procés de navegació on es van seleccionant subnivells de matèries fins trobar el terme que ens interessa *Textile arts*:

[Top Level Subjects](#)

- [Arts & recreation](#)
 - [Drawing & decorative arts](#)
 - Textile arts 
 - Specific materials
 - [Products and processes](#)

d) **Exemple 2.3. "Navegar i saltar"**

'Textile arts' - Related Collections
[Browse Help](#)

Fully equivalent:

[decorative arts: textile arts](#)  (21.86) 

[Textile art: history, general \(677.000\)](#) 

Narrower equivalent:

[Textile art: materials, techniques, research \(677.001\)](#) 

[Textile design, textile industry \(677.002\)](#) 

Broader equivalent:

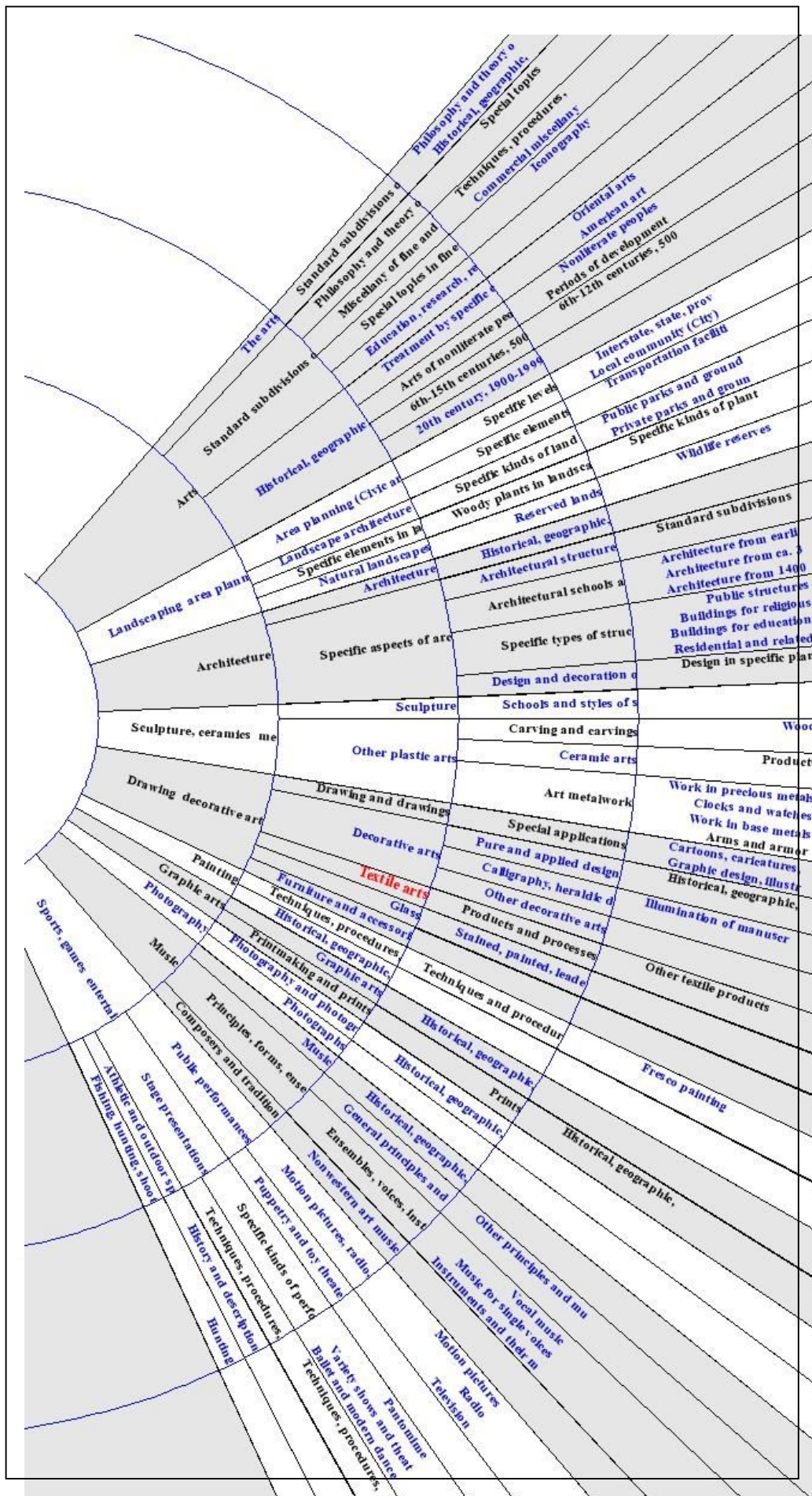
[BASTELN, HANDARBEITEN, HEIMWERKEN \(36\)](#) 

Si seleccionem qualsevol dels termes d'equivalència "saltem" a la col·lecció local de la passarel·la local a través de la qual s'ha recuperat informació. Per exemple, si seleccionem el primer terme d'equivalència plena o "full equivalent", accedim a la passarel·la *DutchESS*:



Com és veu en l'exemple no hi ha cap enllaç que ens permet tornar a la pantalla de navegació de Renardus on s'estava treballant. És una passarel·la molt simple i el nivell d'usabilitat és baix en el sentit de que no hi ha menús d'opcions per facilitar la navegació pel web.

Exemple 3.1. Navegació gràfica Renardus



Annex 3.3. Procés de mapeig intern de Renardus

Mapping Information Format:		
LN Local Classification Notation LC Local Classification Caption RL Relationship Code DN DDC Notation DC DDC Caption UR Local Browsing URL NT Notes :RE: RecordEnd		Legend: Local (white circle), DDC (grey circle)
1. Fully equivalent =	$\frac{DC}{LC} = \sim 100\%$ or $\frac{LC}{DC} = \sim 100\%$	<p>More or less fully, not necessarily totally congruent</p>
2. Narrower equivalent >	$\frac{DC}{LC} > 100\%$	<p>Basically the potential content of the whole local class is comprised in the DDC class: The local class is a true subset of the content of the DDC class</p>
3. Broader equivalent <	$\frac{DC}{LC} < 100\%$	<p>Basically the potential content of the whole DDC class is comprised in the local class: The local class is a true superset of the DDC class</p>
4. Major overlap MJ		<p>The local class has a major overlap with the content of the DDC class, related to the full content of the DDC class</p>
5. Minor overlap MN		<p>The local class has a minor overlap with the content of the DDC class, related to the full content of the DDC class. Normally a local class needs to be mapped to several classes.</p>

Last update: 10 May 2001

Annexes Capítol 4. Estàndards i formats

Annex 4.1. Vocabularis i espais de noms SKOS

El vocabulari de SKOS és un conjunt de URI's, tal i com s'exposa en la columna de l'esquerra de la següent taula:

Taula. Vocabulari SKOS	
URI	Definition
skos:Concept	Section 3. The skos:Concept Class
skos:ConceptScheme	Section 4. Concept Schemes
skos:inScheme	Section 4. Concept Schemes
skos:hasTopConcept	Section 4. Concept Schemes
skos:topConceptOf	Section 4. Concept Schemes
skos:altLabel	Section 5. Lexical Labels
skos:hiddenLabel	Section 5. Lexical Labels
skos:prefLabel	Section 5. Lexical Labels
skos:notation	Section 6. Notations
skos:changeNote	Section 7. Documentation Properties
skos:definition	Section 7. Documentation Properties
skos:editorialNote	Section 7. Documentation Properties
skos:example	Section 7. Documentation Properties
skos:historyNote	Section 7. Documentation Properties
skos:note	Section 7. Documentation Properties
skos:scopeNote	Section 7. Documentation Properties
skos:broader	Section 8. Semantic Relations
skos:broaderTransitive	Section 8. Semantic Relations

skos:narrower	Section 8. Semantic Relations
skos:narrowerTransitive	Section 8. Semantic Relations
skos:related	Section 8. Semantic Relations
skos:semanticRelation	Section 8. Semantic Relations
skos:Collection	Section 9. Concept Collections
skos:OrderedCollection	Section 9. Concept Collections
skos:member	Section 9. Concept Collections
skos:memberList	Section 9. Concept Collections
skos:broadMatch	Section 10. Mapping Properties
skos:closeMatch	Section 10. Mapping Properties
skos:exactMatch	Section 10. Mapping Properties
skos:mappingRelation	Section 10. Mapping Properties
skos:narrowMatch	Section 10. Mapping Properties
skos:relatedMatch	Section 10. Mapping Properties

Font: SKOS “namespace” URI: <http://www.w3.org/2004/02/skos/core#>

Nota: totes les URIs en el vocabulari de SKOS es construeixen afegint un nom local (per exemple, "prefLabel") per a l'espai de noms URI SKOS.

Veure també el resum d'SKOS que apareix en l'[Apèndix B](#) de la Guia de referència SKOS del W3C.

Annexes Capítol 5. Interoperabilitat semàntica en Sistemes d'Informació de Qualitat

Annex 5.1. Subject Gateways Interoperability

SG Title (Country) Year	Interoperability	Metadata	Tesaurus o altres Vocabularis Controlats	Others
Agrigate (Australia) 1998	OAI under consideration	Combination of Dublin Core, AGLS, EdNA and A-Core metadata elements with agricultural metadata extensions.	CABI International Agriculture Tesauro (UK), and the Australian extension of CABI Agterms (Kondinin Group)	Agriculture research. Nowadays isn't operative
<u>ATUA</u> (Australian Trade Union Archives) 2002	<i>No hi ha informació</i>	ISAD(G) - for the archival/resource descriptions; ISAAR(CPF) - for the biographical/corporate body descriptions; DC used to capture online published resources	No ho menciona. Sembla ser que tenen un llistat de paraules clau controlades per a la navegació.	Labour history research
<u>Austlit</u> Australian Literature Gateway (Australia) 2001	Interoperate through XML data	FRBR structure with mappings to XML, plain text, tagged text which can be Mapped to MARC and DC	Custom tesauro of 5000 terms	Australian authors and literacy organizations

<p>Avel Sustainability Knowledge Network (Australia) 1999</p>	<p>Z39.50: search target for the Resource Discovery Pilot</p>	<p>Dublin Core and A-Core, supplemented by some AGLS and EdNA elements <i>Thesauri:</i> Amalgamation</p>	<p>Amalgamation of the EI thesaurus for engineering and the ACM thesaurus for computer science and IT</p>	<p>Portal for engineers and other professionals concernet with sustainable systems</p>
<p>EdNA Online (Australia) 1996</p>	<p>OAI: have trialled exchange using OAI version 1 and hope to do the same with version 2 in 2003.</p>	<p>EdNA Metadata Standard V1.1. This is based on Dublin Core and is interoperable with AGLS.</p>	<p>Multiple Thesauri . SCIS, ScOT, VOCED, ATED, TAGS and AGIFT</p>	<p>Australian education and training community</p>
<p>PADI (Preserving Access to Digital Information) (Australia) 2001</p>	<p>Protocol Z39.50 per a la cerca i recuperació de la Informació</p>	<p>Dublin Core i esquema de metadades propi de PADI</p>	<p>PADI Thesaurus</p>	<p>International gateway for digital preservation resources</p>
<p>WebLaw (Australia) 1999</p>	<p>Dublin Core with some WebLaw-specific elements</p>	<p>Metadata Dublin Core with some WebLawspecific elements</p>	<p>WebLaw Thesaurus</p>	<p>Internet resources for Australian legal researchers, lawyers, students and lay professionals</p>

<p><u>INTUTE</u> (Altis, Artifact, BIOME, EEVL, Gesource, Humbul, PSigate and SOSIG) (UK) 2006</p>	<p>Protocols: IAFA(Whois++); OAI-PMH (intercanvi de metadades); i Z39.50</p>	<p>DublinCore, AARC2, RLLOMAP, RDA (Anglo-American Resource Description and Access</p>	<p>CDU, Tesauro i <i>Subject Headings</i> (navegació), entre d'altres pròpies</p>	<p>Ciència i Tecnologia; Arts i Humanitats; Ciències Socials i Ciències naturals i mèdiques</p>
<p><u>Renardus</u> (Europe) 1996</p>	<p>Z39.50 (research and retrieve protocol), Whois++; International standards such as HTML, Bath Profile, XML and RDF</p>	<p>Dublin Core</p>	<p>DDC and local classificacions and tesauro, etc.</p>	<p>Multidisciplinar; Nowadays is not operable!</p>
<p><u>VASCODA</u> (Alemanya) 2003</p>	<p>OAI harvesting protocol</p>	<p>Metadades (Dublin Core, MARC21),</p>	<p>No ho especifica</p>	<p>Estàndards de formats web (JPEG, MPEG etc.). Ús híbrid de meta-cerca i tecnologia de cerca. Futura navegació amb DDC.</p>

Annex 5.2. Qüestionari realitzat a biblioteques i centres de documentació del àmbit de les ciències aquàtiques i marines

Introduction

The main objective of this questionnaire is to know if the aquatic science librarian community is used the same *controlled vocabularies* in their information systems or not and if these systems could be compatible systems between them in a Subject Gateway project. And one of the most important things to study is the information architecture based in cross-indexing and cross-browsing.

Aquatic Sciences European Subject Gateway (ASESG) is thought as a THESIS project which will design an integrated platform of digital resources and services with high quality, within a research environment like IRTA (Research & Technology Food & Agriculture Institute, a public Institute of the Catalan Government, Spain) where I work or any institution that we can find in EURASLIC or IAMSLIC nets. This system of information management and organization try to be a European initiative, and following other representative projects and international subject gateways like RDN, Intute, Renardus, Carmen and Australian gateways between others.

Questionnaire:

1. Which kinds of *controlled vocabulary* are used in your library or information systems? Could you specify which are?
 - Thesaurus
 - Subject Headings
 - Classifications schemes
 - Others (Specify):

2. Are you using the controlled vocabularies in a digital environment?
 - Yes
 - No

3. If your answer is affirmative, could you specify which they are?

4. If you are working in a digital system (catalogues, data bases, digital libraries, etc.), are you using metadata standards like Dublin Core?
 - Yes (Specify)
 - No (Do you know any standard of metadata? Specify it!)

5. If you have built your *controlled vocabulary*, have you based with some standards, specific vocabularies, dictionaries, etc.? And why did you needed to design or adapt a *controlled vocabulary*?

- Any *controlled vocabulary* covered your collection?
- It was a cost question
- Others

Open-ended questions

In my PhD I have studied *controlled vocabularies* in 'aquatic science' and related topics (environmental, hydrobiology, biodiversity science, etc) which I could access online. But now, I need to know how the real context is.

For example, I worked in a freshwater and aquaculture library and any *controlled vocabulary* that I know covers enough both fields. Overall, because these topics are very heterogeneous and they can include chemistry, engineering, forestry or ecology subjects within others. In fact, we are using our '*controlled vocabulary list*'. So,

6. Are current controlled vocabularies enough to covers your collection in classification and indexing processes? Can you explain it why?

For people who work in a digital environment, I'd be appreciating if you can explain me other technical things that you consider important to share information with others, like:

7. Are your system support Z.39.50 interchange protocol?

8. Which online catalogue or systems are you using (comercial or not)?

9. Others

All information collected in this survey will be confidential and it will only be used for research purposes.

Thanks in advance for your help,

Carmen Reverté Reverté
Aquatic Ecosystems Unit Documentation Center
IRTA
Carme.Reverte@irta.cat

Annexes Capítol 6. Informació de qualitat (IQ): proposta d'un model d'IQ per un sistema d'informació distribuït especialitzat

Annex 6.1. Models IQ

Figura 1. Wang & Strong 1996

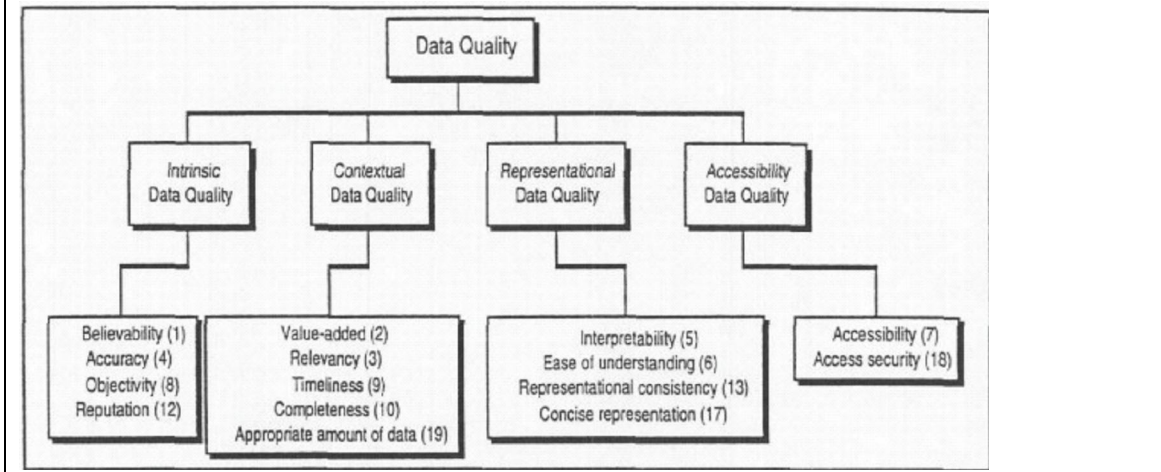


Figura 2. Bovee & Srivastava (2003)

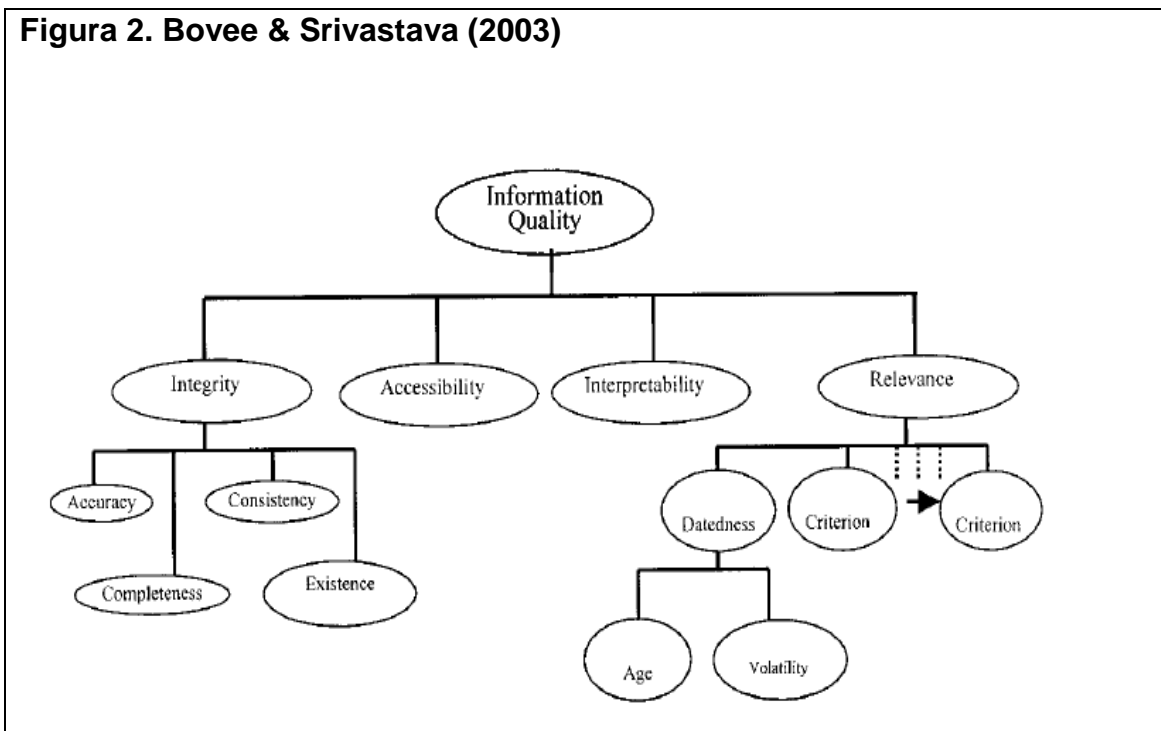


Figura 3: Stivilia, [et al.] (2007)

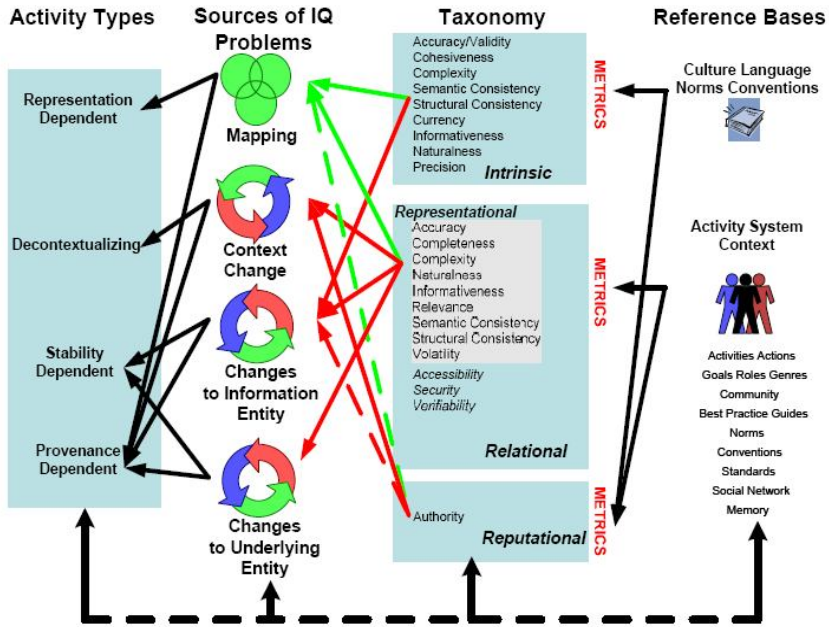


Figure 3: Conceptual model of IQ measurement

Annex 6.2. Taules comparatives d'autors i dimensions d'IQ

Annex 6.2. Taula comparativa d'autors i dimensions d'IQ (1)				
Autors	Models	Categories	Dimencions	Descripció
Stivilia [et a] (2007)	Conceptual	Intrínseca	Accuracy/Validity Cohesiveness Complexity Semantic Consistency Structural Consistency Currency Informativeness Naturalness Precision	En un principi hi havien 32 atributs o dimensions però varen passar a 22. La característica d'aquest model es que dibuixa també els processos que es duen a terme amb el Sistema d'Informació i les problemàtiques que sorgeixen amb els recursos d'informació que proporciona el sistema i les seves activitats.
		Representational	Accuracy Completeness Complexity Naturalness Informativeness Relevance Semantic Consistency Structural Consistency Volatility	
		Relacional	Accessibility Security Verifiability	
		Reputacional	Authority	
Intute	Criteria	Contextual	Audiència Autoritat Comparativa de recursos	
		Contingut	Cobertura Precisió Actualització	

		Format	Accessibilitat Disseny i Etiquetatge Fàcil d'ús
--	--	--------	---

Annex 6.2. Taula comparativa d'autors i dimensions d'IQ (2)				
Autors	Models	Categories	Dimencions	Descripció
Wang & Strong (1996)	Conceptual	Intrínsec	Accuracy Objectivity Believability Reputation	El model representa 15 dimensions perquè són les que els usuaris varen considerar com a més importants però el primer model contenia 20 dimensions que representaven als 179 atributs que havien aparegut de les mostres recollides en els estudis d'usuaris.
		Accessibility	Accessibility Access security	
		Contextual	Relevancy Value-added Timeliness Completeness	
		Representational	Interpretability Ease of understanding Concise representation Consistent representation	
Bovee, M. [et al] (2003)	Conceptual	Integrity (Intrínseca)	Accuracy, Completeness; Consistency, Existency	
		Accessibility (extrínseca)	Timeliness	
		Interpretabilitat (extrínseca)	Understandable, Sense, Significance and Meaning	
		Rellevància (extrínseca)	Dateness (Age; Volatility); Criterion	
Naumann & Rolker (2000)	Classificació (no model)	Subject-Criteria	Believability, Concise representation, Interpretability, Relevancy, Reputation, Understandability, Value-Added	Tres classes amb 22 dimensions (que ell anomena criteris d'IQ)
		Object-Criteria	Completeness, Customer Support, Documentation, Objectivity, Price, Reliability, Security, Timeliness, Verifiability	

		Process-Criteria	Accuracy, Amount of data, Availability, Consistent representation, Latency, Response time
--	--	------------------	---

Annex 6.2. Taula comparativa d'autors i dimensions d'IQ (4)

Autors	Models	Categories	Dimencions	Descripció
Gonçalves, et al. (2007)	Conceptual	Digital object	Accessibility	Societies (actor), Structures (metadata specification), Streams + Structures (structured streams)
			Pertinence	Societies (actor), Scenarios (task)

Annex 6.2. Taula comparativa d'autors i dimensions d'IQ (3)

Autors	Models	Categories	Dimencions	Descripció
Desire (1996-2000)	Criteria	Scope-Criteria (User)	Coverage Access	
		Content-Criteria	Validity Authority/Reputation of Source Substantiveness Accuracy Comprehensiveness Uniqueness Composition and Organisation Currency and Adequacy of Maintenance	
		Form-Criteria (Medium)	Ease of Navigation Use of Recognised Standards Appropriate use of Technology Aesthetics	
		Process-Criteria	Information Integrity Site Integrity System Integrity	
		Collection-Criteria (Service)	Coverage and Balance Availability of Internet Resources Availability of Internet Services	

	(Biblioteca Digital). Les dimensions associades al concepte '5S' (Streams, Structures, Spaces, Scenaris, Societies)		Preservability	Streams, Structures (structural metadata), Scenarios (process (e.g., migration))	
			Relevance	Streams + Structures (structured streams), Structures (query), Spaces (Metric, Probabilistic, Vector)	
			Similarity	Same as in relevance, Structures (citation/link patterns)	
			Conformance	Structure (properties, schema)	
		Collection		Completeness	Structure (collection)
				Consistency	Structure (collection)
		Catalog		Completeness	Structure (collection)
				Consistency	Structure (collection)
		Repository		Completeness	Structure (collection)
				Consistency	Structure (catalog, collection)
		Services		Composability	See Extensibility, reusability
				Efficiency	Streams (time), Spaces (operations, constraints)
				Effectiveness	See Pertinence, Relevance
	Extensibility		Societies + Scenarios (extends, inherits_from, redefines)		
	Reusability		Societies + Scenarios (includes, reuses)		
		Reliability	Societies + Scenarios (uses, executes, invokes)		

Annex 6.2. Taula comparativa d'autors i dimensions d'IQ (5)				
Autors	Models	Categories	Dimencions	Descripció
Codina, 2003	Criteris (llistat d'atributs que anomena com a indicadors)	Contingut (I)	1.1 Identificación y solvencia de la fuente 1.2 Adecuación (entre autoría y agenda temática) 1.3 Comunicación (Contacto) 2 Contenido: 2.1 Tema, público y objetivos 2.2. Interés intrínsec 2.3.Originalidad/Oportunidad	Fase 1 i 2 (Escenari 1, 2 i 3)

	que són agrupats en 4 seccions o dimensions)		<p>2.4. Política editorial</p> <p>2.5. Cantidad (volumen de información)</p> <p>2.6 Rigor (veracidad)</p> <p>2.7 Edición (presentación)</p> <p>2.8 Actualización</p> <p>2.9 Recursos multimedia</p> <p>2.10 Recursos interactivos</p> <p>2.11 Archivo/Hemeroteca</p>	
Codina, 2003		<p>Navegació (II):</p> <p>3 - Acceso a la información: Navegación y Recuperación.</p> <p>4 – Ergonomia: comodidad y facilidad de utilización</p>	<p>3.Navegación y Recuperación de la Información</p> <p>Menú o Sumario</p> <p>3.2 Expresividad (opciones de presentación de contenidos)</p> <p>3.3 Identificación (Identificación básica de secciones: título, autor (o fuente) y fecha)</p> <p>3.4 Recorrido secuencial</p> <p>3.5 Navegación estructural</p> <p>3.6 Orientación (Indicaciones de Contexto)</p> <p>3.7 Jerarquización (importancia de las secciones)</p> <p>3.8 Sumarios locales (secciones)</p> <p>3.9 Índices (contenido o temas en forma de listas)</p> <p>3.10 Navegación Semántica (enlaces entre secciones no relacionadas estructuralmente)</p> <p>3.11 Sistemas de etiquetas (usadas en las secciones)</p> <p>3.12 Recuperación información o Interrogación (sistemas de búsqueda por palabras clave, etc.)</p> <p>3.13 Búsqueda avanzada</p> <p>3.14 Lenguaje documental (CDU, descriptores, etc.)</p> <p>4.Ergonomia</p> <p>4.1 Facilidad de utilización</p> <p>4.2 Flexibilidad</p>	Fase 1 i 2 (Escenaris 2 i 3)

			<p>4.3 Claridad</p> <p>4.4 Legibilidad</p> <p>4.5 Recursos Multimedia</p> <p>4.6 Velocidad (descarga)</p>	
Codina, 2003		Visibilitat (III)	<p>5.Luminosidad</p> <p>5.1 Enlaces (a recursos externos)</p> <p>5.2 Contexto de anclaje (aparición de los anclajes donde actuan los enlaces)</p> <p>5.3. Anticipación (información sobre el enlace)</p> <p>5.4 Oportunidad (adequación de los enlaces)</p> <p>5.5 Calidad enlaces</p> <p>5.6 Actualización</p> <p>5.7 Tratamiento (tipos de enlaces, directorios, etc)</p> <p>6.Ubiquidad</p> <p>6.1 Título (etiqueta html <HEAD> del recurso)</p> <p>6.2 Transparencia (información sobre la temática principal del recurso)</p> <p>6.3 Meta información (etiquetas metadatos incluidas en la sección <HEAD>)</p> <p>6.4 Dublin Core (etiquetas incrustadas en el HEAD)</p> <p>6.5. Popularidad (citaciones que recibe el recurso)</p>	Fase 1 i 2 (Escenaris 2 i 3)
		Usabilitat (IV)		Fase 1 i 2 (Escenaris 2 i 3)

			<ul style="list-style-type: none">7. Procesos<ul style="list-style-type: none">7.1 Visión de Estatus (cuando se realiza una acción, si se puede ver estado del proceo)7.2 Convenciones (uso de lenguaje y convenciones para los usuarios)8. Errores<ul style="list-style-type: none">8.1 Deshacer acciones8.2 Mensajes (en caso de error)8.3 Consecuencias de errores9. Adaptación<ul style="list-style-type: none">9.1 Adaptación (Capacidad de personalización y adaptación a cada usuario)9.2 Redundancia (Capacidad para hacer las mismas acciones por distintos medios)9.3 Acceso (facil de acceder)9.4 Política (políticas y normas del recurso)	
--	--	--	--	--

Annex 6.3. Taula de valors de dimensions d'IQ

Dimensions IQ	Miller	Wang	Bovee	Naumann	Stvilia	Gonçalves	Intute	Desire
Access security		x						
Accessibility	x	x			x	x	x	x
Accuracy	x	x	x	x	x	x	x	x
Adequacy of Maintenance								x
Aesthetics								x
Age			x					
Amount of data				x				
Appropriate use of thecnology								x
Audience							x	
Authority/Reputation					x		x	x
Availability (sources and services)				x				x
Believability		x		x				
Coherence	x							
Comparative of resources							x	
Compatibility								
Completeness	x	x	x	x	x	x		
Complexity					x			
Composability (strcture)						x		
Composition and Organization								x
Comprehensiveness								x
Concise representation		x		x				
Conformance						x		
Consistency (structure)						x		
Consistent representation		x	x	x				
Coverage							x	x
Criterion			x					
Currency							x	x
Customer Support				x				

Dateness			x					
Design and Layout							x	
Documentation				x				
Ease of Navigation								x
Ease of understanding								
Ease of use		x					x	
Effectiveness						x		
Efficency						x		
Existency			x					
Extensibility						x		
Format	x							
Information integrity								x
Informativeness					x			
Interpretability		x		x				
Latency				x				
Meaning			x					
Naturalness					x			
Objectivity		x		x				
Pertinence						x		
Price				x				
Relevancy	x	x		x	x			
Reliability				x		x		
Reputation		x		x				
Response a time				x				
Reusability						x		
Security	x			x	x			
Semantic Consistency					x			
Sense			x					
Significance			x					
Similarity						x		
Site Integrity								x
Structural Consistency					x			
Substantiveness								x

System Integrity								x
Timeliness	x	x	x	x		x		
Undersatable				x				
Uniqueness								x
Use of Recognised Standards								x
Validity	x							x
Value-added		x		x				
Verifiability					x			
Volatility			x		x			

Annex 6.4. Elements Bàsics Dublin Core¹¹⁹

ELEMENTS	DESCRIPCIÓ
Títol	<p>Nom: Títol</p> <p>Identificador:Title</p> <p>Definició: Nom donat al recurs.</p> <p>Comentari: Normalment, el Títol és un nom pel qual es coneix formalment el recurs.</p>
Creador	<p>Nom: Creador</p> <p>Identificador:Creator</p> <p>Definició: Entitat que té la responsabilitat principal de la creació del contingut del recurs.</p> <p>Comentari: Els exemples de Creador inclouen una persona, una organització o un servei. Normalment, el nom d'un Creador s'ha d'utilitzar per a indicar l'entitat.</p>
Matèries i Paraules clau	<p>Nom: Matèries i Paraules clau</p> <p>Identificador:Subject</p> <p>Definició: El tema del contingut del recurs.</p> <p>Comentari: Normalment, les Matèries s'expressen amb paraules clau o frases clau o codis de classificació que descriuen el tema del recurs. Es recomana la bona pràctica de seleccionar un valor d'un vocabulari</p>

¹¹⁹ Conjunt dels elements de metadades del Dublin Core, Versió 1.1: Descripció de referència . [Traducció]. Biblioteca de Catalunya, 2002. DL B-46201-2002. URL:

<http://www.bnc.cat/sprof/normalitzacio/dc/dces.php>; **Uri:** <http://dublincore.org/documents/dces/>

	controlat o d'un esquema de classificació formal.
Descripció	<p>Nom: Descripció</p> <p>Identificador:Description</p> <p>Definició: Explicació del contingut del recurs.</p> <p>Comentari: La Descripció pot incloure un resum, un sumari, una referència a la representació gràfica del contingut o un text lliure del contingut, tot i que no hi està limitada.</p>
Editor	<p>Nom: Editor</p> <p>Identificador:Publisher</p> <p>Definició: Entitat responsable de fer el recurs disponible.</p> <p>Comentari: Els exemples d'Editor inclouen una persona, una organització o un servei. Normalment, el nom d'un Editor s'utilitza per a indicar l'entitat mateixa.</p>
Contribuïdor	<p>Nom: Contribuïdor</p> <p>Identificador:Contributor</p> <p>Definició: Entitat responsable de fer contribucions al contingut del recurs.</p> <p>Comentari: Els exemples de Contribuïdor inclouen una persona, una organització o un servei. Normalment, el nom d'un Contribuïdor s'utilitza per a indicar l'entitat mateixa.</p>
Data	<p>Nom: Data</p> <p>Identificador:Date</p> <p>Definició: Data associada a un esdeveniment del cicle de vida del recurs.</p> <p>Comentari: Normalment, la Data s'associa a la creació o disponibilitat del recurs. Es recomana la bona pràctica d'utilitzar el perfil d'ISO 8601 [W3CDTF], que segueix la forma AAAA-MM-DD, per a la codificació del valor de la data.</p>
Tipus de recurs	<p>Nom: Tipus de recurs</p> <p>Identificador:Type</p> <p>Definició: Naturalesa o gènere del contingut del recurs.</p> <p>Comentari: Tipus inclou termes que descriuen categories, funcions i gèneres generals per al contingut, o nivells d'agregació. Es recomana la bona pràctica de seleccionar un valor d'un vocabulari controlat (per exemple, la llista de treball del Dublin Core Types [DCT1]). Per a descriure una manifestació física o digital del recurs s'utilitza l'element FORMAT.</p>
Format	<p>Nom: Format</p> <p>Identificador:Format</p> <p>Definició: Manifestació física o digital del recurs</p> <p>Comentari: Normalment, el Format pot incloure el tipus de suport o dimensions del recurs. El Format es pot utilitzar per determinar el programari, el maquinari o qualsevol</p>

	<p>altre equipament necessari per a la visualització del recurs o per a fer-lo funcionar. Exemples de dimensions inclouen la mida i la durada. Es recomana la bona pràctica de seleccionar un valor d'un vocabulari controlat (per exemple, la llista d'Internet Media Types [MIME] que defineix els formats de suports electrònics.</p>
Identificador del recurs	<p>Nom: Identificador del recurs Identificador: Identifier Definició: Referència unívoca al recurs en un context donat. Comentari: Es recomana la bona pràctica d'identificar el recurs per mitjà d'una seqüència o un número d'acord a un sistema d'identificació formal. Els exemples de sistemes d'identificació formal inclouen l'URI (Uniform Resource Identifier) (inclòs l'URL (Uniform Resource Locator), el DOI (Digital Object Identifier) i l'ISBN (International Standard Book Number).</p>
Font	<p>Nom: Font Identificador: Source Definició: Referència a un recurs del qual se'n deriva el recurs present. Comentari: El recurs present pot derivar del recurs font en la seva totalitat o en part. Es recomana la bona pràctica de fer referència al recurs per mitjà d'una seqüència o número d'acord a un sistema d'identificació formal.</p>
Llengua	<p>Nom: Llengua Identificador: Language Definició: Llengua del contingut intel·lectual del recurs Comentari: Per als valors de l'element Llengua es recomana el que és definit a RFC 1766 [RFC1766] on s'inclou un codi de llengua de dos caràcters (derivat de l'estàndard ISO 639 [ISO639]), seguit opcionalment d'un codi de país de dos caràcters (derivat de l'estàndard ISO 3166 [ISO3166]). Per exemple, "en" per a l'anglès, "fr" per al francès, "ca" per al català o "en-uk" per a l'anglès del Regne Unit.</p>
Relació	<p>Nom: Relació Identificador: Relation Definició: Referència al recurs relacionat. Comentari: Es recomana la bona pràctica de fer referència al recurs per mitjà d'una seqüència o número d'acord a un sistema d'identificació formal.</p>
Cobertura	<p>Nom: Cobertura Identificador: Coverage</p>

	<p>Definició: Extensió o abast del contingut del recurs.</p> <p>Comentari: Normalment, la Cobertura inclou la localització espacial (el nom d'un lloc o les coordenades geogràfiques), el període de temps (el nom d'un període, una data o un interval de dates) o una jurisdicció (com ara el nom d'una entitat administrativa). Es recomana la bona pràctica de seleccionar un valor d'un vocabulari controlat (per exemple, el Thesaurus of Geographic Names [TGN]) i, quan sigui possible, utilitzar els noms de lloc o períodes de temps preferentment als identificadors numèrics com ara sèries de coordenades o intervals de temps.</p>
Gestió dels drets	<p>Nom: Gestió dels drets</p> <p>Identificador: Rights</p> <p>Definició: Informació dels drets exercits sobre el recurs</p> <p>Comentari: Normalment, l'element Drets inclou una indicació de gestió dels drets sobre el recurs o una referència a un servei que proporciona aquesta informació. La informació sobre els drets sovint comprèn els drets de propietat intel·lectual Intellectual Property Rights (IPR), Copyright i altres drets sobre la propietat. Si l'element Rights és absent, no es poden fer assumpcions sobre l'estatus d'aquests i d'altres drets amb respecte al recurs.</p>

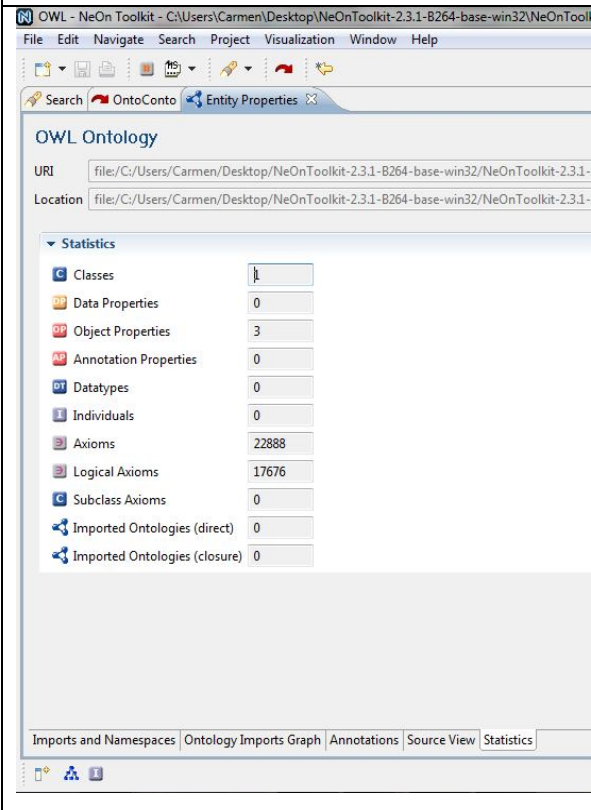
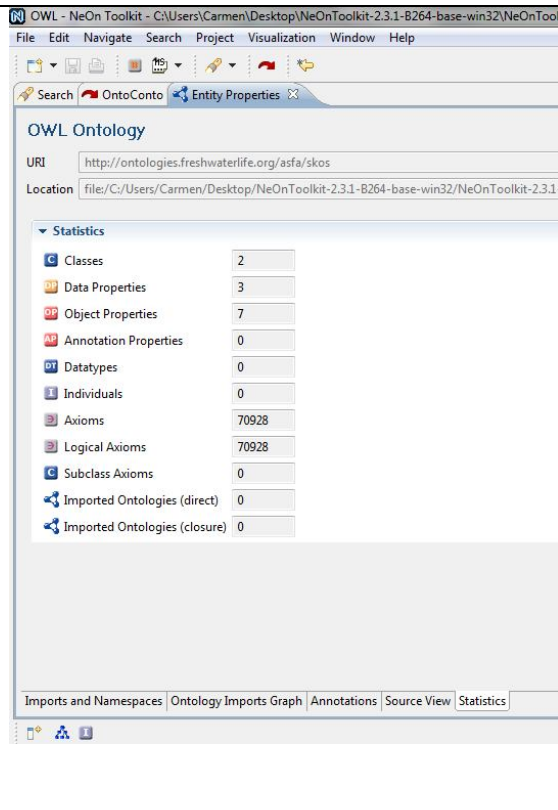
Annexes Capítol 7. Sistemes de Mapeig de vocabularis controlats en ciències aquàtiques

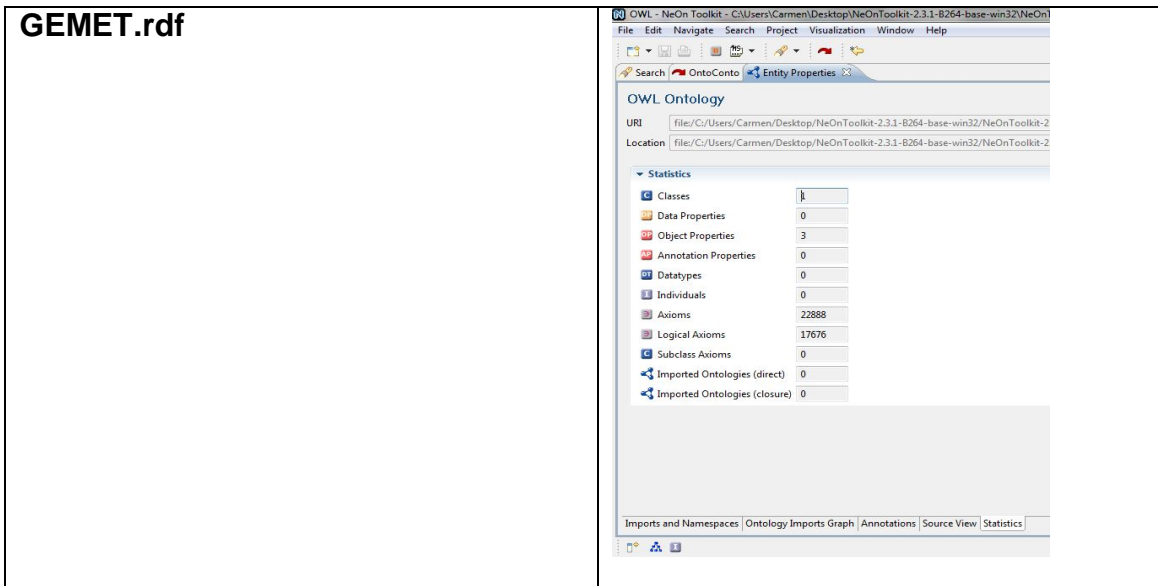
Annex 7.1. Tests amb NeOnToolkit

Les proves realitzades amb el sistema es realitzen entre el 2010 i 2011.

- **Test 1.** Introducció dels tesaurus AGROVOC, ASFA I GEMET en el sistema

En aquesta fase s'observa si el sistema accepta les ontologies. I també les estadístiques generals que genera NeOnToolkit:

AGROVOC.rdf	ASFA.rdf																																																
 <table border="1"><caption>Statistics for AGROVOC.rdf</caption><thead><tr><th>Category</th><th>Count</th></tr></thead><tbody><tr><td>Classes</td><td>1</td></tr><tr><td>Data Properties</td><td>0</td></tr><tr><td>Object Properties</td><td>3</td></tr><tr><td>Annotation Properties</td><td>0</td></tr><tr><td>Datatypes</td><td>0</td></tr><tr><td>Individuals</td><td>0</td></tr><tr><td>Axioms</td><td>22888</td></tr><tr><td>Logical Axioms</td><td>17676</td></tr><tr><td>Subclass Axioms</td><td>0</td></tr><tr><td>Imported Ontologies (direct)</td><td>0</td></tr><tr><td>Imported Ontologies (closure)</td><td>0</td></tr></tbody></table>	Category	Count	Classes	1	Data Properties	0	Object Properties	3	Annotation Properties	0	Datatypes	0	Individuals	0	Axioms	22888	Logical Axioms	17676	Subclass Axioms	0	Imported Ontologies (direct)	0	Imported Ontologies (closure)	0	 <table border="1"><caption>Statistics for ASFA.rdf</caption><thead><tr><th>Category</th><th>Count</th></tr></thead><tbody><tr><td>Classes</td><td>2</td></tr><tr><td>Data Properties</td><td>3</td></tr><tr><td>Object Properties</td><td>7</td></tr><tr><td>Annotation Properties</td><td>0</td></tr><tr><td>Datatypes</td><td>0</td></tr><tr><td>Individuals</td><td>0</td></tr><tr><td>Axioms</td><td>70928</td></tr><tr><td>Logical Axioms</td><td>70928</td></tr><tr><td>Subclass Axioms</td><td>0</td></tr><tr><td>Imported Ontologies (direct)</td><td>0</td></tr><tr><td>Imported Ontologies (closure)</td><td>0</td></tr></tbody></table>	Category	Count	Classes	2	Data Properties	3	Object Properties	7	Annotation Properties	0	Datatypes	0	Individuals	0	Axioms	70928	Logical Axioms	70928	Subclass Axioms	0	Imported Ontologies (direct)	0	Imported Ontologies (closure)	0
Category	Count																																																
Classes	1																																																
Data Properties	0																																																
Object Properties	3																																																
Annotation Properties	0																																																
Datatypes	0																																																
Individuals	0																																																
Axioms	22888																																																
Logical Axioms	17676																																																
Subclass Axioms	0																																																
Imported Ontologies (direct)	0																																																
Imported Ontologies (closure)	0																																																
Category	Count																																																
Classes	2																																																
Data Properties	3																																																
Object Properties	7																																																
Annotation Properties	0																																																
Datatypes	0																																																
Individuals	0																																																
Axioms	70928																																																
Logical Axioms	70928																																																
Subclass Axioms	0																																																
Imported Ontologies (direct)	0																																																
Imported Ontologies (closure)	0																																																



Taula 1. Imatges de les importacions dels tesaurus en el sistema NeOn Toolkit

Els resultats del test 1, mostren com el sistema no identifica correctament els tipus de dades de les ontologies importades, probablement perquè es tracta de tesaurus convertits a ontologies. Cal fixar-se amb els camps de “class”, “data properties” i “object properties”.

➤ **Test 2. Tests amb alineaments des de NeOn Toolkit**

El sistema dona errors continuament amb els “plugins” d’alineaments. Així doncs, es deixen de fer proves amb els vocabularis objecte (AGROVOC, ASFA i GEMET), per veure si el problema és que són “large ontologies”, ja que l’error sembla que té a veure amb l’execució del java: “Unhandled loop exception; Java heap space”

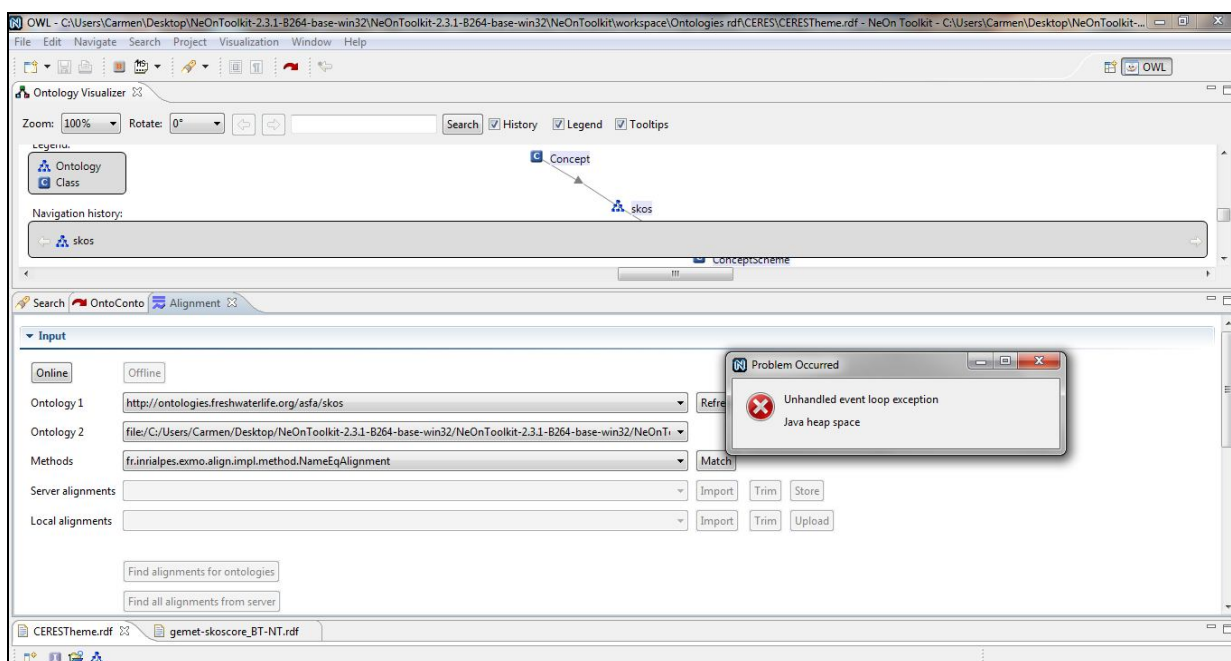
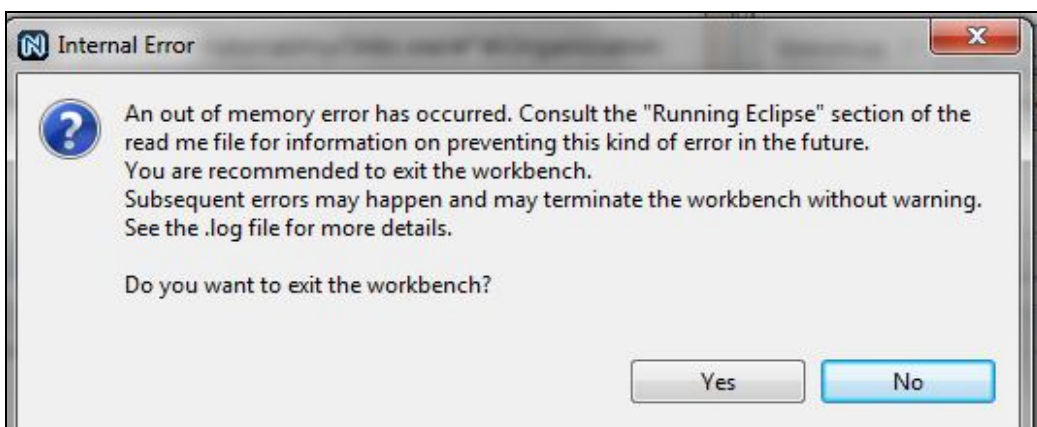
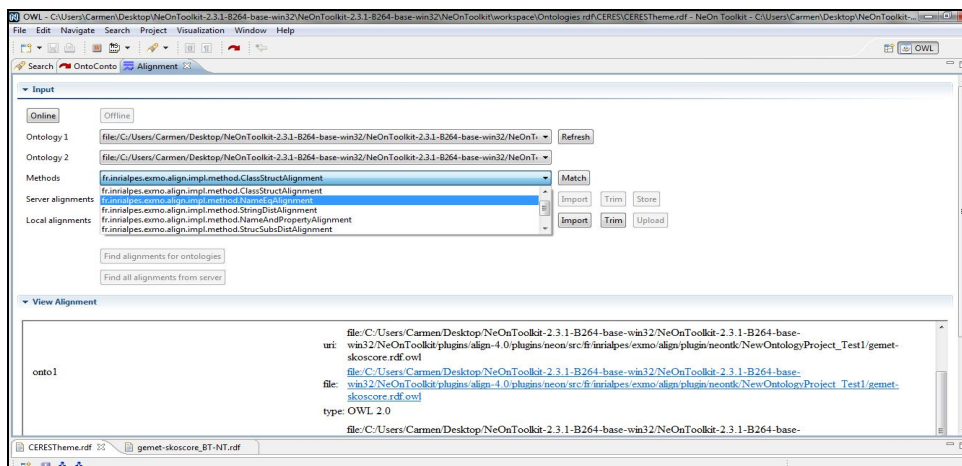


Figura 1. Exemple d'errors

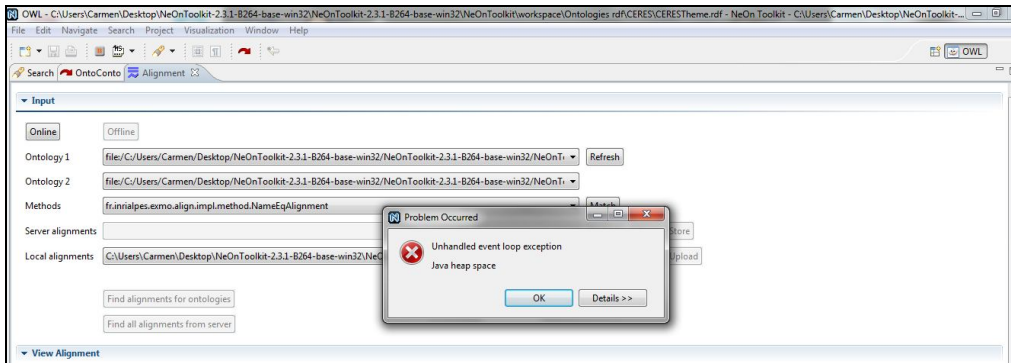
- **Test 3:** vist els errors anteriors es realitza un altre test amb els vocabularis GEMET i CERES. Aquest cop el sistema d'alineament funciona, no obstant, no s'obtenen resultats per tots els mètodes d'aineament.
 - **1r Alignment method – StrucDistNameAlignment:** dona error de memòria, com si el sistema no tinguès suficient capacitat per mostrar els resultats. Cal pensar que estem davant de vocabularis amb una extensió molt gran com són els tesaurus.



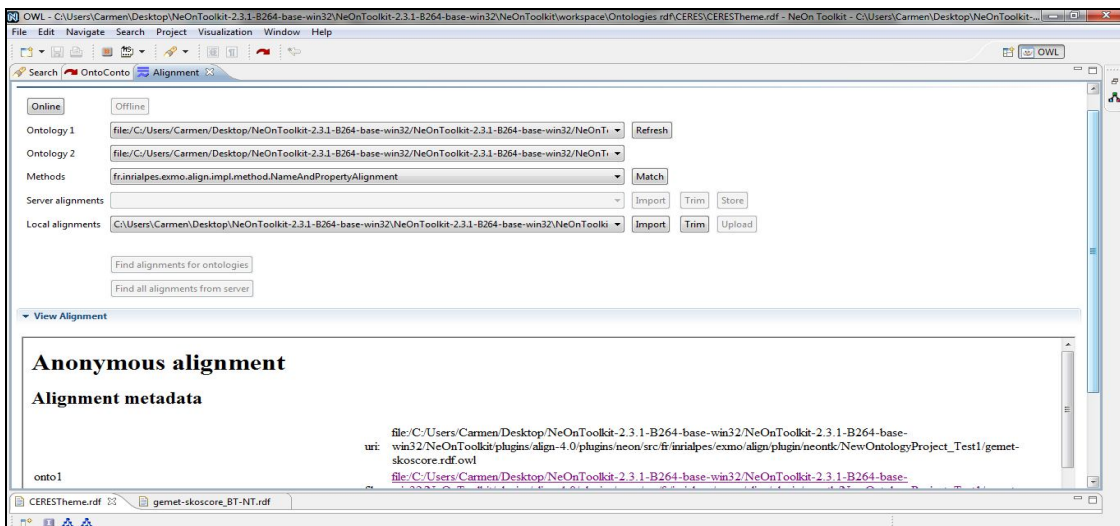
- **2n Alignment method – ClassStructAlignment:** funciona, però la visualització dels alineaments no és molt usable



- **3r Alignment method – NameEqAligment:** no funciona



- **4rt Alignment method – NameAndPropertyAlignment:** funciona correctament



- **5è Alignment method – StringDistAlignment:** Mètode d'alineament que funciona, només apareix 1 alineament:

Onto1	Uri:	file:/C:/Users/Carmen/Desktop/NeOnToolkit-2.3.1-B264-base-win32/NeOnToolkit-2.3.1-B264-base-win32/NeOnToolkit/plugins/align-4.0/plugins/neon/src/fr/inrialpes/exmo/align/plugin/neontk/NewOntologyProject_Test1/gemet-skoscore.rdf.owl
	File:	file:/C:/Users/Carmen/Desktop/NeOnToolkit-2.3.1-B264-base-win32/NeOnToolkit-2.3.1-B264-base-win32/NeOnToolkit/plugins/align-4.0/plugins/neon/src/fr/inrialpes/exmo/align/plugin/neontk/NewOntologyProject_Test1/gemet-

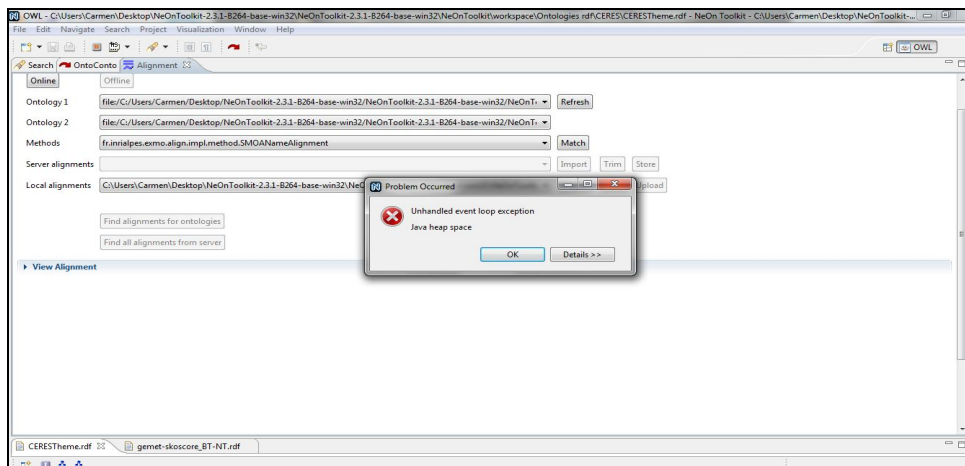
		skoscore.rdf.owl
	Type:	OWL 2.0
Onto2	Uri:	file:/C:/Users/Carmen/Desktop/NeOnToolkit-2.3.1-B264-base-win32/NeOnToolkit-2.3.1-B264-base-win32/NeOnToolkit/plugins/align-4.0/plugins/neon/src/fr/inrialpes/exmo/align/plugin/neontk/NewOntologyProject_Test1/gemet-skoscore.rdf.owl
	File:	file:/C:/Users/Carmen/Desktop/NeOnToolkit-2.3.1-B264-base-win32/NeOnToolkit-2.3.1-B264-base-win32/NeOnToolkit/plugins/align-4.0/plugins/neon/src/fr/inrialpes/exmo/align/plugin/neontk/NewOntologyProject_Test1/gemet-skoscore.rdf.owl
	Type:	OWL 2.0
Level		0
Type		**
http://knowledgeweb.semanticweb.org/heterogeneity/alignment : method		fr.inrialpes.exmo.align.impl.method.StrucSubDistAlignment

- Correspondències exemple mètode 5:

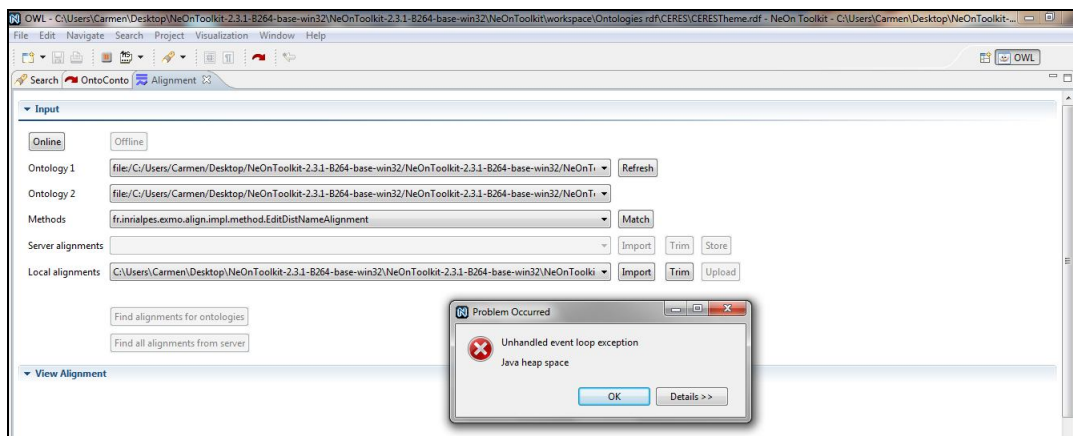
Object1	Relation	Strength	Object2
http://www.w3.org/2004/02/skos/core#narrower	=	1.0	http://www.w3.org/2004/02/skos/core#narrower
http://www.w3.org/2004/02/skos/core#broader	=	1.0	http://www.w3.org/2004/02/skos/core#broader
http://www.w3.org/2004/02/skos/core#Concept	=	1.0	http://www.w3.org/2004/02/skos/core#Concept
http://www.w3.org/2004/02	=	1.0	http://www.w3.org/2004/02/

/skos/core#related			skos/core#related
--------------------	--	--	-------------------

➤ **6è Alignment method - SMOAName Alignment: dóna error**



➤ **7è Alignment Method - EditDistAlignment : dóna error**



▪ **Test 4:** Es realitza una altra prova, on s'agafa un altre parell de vocabularis diferent que en els tests anteriors i que són també d'una mida més petita, "Biodiversity Ontology" and "CERES Ontology". El sistema executa tots els mètodes d'alineament sense errors, però només en quatre d'ells s'obtenen alineaments. Cal afegit que les alineacions han sigut totes de tipus estructural, però no s'obtenen alineaments conceptuals. Mètodes emprats i correspondències dels alineaments trobats:

- NameAndPropertyAlignment: 3 correspondències
- SMOAName Alignment: 9 correspondències
- EditDistAlignment: 10 correspondències
- SubstSistName: 10 correspondències
- NameEqAlignment: 0
- StringDistAlignment: 0
- ClassStructAlignment: 0

Annexes Capítol 8. Prototipus de mapeig: estudi experimental d'alineació de tesaurus en ciències aquàtiques i afins

Annex 8.1. ASFA, AGROVOC i GEMET

La senyalització per colors representa els termes equivalents trobats en la resta de vocabularis analitzats. Cada color correspon al concepte trobat com a equivalent en un segon i/o tercer vocabulari. Si la celda no té un color representat és que no hi ha terme equivalent, i per tant, és visualitzen les diferències quant a cobertura temàtica.

ASFA (T1)	AGROVOC	GEMET
Aqüicultura	Aquaculture	Aquaculture
UF Industria acuícola	UF 37595 - Sea ranching (EN)	broader terms
UF Agricultura acuàtica	UF 37183 - this term does not exist in EN	animal husbandry
UF Acuicultura	UF 8899 - Aquiculture (EN)	narrower terms
NT Aqüicultura d'aigües salobres	BT 8532 - Animal husbandry (EN)	fish farming
NT Aqüicultura d'aigua dolça	BT 2934 - fisheries (EN)	mariculture
NT Aqüicultura marina	NT 15649 - Mariculture (EN)	mussel farming
RT Desenvolupament en Aqüicultura	NT 9001000 - aquaponics (EN)	shellfish farming
RT Economia en l'aqüicultura	NT 8022 - Turtle culture (EN)	spawning ground
RT Enginyeria en l'aqüicultura	NT 7319 - Sponge culture (EN)	related terms
RT Facilitats en l'aqüicultura	NT 7034 - Shellfish culture (EN)	environmental impact of aquaculture
RT Productes en l'aqüicultura	NT 6914 - Seaweed culture (EN)	
RT Regulacions en aqüicultura	NT 5642 - Pearl culture (EN)	Scope note:
RT Estadística en l'aqüicultura	NT 3111 - Frog culture (EN)	scope note is not available
RT Tècniques d'aqüicultura	NT 2918 - Fish culture (EN)	Groups:

RT Aqüicultors	RT 92335 - Aquatic sciences (EN)	AGRICULTURE, FORESTRY; ANIMAL HUSBANDRY; FISHERY
RT Ciències aquàtiques	RT 2929 - Fish ponds (EN)	Themes:
RT Criança	RT 36678 - Fish farms (EN)	animal husbandry
Etc.	RT 28992 - Aquaculture techniques (EN)	fishery
	RT 25745 - Aquaculture equipment (EN)	water
	RT 13936 - Rizipisciculture (EN)	
	RT 212 - Agropisciculture (EN)	
Biologia	Biologia	Biologia
<i>Utilitzar per</i>	UF 27677 - this term does not exist in EN	broader terms
<i>Ciències biològiques</i>	UF 69947 - this term does not exist in EN	life science
<i>Ciències biològiques (biologia)</i>	UF+ 27735 - this term does not exist in EN	narrower terms
NT Anatomia	UF+ 27676 - this term does not exist in EN	aerobiology
NT Botànica	BT 4318 - life sciences (EN)	chorology
NT Criobiologia	NT 7160 - Soil biology (EN)	estuarine biology
NT Citologia	NT 8529 - Zoology (EN)	freshwater biology
NT Embriologia	NT 4800 - Microbiology (EN)	human biology
NT Biologia pesquera	NT 4891 - Molecular biology (EN)	hydrobiology
NT Biologia funcional	NT 5088 - natural history (EN)	limnology
NT Genètica	NT 5509 - Palaeontology (EN)	marine biology
NT Hematologia	NT 10587 - Embryology (EN)	microbiology
NT Histologia	NT 8812 - Animal biology (EN)	molecular biology
NT Hidrobiologia	NT 6427 - Radiobiology (EN)	morphology
NT Microbiologia	NT 2937 - Fishery biology (EN)	parasitology
NT Biologia molecular	NT 3808 - Immunology (EN)	physiology
RT Bioacústica	NT 49903 - Morphology (EN)	related terms

RT Biofísica	NT 49890 - anatomy (EN)	anatomy
RT Biogeografia	NT 1023 - Botany (EN)	
RT Biogeoquímica	NT 5578 - Parasitology (EN)	Scope note:
RT Ecologia	NT 2091 - cytology (EN)	scope note is not available
RT Cicle Vital	NT 49881 - Physiology (EN)	Groups:
	NT 2467 - ecology (EN)	RESEARCH, SCIENCES
	NT 37958 - Bioinformatics (EN)	Themes:
	NT 3056 - Forestry biology (EN)	biology
	NT 3222 - genetics (EN)	research
	NT 3718 - Hydrobiology (EN)	
	NT 12399 - marine biology (EN)	
	RT 29632 - immunotaxonomy (EN)	
	RT 24389 - Biophysics (EN)	
	RT 16165 - Biotechnology (EN)	
	RT 10598 - Endocrinology (EN)	
	RT 915 - Biogeography (EN)	
	RT 929 - biosystematics (EN)	
	RT 2097 - cytotaxonomy (EN)	
	RT 2745 - evolution (EN)	
	RT 868 - Behaviour (EN)	
	RT 1526 - chemotaxonomy (EN)	
	RT 5981 - plant sciences (EN)	
	RT 7631 - taxonomy (EN)	
	RT 4695 - medical sciences (EN)	
ciències aquàtiques	Aquatic Sciences	Aquatic Sciences
NT Ciències de l'aigua dolça	RT 910 - biochemistry (EN)	
NT Ciències marines	Aquatic Sciences	No existeix!
NT Limnologia	BT 37989 - sciences (EN)	
RT Aqüicultura	NT 4343 - Limnology (EN)	

RT Biologia	NT 4613 - Marine sciences (EN)	
RT Hidrosfera	RT 550 - aquaculture (EN)	
RT Medi ambient aquàtic	RT 2438 - Earth sciences (EN)	
Ecologia	Ecology	Ecology
UF Ecologia Aquàtica	UF 10632 - Environmental biology (EN)	broader terms
Bionòmica	UF 10530 - Ecological sciences (EN)	biosphere
Ciències ecològiques	UF 17647 - this term does not exist in EN	life science
NT Autoecologia	UF 20124 - this term does not exist in EN	narrower terms
Ecologia de aigües salobres/estuarines	UF 21971 - this term does not exist in EN	autoecology
Etologia	UF 9270 - Bionomics (EN)	ecological parameter
Ecologia d'aigua dolça	UF 66964 - this term does not exist in EN	ecosystem type
Genecologia	BT 925 - biology (EN)	synecology
Ecologia marina	NT 35783 - Radiation ecology (EN)	
Paleoecologia	NT 35784 - Radioecology (EN)	Scope note:
Parasitologia	NT 24111 - Microbial ecology (EN)	scope note is not available
Fitosociologia	NT 37336 - Population ecology (EN)	Groups:
Plànctologia (UF: estudis planctònics)	NT 37526 - Chemical ecology (EN)	BIOSPHERE (organisms, ecosystems)
Radioecologia	NT 427 - Animal ecology (EN)	Themes:
Sinecologia	NT 34279 - Fire ecology (EN)	biology
RT Biofàcies	NT 7571 - Synecology (EN)	research
Biogeografia	NT 5963 - Plant ecology (EN)	
Biologia	NT 720 - Autecology (EN)	
Condicions mediambientals	NT 92381 - agroecology (EN)	
Distribució ecològica	NT 3044 - Forest ecology (EN)	
Ecologistes	NT 3104 - Freshwater ecology (EN)	
Ecofisiologia	NT 3682 - Human ecology (EN)	

Ecosistemes	NT 5774 - Phenology (EN)	
Ecotoxicologia	NT 4609 - Marine ecology (EN)	
Fenologia	RT 915 - Biogeography (EN)	
Fotoperiodicitat	RT 2464 - ecological succession (EN)	
Espècies	RT 36703 - colonizing ability (EN)	
	RT 36389 - Ecological factors (EN)	
	RT 35788 - Ecotoxicology (EN)	
	RT 2466 - Ecologists (EN)	
	RT 2482 - ecosystems (EN)	
	RT 34821 - Environmental degradation (EN)	
	RT 2593 - environment (EN)	
	RT 33560 - Sustainability (EN)	
	RT 2594 - environmental factors (EN)	
	RT 34929 - balance of nature (EN)	
Ecosistemes	Ecosystems	Ecosystem
RT Balanç ecològic	UF 66960 - this term does not exist in EN	Definition: A community of organisms and their physical environment interacting as an ecological unit. (Source: LBC) broader terms
Comunitats aquàtiques	BT 330985 - systems (EN)	broader terms
Distribució ecològica	NT 36669 - Agroecosystems (EN)	synecology
Ecologia	NT 912 - Biocoenosis (EN)	related terms
Gestió d'ecosistemes	NT 4931 - Moors (EN)	ecosystem type
Medi ambient aquàtic	NT 3456 - Habitats (EN)	
Nivells tròfics	NT 3014 - Food chains (EN)	Scope note:
Estructura tròfica	NT 932 - Biotopes (EN)	scope note is not available
Producció	RT 2056 - cycling (EN)	Groups:

biològica	<p>RT 36230 - Spatial distribution (EN)</p> <p>RT 34328 - primary productivity (EN)</p> <p>RT 33949 - Biodiversity (EN)</p> <p>RT 33481 - ecotones (EN)</p> <p>RT 2467 - ecology (EN)</p> <p>RT 2593 - environment (EN)</p> <p>RT 8176 - vegetation (EN)</p>	<p>BIOSPHERE (organisms, ecosystems)</p> <p>Themes:</p> <p>biology</p> <p>natural areas, landscape, ecosystems</p>
Geologia	Geology	Geology
BT Ciències de la terra	BT 2438 - Earth sciences (EN)	Definition: The study or science of the earth, its history, and its life as recorded in the rocks; includes the study of geologic features of an area, such as the geometry of rock formations, weathering and erosion, and sedimentation. (Source: MGH)
NT Estratigrafia	NT 25341 - Sediment (EN)	broader terms
Geomorfologia	NT 11669 - Hydrogeology (EN)	earth science
Geologia glaciària	NT 7440 - Stratigraphy (EN)	narrower terms
Hidrologia	NT 5748 - Petrology (EN)	geotechnics
Litologia	NT 4856 - Mineralogy (EN)	hydrogeology
Geologia estructural	NT 4612 - Marine geology (EN)	marine geology
tectònica	NT 3233 - Geomorphology (EN)	mining geology
RT Geofísica	NT 31904 - tectonics (EN)	tectonics
Geoquímica	RT 6926 - sedimentation (EN)	volcanology
Història geològica	RT 7188 - Soil sciences (EN)	
Institucions geològiques	RT 3727 - Hydrography (EN)	Scope note:
Geòlegs	RT 5509 - Palaeontology (EN)	scope note is not available
Mineralogia	RT 14650 - Subsidence (EN)	Groups:

Paleontologia	RT 24848 - Geochemistry (EN)	RESEARCH, SCIENCES
Palinologia	RT 3231 - Geologists (EN)	Themes:
	RT 3235 - Geophysics (EN)	natural dynamics
		research
		soil
Gestió	Management	Management
UF Administració	BT 330834 - activities (EN)	Definition: Government, control, superintendence, physical or manual handling or guidance; act of managing by direction or regulation, or administration, as management of family, or of household, etc. (Source: BLACK)
NT Gestió d'ecosistemes	NT 35287 - quality assurance (EN)	
NT Gestió del medi ambient	NT 26994 - project management (EN)	
NT Gestió financera	NT 25301 - Watershed management (EN)	Narrower terms
NT Gestió de la producció	NT 24026 - Fishery management (EN)	agricultural management
NT Gestió de recursos	NT 9000179 - data management (EN)	chain management
RT Màrqueting	NT 16155 - Management by objectives (EN)	decision process
RT Personal	NT 16129 - Forest management (EN)	energy management
RT Pert; Tècniques de revisió i avaluació de programes	NT 16094 - Crop management (EN)	fisheries management
RT Planificació	NT 16093 - Livestock management (EN)	human settlement management
	NT 37875 - environmental management (EN)	integrated management
	NT 37890 - Knowledge management (EN)	management technique

NT 9000088 - integrated management (EN)	policy instrument
NT 331459 - community management (EN)	touristic activity management
NT 331457 - co-management (EN)	
NT 331401 - logistics (EN)	Related terms
NT 330697 - Nutrient management (EN)	environmental management
NT 50192 - coordination (EN)	Scope note:
NT 49838 - information management (EN)	scope note is not available
NT 37934 - risk management (EN)	Groups:
NT 13262 - pest management (EN)	ADMINISTRATION, MANAGEMENT, POLICY, POLITICS, INSTITUTIONS, PLANNING
NT 3364 - Grassland management (EN)	Themes:
NT 2799 - Farm management (EN)	administration
NT 330751 - disease management (EN)	Other relations: Has exact match in AGROVOC: management
NT 2147 - Decision making (EN)	
NT 2143 - decentralization (EN)	
NT 6460 - Rationalization (EN)	
NT 1161 - Business management (EN)	
NT 6521 - Resource allocation (EN)	
NT 6524 - resource management (EN)	
NT 8320 - water management (EN)	
RT 1134 - budgets (EN)	
RT 2491 - Efficiency (EN)	
RT 358 - Depreciation (EN)	

RT 3041 - Forecasting (EN)
 RT 37968 - organizations (EN)
 RT 8440 - work organization (EN)
 RT 4236 - Leadership (EN)
 RT 37887 - entrepreneurship (EN)
 RT 7581 - systems analysis (EN)
 RT 6612 - risk (EN)
 RT 24392 - case studies (EN)
 RT 6205 - production location (EN)
 RT 28011 - Management techniques (EN)
 RT 28714 - Labour relations (EN)
 RT 28908 - Management games (EN)
 RT 28909 - management sciences (EN)

RT 5951 - planning (EN)
 RT 5364 - operations research (EN)
 RT 130 - administration (EN)

Medi ambient	Environment	Environment
(Utilitzar termes més específics)	UF 32706 - this term does not exist in EN	NOTES: No existeix el terme sol!
NT Medi aquàtic	UF 2465 - Ecological zonation (EN)	environmental science
NT Medi ambient tropical	UF 25640 - this term does not exist in EN	Definition: The interdisciplinary study of environmental problems, within the framework of established physical and biological principles, i.e. oriented toward a scientific approach. (Source: UNUN)
NT Mitjans sedimentaris	UF 15704 - Ecological zones (EN)	broader terms
NT Palaeomedis	BT 330892 - entities (EN)	science
RT Condicions mediambientals	NT 26824 - socioeconomic environment (EN)	

RT Efectes sobre el medi ambient	NT 554 - Aquatic environment (EN)	Scope note:
RT mediambientals Factors	NT 35294 - Riparian zones (EN)	scope note is not available
RT mediambientals Gràfics	NT 331336 - Natural environment (EN)	Groups:
	NT 24950 - sociocultural environment (EN)	RESEARCH, SCIENCES
	NT 24949 - economic environment (EN)	Themes:
	RT 23990 - Anaerobiosis (EN)	research
	RT 9000054 - Environmental information technology (EN)	
	RT 9000053 - Environmental extension (EN)	
	RT 2467 - ecology (EN)	
	RT 2482 - ecosystems (EN)	
	RT 92350 - environmental economics (EN)	
	RT 37875 - environmental management (EN)	
	RT 36263 - Environmental legislation (EN)	
	RT 2594 - environmental factors (EN)	
	RT 34821 - Environmental degradation (EN)	
	RT 33483 - Environmental impact assessment (EN)	
	RT 4185 - Landscape (EN)	
	RT 4802 - microclimate (EN)	
	RT 23988 - Aerobiosis (EN)	
	RT 9000056 - Environmental modelling (EN)	
Oceanografia	Oceanography	Oceanography

UF Oceanology	UF 5300 - Oceanology (EN)	Definition: The scientific study and exploration of the oceans and seas in all their aspects. (Source: MGH)
BT Ciències de la terra	BT 2438 - Earth sciences (EN)	broader terms
Ciències marines	BT 4613 - Marine sciences (EN)	water science
NT Oceanografia química	NT 1516 - Chemical oceanography (EN)	narrower terms
Oceanografia litoral	NT 2938 - Fishery oceanography (EN)	chemical oceanography
Oceanografia dinàmica	NT 5830 - Physical oceanography (EN)	estuarine oceanography
Oceanografia pesquera	RT 4612 - Marine geology (EN)	physical oceanography
Oceanografia militar	RT 4610 - Marine environment (EN)	
Paleoceanografia	RT 4609 - Marine ecology (EN)	Scope note:
Oceanografia física	RT 6903 - Sea water (EN)	scope note is not available
Oceanografia polar		Groups:
Oceanografia per ràdio		RESEARCH, SCIENCES
Oceanografia tropical		Themes:
RT Atlas oceanogràfics		research
Ecologia marina		water
Estudis oceanogràfics		
Geologia marina		
Medi ambient marí		
Meteorologia		
Recursos Naturals	Natural Resources	Natural Resource
UF Recursos naturals aquàtics	BT 9001017 - resources (EN)	Definition: A feature or component of the natural environment that is of value in serving human needs, e.g. soil, water, plantlife, wildlife,

		etc. Some natural resources have an economic value (e.g. timber) while others have a "noneconomic" value (e.g. scenic beauty). (Source: UNUN)
BT Recursos	NT 92341 - Biological resources (EN)	broader terms
NT Recursos alimentaris	NT 35106 - Feed resources (EN)	resource
Recursos de propietat comú	NT 28012 - Marine resources (EN)	narrower terms
Recursos energètics	NT 28006 - Energy resources (EN)	biological resource
Recursos hídrics	NT 28001 - Agricultural resources (EN)	climate resource
Recursos marins	NT 24056 - Nonrenewable resources (EN)	mineral resource
Recursos minerals	NT 7187 - Soil resources (EN)	sea resource
Recursos vivents	NT 3026 - Food resources (EN)	soil resource
Recursos renovables	NT 3050 - Forest resources (EN)	water resource
Recursos renovables no	NT 3218 - genetic resources (EN)	related terms
Recursos tradicionals no	NT 2942 - Fishery resources (EN)	natural material
RT Conservació de recursos	NT 6502 - Renewable resources (EN)	nature protection
Gestió de recursos	RT 36649 - Biological production (EN)	
Matèries primes	RT 2476 - Economic resources (EN)	Scope note:
Recursos escassos	RT 34268 - Indigenous organisms (EN)	scope note is not available
Ús múltiple de recursos	RT 9000115 - Natural resources management (EN)	Groups:
Recursos protegits	RT 33560 - Sustainability (EN)	RESOURCES (utilisation of resources)
	RT 28020 - Resource exploration (EN)	Themes:
	RT 28019 - Resource evaluation	resources

	(EN)	
	RT 6523 - Resource conservation (EN)	
	RT 24887 - raw materials (EN)	
	RT 6207 - production possibilities (EN)	
	RT 16141 - Nature reserves (EN)	
	RT 8388 - Wilderness areas (EN)	
	RT 6524 - resource management (EN)	
	RT 926 - biomass (EN)	
GIS (Sistemes d'informació Geogràfica)	GIS (Non-descriptor); USE: Geographical information systems	GIS digital system
UF Geographic information systems	USE FOR: Geographic information systems	Definition: An organized collection of computer hardware, software, geographic data, and personnel designed to efficiently capture, store, update, manipulate, analyze, and display all forms of geographically referenced information that can be drawn from different sources, both statistical and mapped. (Source: SFRC / GILP96) Broader terms
BT Sistemes d'informació (Information Systems)	BT 11769 - Information systems (EN)	geographic information system
	RT 36733 – Global positioning systems	Scope note:
	RT 92332 - geomatics (EN)	geographic information system
	RT 92363 - precision agriculture (EN)	Groups: INFORMATION, EDUCATION, CULTURE, ENVIRONMENTAL AWARENESS
		Themes: Geography,

Remote Sensing (Teledetection)	Remote Sensing	Remote Sensing
Geodetecció: <i>(Use for remote sensing of earth surface from space. Before 1986 search). For sensing from space use GEOSENSING</i>	UF 17590 - this term does not exist in EN	Definition: 1) The scientific detection, recognition, inventory and analysis of land and water area by the use of distant sensors or recording devices such as photography, thermal scanners, radar, etc. 2) Complex of techniques for the remote measure of electromagnetic energy emitted by objects. (Source: LANDY / ZINZAN)
UF Teledetecció terrestre	BT 7536 - surveying (EN)	
UF Teledetecció (Terra)	RT 156 - Aerial surveying (EN)	
UF Detecció per satèl·lit	RT 9000009 - AVHRR (EN)	
UF Tècniques de teledetecció		
BT Teledetecció (Remote Sensing)	RT 37359 - Image processing (EN)	broader terms
NT Teledetecció des de l'aire	RT 36760 - Imagery (EN)	telemetry
NT Teledetecció per satèl·lit	RT 36733 - Global positioning systems (EN)	
NT Geodetecció		
RT Radiació electromagnètica	RT 28279 - Sensors (EN)	related terms
RT Satèl·lits científics	RT 26062 - Remote sensors (EN)	digital image processing technique
RT Captura de dades	RT 24906 - telecommunications (EN)	observation satellite
RT Detectors d'infrarojos	RT 24134 - Telemetry (EN)	Scope note:
RT Ecosondeig	RT 24074 - Remote control (EN)	scope note is not available
RT Equip de teledetecció	RT 24071 - Radar (EN)	Groups:
RT Formació d'imatges	RT 16343 - Spot image (EN)	RESEARCH, SCIENCES

RT Oceanografia per ràdio	RT 14093 - Satellites (EN)	Themes: environmental policy geography information natural areas, landscape, ecosystems research space
RT Radiació electromagnètica	RT 5807 - Photogrammetry (EN)	
	RT 9000039 - Earth observation satellites (EN)	
SUB DISCIPLINES	SUB DISCIPLINES	SUB DISCIPLINES
Biodiversitat	Biodiversity	Biodiversity
UF Diversitat del Ecosistema	UF 33954 - Species diversity (EN)	Definition: 1) Genetic diversity: the variation between individuals and between populations within a species; species diversity: the different types of plants, animals and other life forms within a region; community or ecosystem diversity: the variety of habitats found within an area (grassland, marsh, and woodland for instance. 2) An umbrella term to describe collectively the variety and variability of nature. It encompasses three basic levels of organisation in living systems: the genetic, species, and ecosystem levels. Plant and animal species are the most commonly recognized units of biological diversity, thus public concern has been mainly devoted to conserving species diversity. (Source: WRES / GILP96)
UF Diversitat de hàbitats	UF 33951 - Ecosystem diversity (EN)	broader terms

RT Diversitat genètica	UF 33950 - Biological diversity (EN)	ecological parameter
RT Diversitat de espècies	BT 92341 - Biological resources (EN)	related terms
	NT 33952 - Genetic diversity as resource (EN)	genetic diversity
	NT 37977 - Agrobiodiversity (EN)	
	NT 37967 - Natural heritage (EN)	Scope note:
	RT 7280 - Species (EN)	scope note is not available
	RT 50344 - bioprospecting (EN)	Groups:
	RT 35030 - Biosafety (EN)	BIOSPHERE (organisms, ecosystems)
	RT 33955 - Biological diversity conservation (EN)	Themes:
	RT 3212 - Gene pools (EN)	biology
	RT 3249 - Germplasm (EN)	
	RT 2482 - ecosystems (EN)	
Climatic Changes (Canvi Climàtic)	Climatic Change	Climatic Change
UF Canvis climàtics	UF 19960 - this term does not exist in EN	Definition: The long-term fluctuations in temperature, precipitation, wind, and all other aspects of the Earth's climate. External processes, such as solar-irradiance variations, variations of the Earth's orbital parameters (eccentricity, precession, and inclination), lithosphere motions, and volcanic activity, are factors in climatic variation. Internal variations of the climate system, e.g., changes in the abundance of greenhouse gases, also may produce fluctuations of sufficient magnitude and variability to explain observed climate change through the feedback processes interrelating the components of the climate system. (Source: GSFC)

RT Activitat solar terrestre	UF 19961 - this term does not exist in EN	broader terms
RT Canvis a llarg temps		climate
RT Canvis en el nivell del mar	UF 73163 - this term does not exist in EN	Narrower terms
RT Canvis eustàtics		climate change adaptation
RT Clima	UF 35367 - Climate change (EN)	related terms
RT Climatologia	BT 330740 - natural phenomena (EN)	climatic alteration
RT Constant solar		deforestation
RT Desglaciació	RT 35921 - carbon dioxide enrichment (EN)	Man-made climate change
RT Efecte Hivernacle	RT 35249 - Ozone depletion (EN)	Groups:
RT Extincions massives	RT 34841 - Greenhouse gases (EN)	ATMOSPHERE (air, climate)
RT Glaciació	NT 34836 - Global warming (EN)	Themes:
RT Paleoclima	RT 1665 - climate (EN)	climate, natural dynamics
RT Paleotemperatura		Scope note:
RT Pol·lució aèria		scope note is not available
RT Química atmosfèrica		
No existeix el terme.	Aquatic Ecosystem	Aquatic Ecosystem
		Definition: Any watery environment, from small to large, from pond to ocean, in which plants and animals interact with the chemical and physical features of the environment. (Source: GILP96)
<i>Un terme relacionat seria: "Aquatic Ecology" i "Aquatic environment"</i>	<i>Un terme relacionat seria: "Aquatic Ecology" i "Aquatic environment"</i>	broader terms
	<i>Un terme Genèric: "Aquatic sciences"</i>	ecosystem type
<i>Un terme Genèric: "Aquatic sciences"</i>		narrower terms
		coastal ecosystem
		estuarine ecosystem

		freshwater ecosystem
		marine ecosystem
		Scope note: scope note is not available
		Groups: BIOSPHERE (organisms, ecosystems)
		Themes: biology
		water
Limnologia	Limnology	Limnology
BT Ciències aquàtiques	SNR 3104 - Freshwater ecology (EN)	Definition: The study of bodies of fresh water with reference to their plant and animal life, physical properties, geographical features, etc. (Source: CED)
NT Limnologia química	BT 92335 - Aquatic sciences (EN)	broader terms
Limnologia de peixos	RT 3875 - Inland water environment (EN)	biology
Limnologia física		
Paleolimnologia		
RT Ciència de l'aigua dolça		Scope note: scope note is not available
Dades limnològiques		Groups: RESEARCH, SCIENCES
Equipament limnològic		Themes: research
Estanys		water
Hidrografia		
Hidrologia		
Institucions limnològiques		
Llacs		
Reserves d'aigua		

Hidrologia (Use for studies of continental surface water and hydrogeology).	Hydrology	Hydrology
BT Geologia	UF 92095 - this term does not exist in EN	Definition: The science that treats the occurrence, circulation, distribution, and properties of the waters of the earth, and their reaction with the environment. (Source: MGH)
RT Ciències de l'aigua dolça (freshwater)	BT 2438 - Earth sciences (EN)	broader terms
Geoquímica	NT 3727 - Hydrography (EN)	water science
Geomorfologia	NT 11669 - Hydrogeology (EN)	
Cicle hidrològic	RT 3718 - Hydrobiology (EN)	Scope note:
Hidrosfera	RT 3723 - Hydrodynamics (EN)	scope note is not available
Limnologia	RT 6161 - precipitation (EN)	Groups:
Aigua	RT 11670 - Hydrological cycle (EN)	RESEARCH, SCIENCES
<i>Economia de l'aigua (Water Budget)</i>		Themes:
		research
		water
Hidrobiologia	Hydrobiology	Hydrobiology
UF Biologia Aquàtica	UF 92053 - this term does not exist in EN	Definition: Study of organisms living in water. (Source: ZINZAN)
BT Biologia	BT 925 - biology (EN)	broader terms
RT Algologia	RT 262 - Algology (EN)	biology
Biologia de peixos	RT 2937 - Fishery biology (EN)	
Ciències marines	RT 3731 - Hydrology (EN)	Scope note:
Ictiologia	RT 4613 - Marine sciences (EN)	scope note is not available
Malacologia		Groups:
		RESEARCH, SCIENCES
		Themes:

		biology research water
Hidrodinàmica	Hydrodynamics	Hydrodynamics
BT Mecànica de fluids RT Capes limítrofes Circulació del aigua Cossos "acoplats" Equacions de Navier-Stokes Equacions hidrodinàmiques Fluxos de corrent Forces de les onades Hidrostatica Limnologia física Oceanografia física Vorticitat	BT 330861 - Fluid dynamics (EN) RT 3731 - Hydrology (EN) RT 33494 - Friction (EN) RT 35220 - Mechanics (EN)	No existeix
Hidrogeologia	Hydrogeology	Hydrogeology
[No Descriptor] <u>Definició: Rama de la hidrologia que tracta de les aigües subterrànies, tenint en compte les condicions geològiques. [1]</u>	[No Descriptor]	[No Descriptor]
Hidrogeomorfologia	Hydrogeomorphology	Hydrogeomorphology
[No descriptor] <u>Definició: Disciplina que estudia les relacions entre el cicle hídic i el relleu.[2]</u>	[No descriptor]	[No descriptor]
ALTRES DESCRIPTORS	ALTRES DESCRIPTORS INTERESSANTS	ALTRES DESCRIPTORS INTERESSANTS

INTERESSANTS		
Medi Aquàtic	Aquatic environment	Aquatic environment
BT Medi Ambient	UF 22932 - this term does not exist in EN	Definition: Waters, including wetlands, that serve as habitat for interrelated and interacting communities and populations of plants and animals. (Source: LEE)
Medi bentònic	BT 331336 - Natural environment (EN)	broader terms
Medi ambient d'aigües salobres	BT 2593 - environment (EN)	natural environment
Medi ambient epònic	NT 12208 - Littoral environment (EN)	
Medi ambient de las aigües continentals	NT 5658 - Pelagic environment (EN)	Scope note:
Medi intersticial	NT 4610 - Marine environment (EN)	scope note is not available
Medi ambient marí	NT 3875 - Inland water environment (EN)	Groups:
Medi ambient pelàgic	NT 31 - Abyssal environment (EN)	ENVIRONMENT (natural environment, anthropic environment)
RT Aigua	NT 1050 - Brackishwater environment (EN)	Themes:
Biòtops	NT 877 - Benthic environment (EN)	natural areas, landscape, ecosystems
Ciències de l'aigua	NT 840 - Bathyal zone (EN)	water
Comunitats aquàtiques	RT 553 - Aquatic communities (EN)	
Ecosistemes	RT 3104 - Freshwater ecology (EN)	
Estudis del medi ambient		
Gestió del medi ambient		
Hàbitat		
Masses d'aigua		

Water Masses	Body Water	Water body
NT Aigua subàlvia	UF 20243 - this term does not exist in EN	Definition: Any mass of water having definite hydrological, physical, chemical and biological characteristics and which can be employed for one or several purposes. (Source: FLGISA)
NT Aigües continentals	BT 979 - Body fluids (EN)	broader terms
Aigües costeres	RT 5316 - Oedema (EN)	hydrosphere
Aigües praderes	RT 8309 - water (EN)	narrower terms
Desaigües		aquifer
Llacunes		water bottom
Oceans		
Masses d'aigües profundes		Scope note:
Masses d'aigua intermèdia		scope note is not available
Masses d'aigua de superfície		Groups:
RT Aigua		HYDROSPHERE (freshwater, marine water, waters)
Aigües recreatives		Themes:
Canals		soil
Capes testimoni (aigua)		water
Circulació del aigua		
Classificació òptica		
Columna d'aigua		
Diagrames T/S		
Hidrografia		
Hidrosfera		
Intrusions de masses d'aigua		

Medi aquàtic		
Recursos hídrics		
Mescles d'aigua		
Propietats del aigua		
Zones humides	Wetlands	Wetlands ecosystem
BT Aigües continentals	BT 5834 - physiographic features (EN)	Definition: Ecosystems of areas largely inundated with water but offering elevated lands as a habitat for wildlife. This areas include swamps, both seasonal and permanent, marsh, open fresh water, shallow saline lagoons, the estuaries of rivers, floodplains and coastal sand dunes. They provide food, breeding grounds, water and sanctuary for many forms of fish, birds and other animal and plant life. They are among the most productive ecosystems producing timber, peat moss and crops such as rice and a variety of berries. (Source: GILP96 / WRIGHT)
NT Maresmes	NT 987 - Bogs (EN)	broader terms
Pantans	NT 2978 - Flooded land (EN)	terrestrial ecosystem
RT Aigua estancada	NT 4631 - Marshes (EN)	Scope note: scope note is not available Groups: BIOSPHERE (organisms, ecosystems) Themes: biology natural areas, landscape,
<i>Cheniers</i>	NT 7544 - Swamps (EN)	
Deltas	NT 331329 - constructed wetlands (EN)	
Inundació; Enfonsament	RT 2372 - drainage (EN)	
Recuperació de sols	RT 8333 - Waterlogging (EN) RT 16151 - Water tolerance (EN)	

DISCIPLINES O FACETES	DISCIPLINES O FACETES	DISCIPLINES O FACETES
Materials [no fa referència a cap material relacionat amb els ecosistemes aquàtics]	Materials	Materials (no és un concepte en si mateix)
NT Alloy	BT 330705 - substances (EN)	materials science
NT Biogenic material	NT 34300 - insulating materials (EN)	scope note: applied science
NT Buyoyancy materials	NT 32440 - Envelope materials (EN)	materials science
NT Ceramics	NT 29036 - Metals (EN)	scope note: science
NT Coating materials	NT 28376 - Laminates (EN)	materials technology
NT Construction materials	NT 26764 - ceramics (EN)	scope note: environmental technology
NT Fiber glass	NT 26181 - Sheet metal (EN)	
NT Gear materials	NT 25740 - protective materials (EN)	
NT Hazardous materials	NT 16008 - packaging materials (EN)	
NT Insulating materials	NT 8421 - Wood (EN)	
NT Isotropic materials	NT 5998 - Plastics (EN)	
NT Packing materials	NT 3393 - growing media (EN)	
NT Plastics	NT 3272 - glass (EN)	
NT Radioactive materials	NT 1141 - Construction materials (EN)	
NT Raw materials	NT 434 - Litter for animals (EN)	
NT Rubber materials		
NT Wood		
RT Components		
RT Material technology		
RT Materials testing		
Modelització	Modelling	Modelling

Modelatge	[No existeix]	Definition: An investigative technique using a mathematical or physical representation of a system or theory that accounts for all or some its known properties. Models are often used to test the effect of changes of system components on the overall performance of the system. (Source: LEE)
TR Programació matemàtica		broader terms
TR Models		simulation
TR Simulació		narrower terms
		model
		Scope note: not available
		Groups: RESEARCH, SCIENCES Themes: research
Monitorització	Monitoring	Monitoring (no correspon a un terme sol)!
NT Monitoreig mediambiental	UF 90019 - this term does not exist in EN	monitoring criterion
TR Canvis a llarg termini	UF 32876 - this term does not exist in EN	scope note: monitoring technique
TR Control	UF 29424 - this term does not exist in EN	monitoring data
TR Estudis de valor basal	UF 37006 - this term does not exist in EN	scope note: environmental data
TR Inspecció	UF 21389 - this term does not exist in EN	monitoring equipment
TR Sistemes de monitorització; Sistemes de vigilància	UF 61576 - this term does not exist in EN	scope note: equipment
	UF 89384 - this term does not exist in EN	monitoring network

BT 330834 - activities (EN)	scope note: monitoring technique
NT 37876 - Environmental monitoring (EN)	monitoring station
NT 28322 - Process control (EN)	scope note: industrial site
NT 37663 - Pest monitoring (EN)	monitoring system
NT 37212 - Biomonitoring (EN)	scope note: monitoring technique
NT 37838 - Crop monitoring (EN)	monitoring technique
RT 37871 - early warning systems (EN)	scope note: monitoring
RT 31862 - control systems (EN)	
RT 31861 - surveillance systems (EN)	
RT 28279 - Sensors (EN)	
RT 15855 - Automation (EN)	
RT 10289 - Data processing (EN)	
RT 7536 - surveying (EN)	
RT 6774 - Sampling (EN)	
RT 4788 - methods (EN)	
RT 2128 - data collection (EN)	
<p>^[1] Glossario Hidrológico Internacional. EUA : UNESCO; World Meteorological Organization (WMO). URL: http://www.cig.ensmp.fr/~hubert/glu/HINDEST.HTM</p> <p>^[2] VADEMECUM REMER (Glosario geológico). URL: España : Dirección General de Protección Civil y Emergencias. Ministerio del Interior. URL: http://www.proteccioncivil.org/vademecum/vdm02514.htm</p>	

Annex 8.1.1. Biocomplexity Tesaurus NBII Thesaurus

Biocomplexity Tesaurus – NBII (T2)
Aquaculture
BT Husbandry
NT Aquaculture practices
NT Fish culture
NT Freshwater aquaculture
NT Marine Aquaculture
RT Aquaculture development
RT Aquaculture effluents
RT Aquaculture engineering
RT Aquaculture equipment
RT Aquaculture facilities
RT Aquaculture products
RT Aquatic sciences
RT Breeding
RT Brood care
RT Fish ponds
RT Fishes
RT Ocean ranching
RT Rearing
RT Resource management
SC ASF
SC LSC
SC POL
UF Aquatic agriculture

UF Aquiculture
Biologia
NT Anatomia
NT Bioquímica
NT Bioenergètica
NT Botànica
NT Criobiologia
NT Cronobiologia
NT Citologia
NT Embriologia
NT Etnobiologia
NT Biologia pesquera
NT Biologia funcional
NT Biologia del sol
NT Genètica
NT Hematologia
NT Histologia
NT Hidrobiologia
NT Microbiologia
NT Morfologia (Organismes)
NT Biologia molecular
NT Paleontologia
NT Parasitologia
NT Zoologia
RT Activitat Biològica
RT Anàlisis biològics
RT Bioacústica

RT Biofísica
RT Biogeografia
RT Biogeoquímica
RT Bioinformàtica
RT Cicle Vital
RT ciències aquàtiques
RT Ecologia
RT Fenomen biològic
RT Limnologia
Etc.
SC ASF
SC LSC
SN És recomana utilitzar el terme més específic
UF Ciències Biològiques
UF Ciències de la vida
ciències aquàtiques
NT Ciències de l'aigua dolça
NT Ciències marines
NT Limnologia
RT Aqüicultura
RT Biologia
RT Hidrosfera
RT Medi ambient aquàtic
SC ASF
Ecology
NT Autoecology

NT Chorology
NT Ecophysiology
NT Estuarine ecology
NT Ethology
NT Genecology
NT Freshwater ecology
NT Habitat utilization
NT Historical ecology
NT Human ecology
NT Marine ecology
NT Ordination
NT Paleoecology
NT Phytosociology
NT Planktonology
NT Population ecology
NT Soil ecology
NT Synecology
NT Systems ecology
RT Biocomplexity
RT Biogeography
RT Biology
RT Ecological adaptations
RT Ecological distribution
RT Ecological evaluation
RT Ecological genetics
RT Ecologists
RT Ecosystem analysis

RT Ecosystem stability

RT Ecosystems

RT Ecotourism

RT Ecotoxicology

RT Environmental conditions

RT Environmental education

RT Limnology

RT Parasitology

RT Phenology

RT Photoperiodicity

RT Species

RT Water relations

SC ASF

SC LSC

SC POL

UF Bionomics

UF Ecological sciences

Ecosystems

NT Aquatic Ecosystems

NT Assembles (communities)

NT Biocoenoses

NT Climax communities

NT Coastal Ecosystems

NT Faunal provinces

NT Microbial mats

NT Microecosystems

NT Plant communities

NT Roadsides

NT Terrestrial ecosystems

RT Bioenergetics

RT Biological production

RT Biosphere

RT Community composition

RT Ecological adaptations

RT Ecological association

RT Ecological distribution

RT Ecology

RT Ecosystem analysis

RT Ecosystem disturbance

RT Ecosystems dynamics

RT Ecosystems management

RT Ecosystem models

RT Ecosystem recovery

RT Ecosystem resilience

RT Ecosystem stability

RT Ecosystem structure

RT Environments

RT Food webs

RT Habitats

RT Microcosms

RT Niches

RT Populations

RT Prebiosis

RT Synecology

RT Systems ecology

RT Trophic levels

RT Water relations

SC ASF

SC LSC

SC POL

UF Ecological communities

UF Ecozones

Geology

BT Herath sciences

NT Geochemistry

NT Geomorphology

NT Geophysics

NT Glaciology

Hydrology

Lithology

Marine geology

Mineralogy

Petroleum geology

Petrology

Sedimentology

Soil science

Speleology

Stratigraphy

Structural geology

tectonics

RT Basins (geology)

Biochemistry

Deltas

Evolution (biology)

Geography

Geologists

Geotechnology

Lithification

Natural resources

Origin

Paleoclimatology

Paleoenvironemnts

Paleolimnology

Paleontology

Palynology

Soil analysis

Soil surveys

SC ASF

SC LSC

SC POL

UF Earth history

Management

NT Emergency preparedness

NT Environmental management

NT Financial management

Fishery management

Production management

Resource management

Risk assessment

Total quality management

Waste management

Wildlife management

RT Bureaucracy

RT Conservation

RT Legislation

Management tools

Marketing

Personnel

PERT

Planning

Property rights

Silviculture

SC ASF

SC LSC

SC POL

SN Utilitzar el terme més específic

UF Administration

Environment

NT Anaerobic environments

NT Aquatic environments

NT Coastal environments

NT Dystrophic environments

NT Geothermal areas

NT Habitats

NT Harsh environments

NT Mesocosms

NT Microcosms

NT Microenvironments

NT Paleoenvironments

NT Polar environments

NT Polluted environments

NT Refugia

NT Sedimentary environments

NT Slope environments

NT Social environments

NT Subpolar environments

NT Terrestrial environments

NT Tropical environments

NT Wilderness

RT Biosphere

RT Climatic zones

RT Ecosystems

RT Environmental chambers

RT Environmental changes

RT Environmental conditions

RT Environmental effects

RT Environmental factors

RT Environmental hygiene

RT Environmental impact

RT Environmental management

RT Environmental monitoring

RT Environmental protection

RT Environmental quality

RT Environmental regulations

RT Environmental requirements

RT Environmental restoration

RT Environmental sociology

RT Environmental stress

RT Environmental studies

RT Environmental surveys

RT Environmental utilization

RT Geography

RT Natural areas protection

SC ASF; LSC

SN Use of a more specific term is recommended

Oceanography

BT Marine sciences

NT Chemical oceanography

NT Coastal oceanography

NT Paleoceanography

NT Physical oceanography

NT Polar oceanography

NT Tropical oceanography

RT Coastal environments

RT Hydrology

RT Limnology

RT Marine ecology

RT Marine geology
RT Marine scientists
RT Meteorology
RT Natural resources
RT Ocean floor
RT Oceans
RT Sea level
SC ASF
SC LSC
SC POL
UF Oceanology
Recursos Naturals
NT Agricultural resources
Coastal resources
Common property resources
Desert resources
Fishery resources
Food resources
Forest resources
Marine resources
Mineral resources
Nonrenewable resources
Plant resources
Rare resources
Renewable energy resources
Water resources
RT Biodiversity

Conservation
Energy resources
Geology
Multiple use resources
Oceanography
Protected areas
Raw materials
Resource conservation
Resource management
Rights of future generations
Rocks
Soils
Rocks
Watersheds
SC ASF
SC POL
UF Biological resource
GIS – USE: Geographic information systems
BT Information systems
RT Geography
RT Remote sensing
SC ASF
SC LSC
SC POL
UF GIS
Remote sensing

NT Airborne sensing
Satellite sensinG
RT Aerial photography
Biotelemetry
Data acquisition
Echosounding
Electromagnetic radiation
→
Geographic information systems
Imaging
Landsat
Lidat
Mapping
Monitoring
Monitoring systems
Photography
Photointerpretation
Radar
Remote sensing equipment
Satellites
Surveys
Telemetry
Tracking
SC ASF
SC LSC
UF Image sensors
UF Remote sensors

SUB DISCIPLINES

Biodiversitat

NT Genetic diversity

NT Species diversity

RT Biocomplexity

RT Biological control

RT Biomass

RT Natural Resources

SC ASF

SC CERES

SC LSC

SC POL

UF Biological diversity

UF Ecosystem diversity

UF Habitat diversity

Climatic change

BT Changes

NT Glaciation

NT Global warming

NT Greenhouse effect

RT Air pollution

RT Atmospheric chemistry

RT Climate

RT Climatology

RT Deglaciation

RT Erth rotation

RT Greenhouse gases

...

RT Meteorology

RT Paleoclimate

RT Sea level changes

RT Seasonal variations

SC ASF

SC LSC

SC POL

Ecosistemes Aquàtics

BT Biosphere

BT Ecosystems

NT Benthos

NT Epipsammom

NT Estuarine Ecosystems

NT Freshwater ecosystems

NT Marine ecosystems

NT Periphyton

NT Plankton

NT Pleuston

NT Psammon

NT Season

RT Aquatic environments

RT Aquatic organisms

RT Biota

RT Estuarine ecology

RT Water
SC CERES
SC LSC
Limnology
BT Aquatic sciences
NT Paleolimnology
RT Biology
RT Ecology
RT Ecosystem analysis
RT Freshwater ecology
RT Freshwater ecosystems
RT Freshwater environments
RT Freshwater sciences
RT Freshwater scientists
RT Hydrography
RT Hydrology
RT Inland waters
RT Lakes
RT Limnology equipment
RT Oceanography
RT Ponds
RT Reservoirs
RT Streams
SC ASF
SC LSC
SC POL

Hydrology
BT Geology
NT Soil hydrology
RT Flood control
RT Floods
RT Freshwater sciences
RT Geochemistry
RT Geomorphology
RT Hydrography
RT Hydrologic cycle
RT Hydrosphere
RT Limnology
RT Oceanography
RT Surface water
RT Water
RT Water Resources
RT Water supplies
SC ASF
SC LSC
SC POL
SN Use for studies of continental surface water and hydrogeology
Hidrobiología
BT Biology
RT Freshwater sciences
RT Marine sciences
RT Phycology

SC ASF
UF Aquatic biology
Hidrodinàmica
BT Dynamics
BT Fluid dynamics
RT Aerodynamics
Boundary layers
Hydraulics
Hydrostatics
Physical oceanography
Stream flow
Water circulation
SC ASF
SC LSC
SC POL
Hidrogeologia [No Descriptor]
<i>Nota: En el T2, dintre del descriptor 'Hidrologia', ens diuen que és pot utilitzar per als estudis de "Hidrogeologia".</i>
Hidrogeomorfologia [No descriptor]
ALTRES DESCRIPTORS INTERESSANTS
Aquatic environment
BT Environments
NT Aphotic zone
NT Aquatic saline environments

NT Bentic environments
NT Compensation depth
NT Epontic environments
NT Eutrophic environments
NT Inland water environments
NT Interstitial environments
NT Mesotrophic environments
NT Oligotrophic environments
RT Aquatic ecosystems
RT Aquatic sciences
RT Boats
RT Dams
RT Deep water
RT Environmental management
RT Flood plains
RT Hydrosphere
RT Water
RT Water bodies
RT Water column
SC ASF
SC LSC
SC POL
SN Environments of all types of hydrosphere
Water bodies
NT Anoxic waters
NT Coastal water

NT Inland water
NT Lagoons
NT Oceans
NT Pools
NT Puddles
NT Water courses
NT Water springs
RT Aquatic environments
RT Hydrosphere
NT Recreational waters
NT Water
NT Water budget
NT Water column
NT Water resources
SC ASF
SN Surface waters of the Earth. Use of a narrower term is recommended
SN Surface water bodies
Wetlands
BT terrestrial ecosystems
NT Artificial wetlands
NT Bogd
NT Fens
NT Forested wetlands
NT Marshes
NT Mud flads
NT Peatlands

NT Swamps
RT Deltas
RT Flood irrigation
RT Flooded areas
RT Inland water environments
RT Rice fields
RT Stagnant water
RT Waterlogging
RT Wetland agriculture
SC ASF
SC CERES
SC LSC
SC POL
DISCIPLINES O FACETES
Materials [Només és citaran els materials relacionats amb els ecosistemes i medi aquàtic]
NT Biodegradable substance
Biogenic materials
Biomass
Biomaterials
Chemicals
Fertilizers
Fluids
Fuels
Minerals

Nutrients

Matter organic

Plastics

Pollutants

Radioactive materials

Repellents

Residues

Sediments

Soils

Wastes

Etc.

RT Materials testing

SC ASF

SC LSC

SN Use of a more specific term is recommended

Modelling

[No existeix]

Monitoring

NT Biomonitoring

NT Environmental monitoring

Surveillance (monitoring)

RT Baseline studies

Follow-up studies

Inspection

Laboratory methods

Long term changes

Measuring techniques

Monitoring systems

Observation methods

Quality control

Remote sensing [= *Teledetecció*]

Sampling

Surveys

Telemetry

SC ASF

SC LSC

UF Monitoring methods

Annex 8.2. Resultats mapeig semi automàtic: anàlisi qualitativa dels alineaments produïts i realitzades per avaluadors humans

8.2.1. Resultats de l'anàlisi qualitativa de l'avaluador 1

Stratum	Exact Match	Close Match	Broader Match	Narrower Match	Related Match	Unrelated Match	Unsure
1	7	4	5	4	8	12	0
2	12	3	1	1	1	2	0
3	15	3	0	0	0	1	1
4	17	2	0	0	1	0	0
5	18	1	0	0	0	1	0
6	17	2	0	0	1	0	0
7	20	0	0	0	0	0	0
8	19	0	0	1	0	0	0
9	19	1	0	0	0	0	0
10	3	3	1	0	1	0	0

Nota: els resultats de l'avaluador 1, corresponen al doctorand d'aquesta tesi

8.2.2. Resultats de l'anàlisi qualitativa de l'avaluador 2

Stratum	Exact Match	Close Match	Broader Match	Narrower Match	Related Match	Unrelated Match	Unsure Match
1	6	4	8	4	9	9	0
2	8	2	2	3	4	0	1
3	17	1	0	1	0	1	0
4	17	2	0	0	1	0	0

5	18	1	1	0	0	0	0
9	19	1	0	0	0	0	0

Nota: els resultats de l'avaluador 2, corresponen a un col·laborador de la Unicersitat d'Àmsterdam.

Annex 8.3. Anàlisi qualitativa de l'experiment d'alineament de tesaurus multilingües (Avaluador 1, 10 estrats)

Els resultats poden ser visualitzats en diferents modes, mode de gràfic, mode d'avaluació, mode *turtle* o *tripletes* i mode local (exemplificació de l'ontologia en format rdf/xml). Els experiments es realitzen durant el 2011 amb versions del tesaurus del 2010, però les anàlisis dels experiments són duts a terme durant el 2011, per tant, ja es comencen a veure algunes diferències d'actualització dels tesaurus. En alguns casos d'avaluacions, per tal de justificar els processos d'equivalència dels conceptes, es mostra la versió actualitzada dels conceptes i la seva jerarquia extrets de la versió online dels tesaurus. Però en general els annexes representen en mode de la interfície d'avaluació d'Amalgame.

Aquests annexes representen només les avaluacions del avaluador 1 (Carmen Reverté, Universitat de Barcelona), on apareixen observacions i informació qualitativa que han servit per elaborar les avaluacions finals. A continuació es mostren els 10 estrats agafats com a mostra d'estudi de mapeig multilingüe:

Stratum 1: Mostra aleatòria de 40 alineacions

1.1 Alineament 1: Camping (GEMET) versus Recreation (AGROVOC)

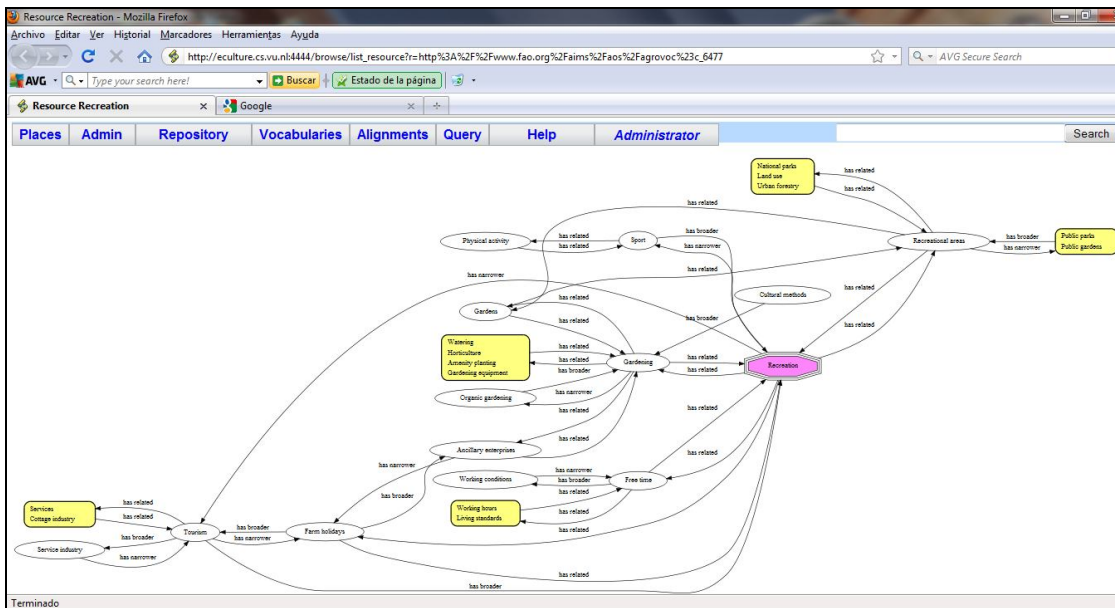
- **Resultat d'avaluació:** Camping és considerat en l'avaluació com un terme específic o NT de "Recreation", ja que GEMET representa el terme "Recreation" com un concepte superior al de Camping.

Camping (GEMET). Url: http://www.eionet.europa.eu/gemet/concept/1121	Recreation (AGROVOC). Url: http://aims.fao.org/en/agrovoc/c-term-info?mytermcode=6477&mylang_interface=en&myLanguage=EN
Definition: Guarded area equipped with sanitary facilities where holiday-makers may pitch a tent and camp by	UF: 29031 - leisure (recreation) (EN) UF: 29314 - this term does

<p>paying a daily rate. (Source: ZINZAN)</p> <p>Broader Alineament: leisure activity</p> <p>Scope note: not available</p> <p>Groups: RECREATION, TOURISM</p> <p>Themes: social aspects, population tourism</p> <p>Other relations:</p> <p>Has exact match</p> <hr/> <p>AGROVOC: Camping</p> <hr/> <p>EuroVoc: camping</p> <hr/> <p>Wikipedia article</p> <hr/> <p>Camping</p> <hr/> <p>Has close match</p> <hr/> <p>UMTHES: Camping</p>	<p>not exist in EN</p> <p>BT: 330834 - activities (EN)</p> <p>NT: 7328 - Sport (EN)</p> <p>NT: 7822 - tourism (EN)</p> <p>NT: 25225 - Camping (EN)</p> <p>RT: 29032 - Free time (EN)</p> <p>RT: 34092 - Recreational fishing (EN)</p> <p>RT: 16198 - recreational areas (EN)</p> <p>RT: 3195 - Gardening (EN)</p>
--	--

*Visualització dels propis tesaurus online (darrera consulta 2013):

- **Representació Gràfica amb Amalgame del concepte “Recreation” en AGROVOC.** La versió d’AGROVOC emprada en Amalgame, 2011, no apareixia el terme “Camping” com a un terme específic de “Recreation”, però en canvi, en GEMET si apareixia el concepte RECREATION com a terme superior (Group).



1.2 Alineament 2: Draina versus Watersheds

The screenshot shows the 'Amalgame alignment evaluator' interface. The search term is 'drainage system'. The interface displays two main sections:

- Left Panel (Source: GEMET):** A hierarchical tree structure showing 'drainage system' under 'HYDROSPHERE (freshwater, marine water, waters)'. The tree includes:
 - hydrologic cycle
 - hydrologic balance
 - runoff
 - drainage
 - drainage system** (highlighted)
 - natural drainage system
- Right Panel (Target: AGROVOC):** A search result for 'Watersheds, Catchment areas'. It shows 'target concept 1 (was: closeMatch?) out of 1 mappings: Exact match (=)'. Below this, there are options for 'Broader (b)', 'narrower (n)', 'related match (r)', and '? I'm not sure (u)?'. A 'comment (optional):' field is also present.

 The interface includes a search bar, navigation tabs, and a status bar at the bottom with the version 'ClioPatris (version epoch-682-g4a458ba)'.

- **Avaluació:** es poden considerar conceptes relacionats, perquè les conques hidrogràfiques poden ser un sistema de drenatge natural. Però en aquests dos tesaurus representen conceptes totalment diferents. GEMET parla de sistemes de drenatges com a mètode i AGROVOC, contempla el concepte com a “conques” o una representació hidrogràfica d’un riu.
 - i. Cas d’exemple, on la institució o el tipus de col·lecció continguda en una institució determinaran el tipus de relació a establir (*unrelated* o *related*)

1.3 Alineament 3: Rural Law versus Agricultural and rural legislation

<p>Rural Law. Url: http://www.eionet.europa.eu/gemet/concept?cp=11476&langcode=en&ns=1</p>	<p>Agricultural and rural legislation. Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=28633&mylanginterface=en&myLanguage=EN</p>
<p>Definició: A binding rule or body of rules prescribed by a government pertaining to matters of importance to residents of sparsely populated regions, especially agricultural and other economic issues. (Source: RUR)</p> <p>Broader terms: agricultural legislation</p> <p>Scope note: is not available</p> <p>Groups: LEGISLATION, NORMS, CONVENTIONS</p> <p>Themes: agricultura legislation</p>	<p>UF 28822 - Agricultural and rural law (EN)</p> <p>UF 28824 - Agricultural law (EN)</p> <p>UF 36950 - Agrarian law (EN)</p> <p>BT 4253 - legislation (EN)</p>

- **Avaluació:** tal i com queda representat en l'estructura de GMET, "Agricultural legislation" és un BT (concepte superior o genèric) de Rural Law.

1.4Alineament 4: justice versus Justicia

The screenshot shows the Amalgame alignment evaluator interface. On the left, the source concept is 'justice' from the URL <http://www.eionet.eu.int/gemet/concept/11583>. Its definition is 'The correct application of law as opposed to arbitrariness. (Source: CURZON)'. The scope is 'legislation'. The hierarchy includes 'SOCIAL ASPECTS, ENVIRONMENTAL POLICY MEASURES' and 'LEGISLATION, NORMS, CONVENTIONS', with 'justice' as a sub-concept. Other sub-concepts listed are 'access to the courts', 'crime', 'criminality', 'legal procedure', and 'ruling'.

On the right, the target concept is 'Justicia' from the URL http://www.fao.org/aims/aos/agrovoc/Rc_27220. The interface shows a 'target concept 1 (was: closeMatch?) out of 1 mappings' with options for 'Exact match (e)', 'close match (c)', and 'not related (-)'. Below this, there are options for 'Broader (b)', 'narrower (n)', 'related match (r)', and '? I'm not sure (u)?'. A 'comment (optional):' field is also present. The target hierarchy includes 'Organisms', 'Plants', and 'Acanthaceae', with 'Justicia' as a sub-concept.

At the bottom of the interface, there is a 'reset all()' button and a version string 'ClioPatria (version epoch-682-g4a458ba)'.

- **Avaluació:** el primer concepte fa referència a un terme jurídic (rural law), i el segon concepte fa referència al nom científic d'una planta. Exemple clar de que no existeix relació entre els conceptes.

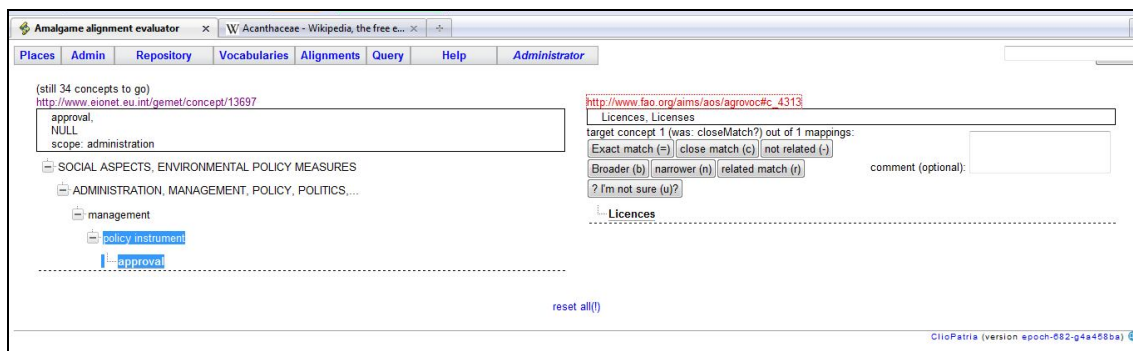
1.5 Alineament 5: Prescription versus Common Law

- **Avaluació:** en aquest context són termes no relacionats, però en altres contexts podrien ser termes relacionats (RT). És a dir, una llei comú o “common law” pot incloure “personal law or prescriptions”. Ja que una prescripció en termes de dret comú o “common law” significa dret civil, i per tant, són termes relacionats quant a la matèria, no pas amb el significat.

1.6 Alineament 6: underground disposal versus Runoff

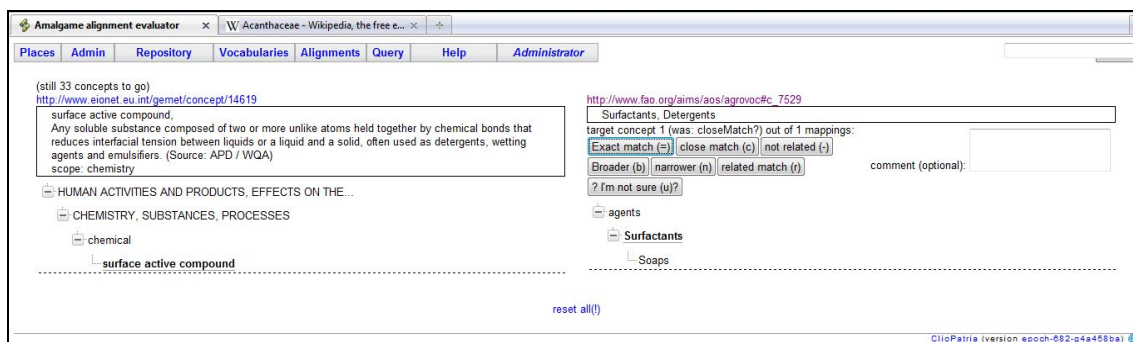
- **Comentari:** termes no relacionats. El terme equivalent a “Underground disposal” de GEMET és “Waste disposal” en AGROVOC, però no apareix en aquest mapeig.

1.7 Alineament 7: Approval versus licences



- **Avaluació:** una llicència és un producte d'una aprovació (instrument polític). Així doncs, "licences" es pot considerar com un terme específic o NT d'"Approval". Seria un exemple similar a dir que un formatge és un "producte lacti". Aquest exemple podria servir per millorar la cobertura i exhaustivitat d'un nou vocabulari controlat.

1.8 Alineament 8: Surface active compound *versus* surfactants



- **Avaluació:** és pot considerar una equivalència exacta o "Exact Match". Si s'entra a la Url del concepte "Surfactants" d'AGROVOC, es visualitza que el terme representat per GEMET ("surface active compound") és un terme no acceptat de "Surfactants" en el thesaurus AGROVOC. En un tercer/nou thesaurus s'hauria de decidir quin d'ambdós termes es considerat el terme preferit (prelabel) i quin el no acceptat o alternatiu (altlabel).

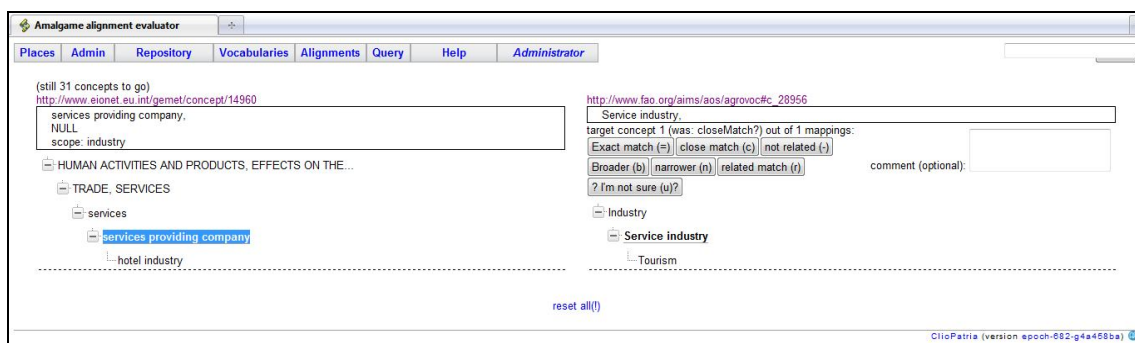
1.9 Alineament 9: longitude *versus* length:

Longitude (GEMET). Url:	Length (AGROVOC). Url:
http://www.eionet.europa.eu/gemet/concept?cp=14937&langcode=en&ns=1	http://aims.fao.org/en/agrovoc-term-info?mytermcode=4260&mylanginterface=en&myLanguage=EN
Definition: Distance in degrees east or west of the prime meridian at 0° measured by the angle between the plane of the prime	UF: 1327908668649 - this term does not exist in EN UF: 22624 (largo - ES) - this

<p>meridian and that of the meridian through the point in question, or by the corresponding time difference. (Source: CED)</p> <p>Broader terms: co-ordinate system</p> <p>Scope note: not available</p> <p>Groups: RESEARCH, SCIENCES</p> <p>Themes: research</p> <p>Other relations:</p> <p>Wikipedia article: Longitude</p>	<p>term does not exist in EN</p> <p>BT: 2283 - Dimensions (EN)</p> <p>NT: 14581 - Stem length (EN)</p> <p>NT: 24232 - Root length (EN)</p>
---	--

- **Avaluació:** ambdós termes es poden traduir com a longitud i per tant, les etiquetes d'idioma (per exemple, castellà) poden haver-los identificat com a conceptes equivalents. Però el primer terme es refereix a la distància amb la que es mesuren les coordenades geogràfiques i el segon terme és refereix a la longitud o llargària com una mesura de dimensió. Ambdós són tipus de mesures, podrien ser termes relacionats però mai equivalents per que parlen de contextos diferents, per tant, en aquests casos és la col·lecció a indexar la que determina si són termes relacionats o no.

1.10 Alineament 10: services providing company *versus* service industry



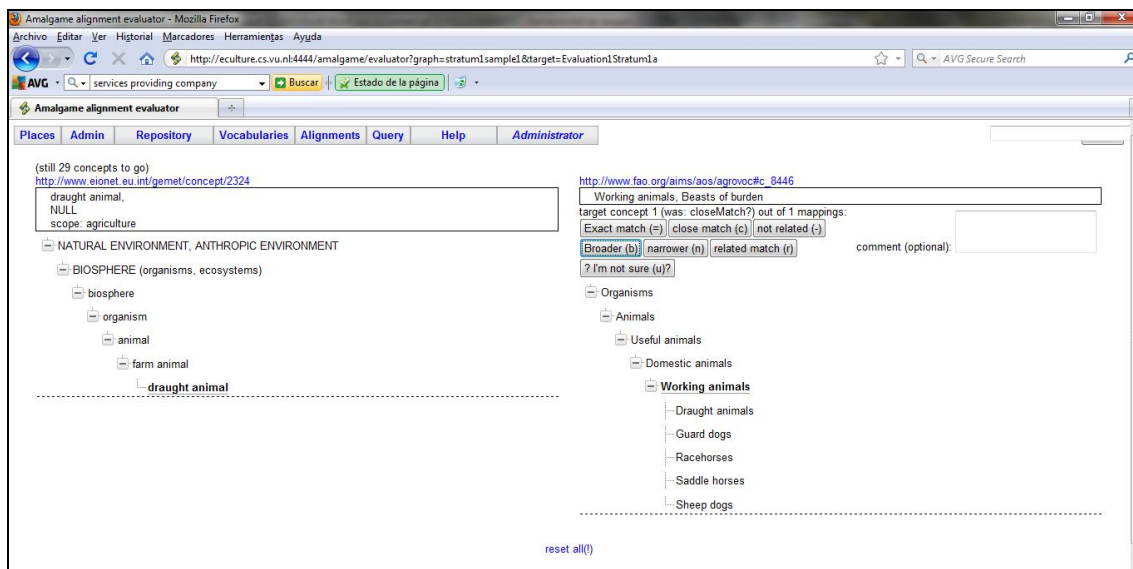
- **Avaluació:** relació d'equivalència exacta o "exact match". La decisió es presa perquè la resta de conceptes de la jerarquia ajuden a representar el mateix concepte. Ambdós vocabularis empen conceptes diferents però amb el mateix significat, serveis provinents de la indústria turística o hotelera.

1.11 Alineament 11: coal *versus* carbon



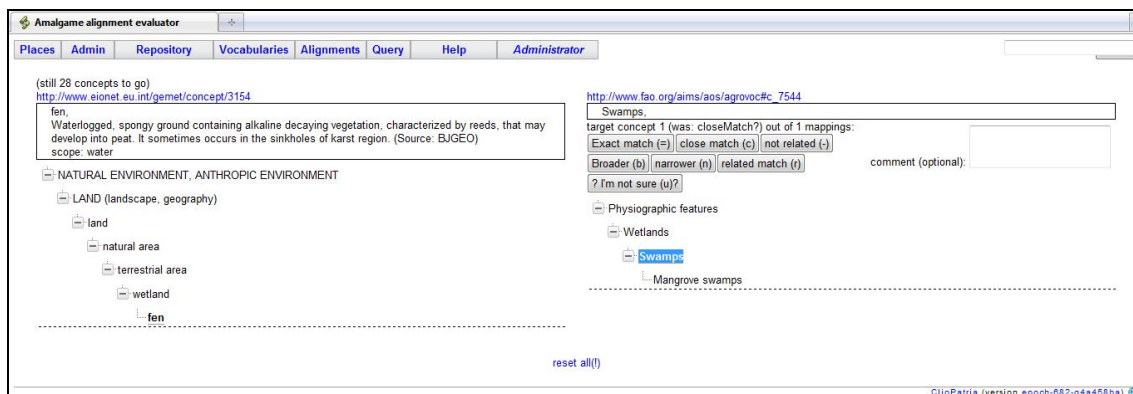
- **Avaluació:** Pot ser considerat tant un terme equivalent exacte com un terme quasi-equivalent. Ja que tot i ser conceptes diferents, un representa material i l'altre el símbol que representa el material com a element químic. No obstant això, en aquest cas potser considerat com a equivalència exacta, perquè AGROVOC representa el concepte “coal” dintre la jerarquia del concepte “Carbon C (Symbol)”, com un terme no acceptat, i com a termes preferit o acceptats considera els conceptes que representen els elements. Per tant, l'ús del concepte per part dels dos tesaurus és el mateix.

1.12 Alineament 12: draught animal *versus* working animals



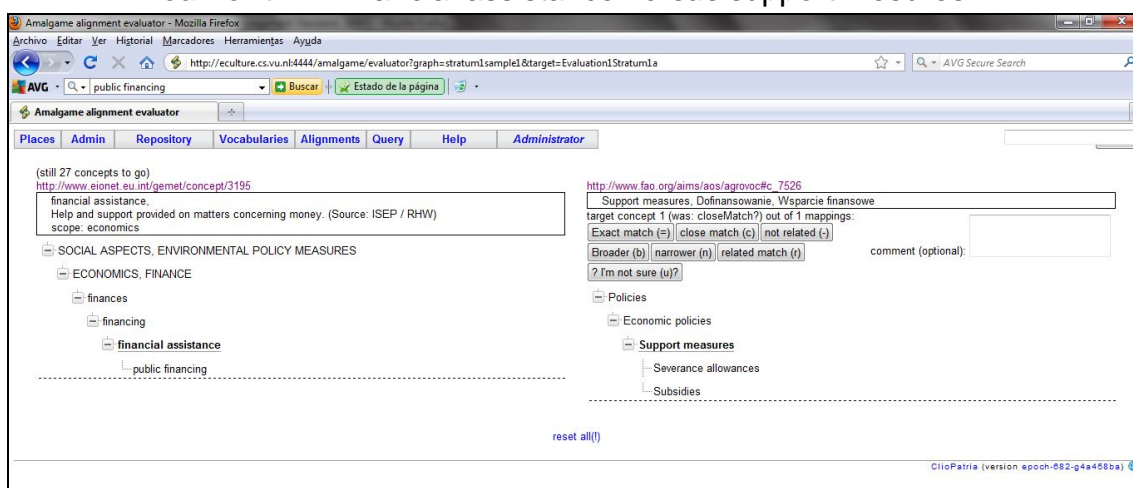
- **Comentari:** alineament de tipus de relació BT (terme superior). Tal i com s'observa en la jerarquia d'AGROVOC (representada en el concepte de la part dreta del gràfic), “draught animal” és un tipus de “Working Animal”.

1.13 Alineament 13: Fen *versus* Swamps



- **Avaluació:** els dos conceptes semblen sinònims perquè segons la llengua potser traduït amb el mateix terme (pantà/pantans), però en realitat són dos tipus de zones humides (wetlands) diferents¹²⁰. És pot dir que són **termes relacionats**, i com s'observa en la imatge superior, ambdós termes pengen del mateix concepte superior o BT, Wetlands. Un cop més, s'observa com les ontologies són una eina per trobar nous conceptes i dotar d'una major cobertura el vocabulari controlat principal.

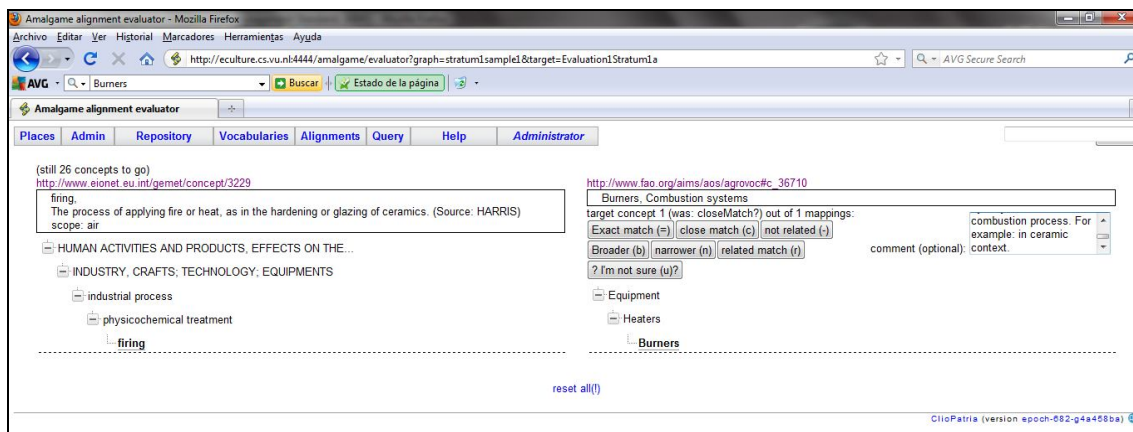
1.14 Alineament 14: financial assistance *versus* support measures



- **Avaluació:** Ambdós conceptes fan referència a mesures de suport econòmic, però es considera un terme quasi sinònim (*close match*), perquè ambdós termes tenen significats diferents tot i que poden ser emprats per designar o descriure el mateix tipus d'informació, tal i com s'observa en la representació de termes específics similars: finançament públic *versus* subsidis.

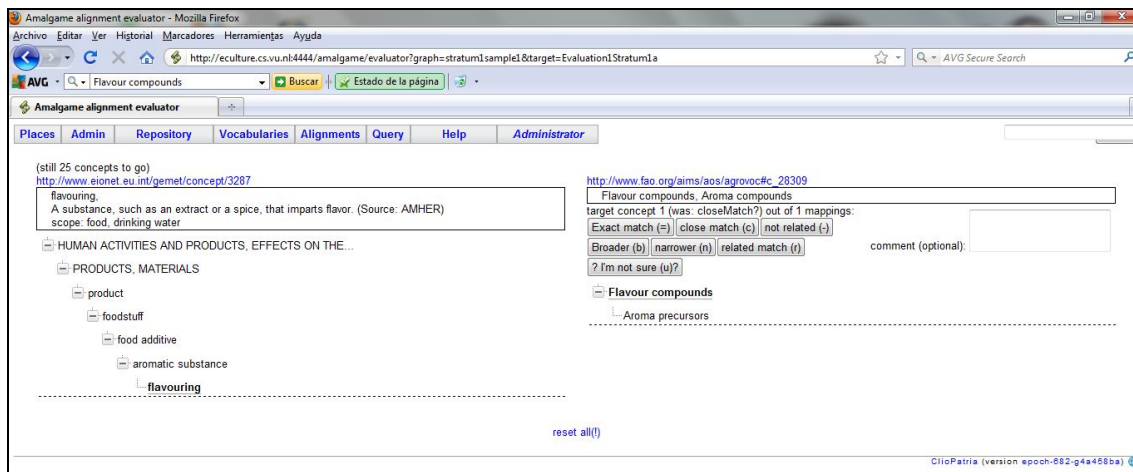
1.15 Alineament 15: Firing *versus* Burners

¹²⁰ About Wetlands. Environment Canada. Canada. Government. Darrera modificació: 17/07/2013. Url: <http://www.ec.gc.ca/tho-wlo/default.asp?lang=En&n=B4669525-1>; [Darrera Consulta: 3/11/2013]



- **Avaluació:** ambdós termes poden ser considerats sinònims, sobretot com a processos de combustió (per exemple en el context ceràmic). Però com tenen significats diferents es consideren quasi sinònims o “Close Match”.

1.16 Alineament 16: Flavour compounds versus flavouring



- **Avaluació:** ambdós conceptes estan relacionats amb el concepte de “substàncies aromàtiques”, concepte genèric de “flavouring”. Però ambdós tesaurus tenen un enfocament diferent del terme. GEMET considera “flavouring” com un producte o material d’àmbit alimentari. I AGROVOC contempla el terme com un component orgànic, com un compost del sabor.

1.17 Alineament 17: flue gas versus smokes

Flue gas (GEMET). Url: http://www.eionet.europa.eu/gemet/concept?cp=3319&langcode=en&ns=1	Smokes (AGROVOC). Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=7117&mylang_interface=en&myLanguage=EN
--	--

<p>Broader terms: combustion gas</p> <p>Scope note: not available</p> <p>Groups: WASTES, POLLUTANTS, POLLUTION</p> <p>Themes: air</p> <p>Other relations:</p> <hr/> <ul style="list-style-type: none"> - Wikipedia article: Flue gas - Has close match: UMTHES: Rauchgas 	<p>UF 1327917407562 - this term does not exist in EN</p> <p>UF 1339129807481 - this term does not exist in EN</p> <p>BT 2969 - Flavourings (EN)</p> <p>RT 3142 - Fumigation (EN)</p> <hr/> <p>RT 6172 - Preservatives (EN)</p>
---	--

- **Avaluació:** el primer concepte es refereix a un producte gasós de combustió que és generat per un forn, i el segon terme es refereix a un producte alimentari, de fet el seu terme genèric és un Flavourings.

1.18 Alineament 18: Forest management versus Forestry

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'forest management' is defined as 'Planning of forest utilization for wood production, conservation purposes, fauna and flora protection, recreation and water supply. (Source: RFA) scope: forestry'. On the right, the AGROVOC concept 'Forestry' is defined as 'Forestry, Dasonómia. Station forestària scope: Science, business and art of creating, conserving and managing forests'. The interface also shows a tree structure for 'Forestry' with sub-terms: 'Community forestry', 'Silviculture', and 'Urban forestry'. Mapping options include 'Exact match (=)', 'close match (c)', 'not related (-)', 'Broader (b)', 'narrower (n)', 'related match (r)', and '? I'm not sure (u)?'. A 'reset all()' button is at the bottom.

- **Avaluació:** són termes relacionats, tal i com es defineix en la versió actualitzada d'ambdós tesaurus. Visualitza la taula següent

<p>Forest Management (GEMET). Url: http://www.eionet.europa.eu/gemet/concept?cp=3425&langcode=en&ns=1</p>	<p>Forestry (AGROVOC). Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=3055&mylang_interface=en&myLanguage=EN</p>
<p>Definition: Planning of forest utilization for wood production, conservation purposes, fauna and</p>	<p>BT 330834 - activities (EN) BT 37989 - sciences (EN) NT 32620 – Urban forestry (EN) NT 7071 - silviculture (EN)</p>

<p>flora protection, recreation and water supply. (Source: RFA)</p> <p>Broader terms: environmental management</p> <p>Related terms: forestry</p> <p>Scope note: is not available</p> <p>Groups: Environmental policy</p> <p>Themes: environment policy; forestry</p> <p>Other relations:</p> <ul style="list-style-type: none"> - Has exact match: AGROVOC: Forest management - Wikipedia article: Forest management - Has close match: UMTHES: Forstwirtschaft 	<p>NT 16532 - Community forestry (EN)</p> <p>RT 34830 - Forest nurseries (EN)</p> <p>RT 36729 - Forest rehabilitation (EN)</p> <p>RT 50212 – forest USER GROUPS (EN)</p> <p>RT 3060 - Forestry policies (EN)</p> <p>RT 3059 - forestry operations (EN)</p> <p>RT 3057 - Forestry development (EN)</p> <p>RT 3056 - Forestry biology</p> <p>RT 3052 - forest trees</p> <p>RT 3049 - Forest products</p> <p>RT 16129 - Forest management</p> <p>RT 16209 - Forest mensuration</p> <p>RT 16550 - Forest surveys (EN)</p> <p>RT 207 - agroforestry (EN)</p> <p>RT 28075 - Forest protection (EN)</p> <p>RT 28077 - Forest returns (EN)</p> <p>RT 28078 - forest pathology (EN)</p> <p>RT 28079 - forest regulations (EN)</p> <p>RT 28173 - Tree classes (EN)</p> <p>RT 1336142720153 - forestry statistics (EN)</p>
---	---

*Darrera consulta 3/11/2013

1.19 Alineament 19: Gastropod versus snails

The screenshot shows the Amalgame alignment evaluator interface. On the left, a concept 'gastropod' is selected from a hierarchical tree under 'NATURAL ENVIRONMENT, ANTHROPIC ENVIRONMENT' > 'BIOSPHERE' > 'organism' > 'animal' > 'invertebrate' > 'mollusc'. The definition states: 'Any mollusc of the class Gastropoda, typically having a flattened muscular foot for locomotion and a head that bears stalked eyes. (Source: CED)'. On the right, a target concept 'Snails, Land snails' is selected from 'http://www.fao.org/aims/aos/agrovoc/c_7121'. The definition states: 'Snails, Land snails scope: Chiocciolle'. The interface shows a 'target concept 1 (was: closeMatch?) out of 1 mappings:' with a selected 'Exact match (=)' and other options like 'close match (c)', 'not related (-)', 'Broader (b)', 'narrower (n)', 'related match (r)', and '? I'm not sure (u)?'. A 'comment (optional):' field is also present.

- **Avaluació:** els cargols són un tipus de mol·lusc gastròpode. Aquí doncs “Gastropod” és un terme genèric (BT) de “snails”.

1.20 Alineament 20: incineration versus burning, flaming

The screenshot shows the Amalgam alignment evaluator interface. On the left, a concept from GEMET is displayed: 'incineration' (http://www.eionet.eu.int/gemet/concept/4182), defined as 'Controlled process by which solid, liquid, or gaseous combustible wastes are burned and changed into gases, residue produced contains little or no combustible material. (Source: LANDY)'. Its scope is 'air'. The hierarchy includes 'HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE...', 'INDUSTRY, CRAFTS, TECHNOLOGY, EQUIPMENTS', 'industrial process', 'physical treatment', 'incineration', and 'co-incineration'. On the right, a target concept from FAO is shown: 'Burning, Flaming' (http://www.fao.org/aims/aos/agrovoc/c_9499), with a scope of 'Foods that are more or less ready for consumption without further preparation other than heating'. The alignment is 'close match (c)'. A comment box contains the text: 'They are not synonyms but they almost mean the same thing.' The interface also includes navigation tabs (Places, Admin, Repository, Vocabularies, Alignments, Query, Help, Administrator) and a 'reset all()' button.

- **Avaluació:** no són sinònims però gairebé tenen el mateix significat. Es considera que poden tenir una relació de tipus “close match” (quasi sinònims), que segons el punt de vista podria ser considerada també com a un tipus de relació jeràrquica, considerant “Burning” com a terme superior, ja que la incineració pot ser definida com un tipus de crema (cremar fins arribar a les cendres).

1.21 Alineament 21: intermediate goods versus prepared foods

The screenshot shows the Amalgam alignment evaluator interface. On the left, a concept from GEMET is displayed: 'intermediate goods' (http://www.eionet.eu.int/gemet/concept/4410), defined as 'Partly finished goods or products that re-enter into production elsewhere. (Source: ISEP / MGHME)'. Its scope is 'economics'. The hierarchy includes 'SOCIAL ASPECTS, ENVIRONMENTAL POLICY MEASURES', 'ECONOMICS, FINANCE', 'economy', 'goods and services', 'goods', and 'intermediate goods'. On the right, a target concept from FAO is shown: 'Prepared foods, Convenience foods' (http://www.fao.org/aims/aos/agrovoc/c_6170), with a scope of 'Foods that are more or less ready for consumption without further preparation other than heating'. The alignment is 'close match (c)'. A comment box is empty. The hierarchy includes 'Processed products', 'Prepared foods', 'Icecream', 'Infant foods', 'Instant foods', and 'Soyfoods'. The interface also includes navigation tabs (Places, Admin, Repository, Vocabularies, Alignments, Query, Help, Administrator) and a 'reset all()' button.

- **Avaluació:** termes no relacionats, el concepte de GEMET fa referència a un producte un bé dintre l'àmbit econòmic, mentre que “prepared foods” és un terme relacionat amb un tipus de producte alimentari.

1.22 Alineament 22: land planning versus regional planning, interregional planning

The screenshot shows the Amalgame alignment evaluator interface. On the left, a concept from GEMET is displayed: 'land planning' (concept 4635), defined as 'The activity of designing, organizing or preparing for the future use of solid areas of the earth's surface, especially regions valued for natural resources, utilized as agricultural resources or considered for human settlement. (Source: RHW) scope: natural areas, landscape, ecosystems'. It is categorized under 'SOCIAL ASPECTS, ENVIRONMENTAL POLICY MEASURES' and 'ENVIRONMENTAL POLICY', with a hierarchical structure: environmental planning > physical planning > land planning > building planning, coastal zone planning, land allotment, land use planning, urban planning. On the right, a concept from AGROVOC is displayed: 'Regional planning, Interregional planning' (concept 6490). The interface shows a 'target concept 1 (was: closeMatch?) out of 1 mappings' with options for 'Exact match (=)', 'close match (c)', and 'not related (-)'. There are also buttons for 'Broader (b)', 'narrower (n)', 'related match (r)', and '? I'm not sure (u)?'. A 'comment (optional):' field is present. The status bar at the bottom indicates 'ClicPatris (version epoch-682-g4a458ba)'.

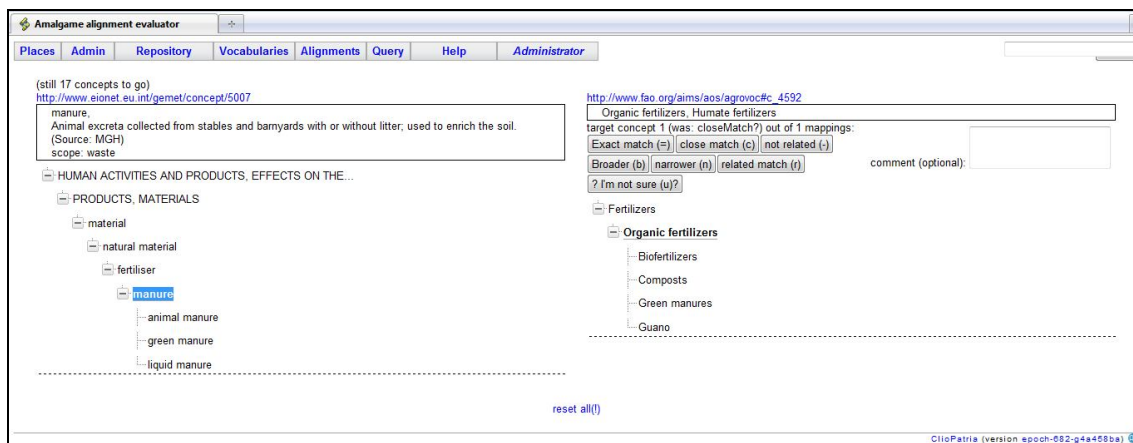
- **Avaluació:** “land planning” es traduït per GEMET al espanyol com a “ordenació del territori”, terme referent a la superfície de la terra especialment a regions de recursos naturals. En canvi, el concepte de “planificació regional” fa referència a un concepte més polític o socioeconòmic. No són termes relacionats.

1.23 Alineament 23: landscape versus landscaping

The screenshot shows the Amalgame alignment evaluator interface. On the left, a concept from GEMET is displayed: 'landscape management' (concept 4658), defined as 'Measures aiming at preserving landscape or controlling its transformations caused by anthropic activities or natural events. (Source: DIFID) scope: environmental policy'. It is categorized under 'SOCIAL ASPECTS, ENVIRONMENTAL POLICY MEASURES' and 'ENVIRONMENTAL POLICY', with a hierarchical structure: environmental management > landscape management. On the right, a concept from AGROVOC is displayed: 'Landscaping, Landscape architecture' (concept 4196). The interface shows a 'target concept 1 (was: closeMatch?) out of 1 mappings' with options for 'Exact match (=)', 'close match (c)', and 'not related (-)'. There are also buttons for 'Broader (b)', 'narrower (n)', 'related match (r)', and '? I'm not sure (u)?'. A 'comment (optional):' field is present. The status bar at the bottom indicates 'ClicPatris (version epoch-682-g4a458ba)'.

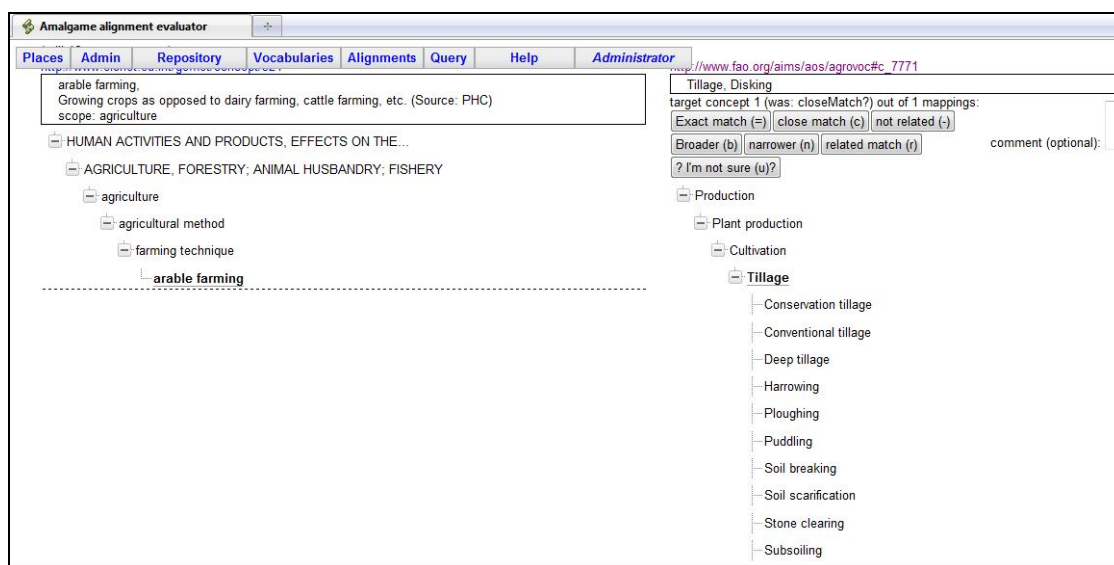
- **Avaluació:** Són conceptes relacionats. El primer terme, gestió del paisatge, tal i com és descrit de GEMET, fa referència a la preservació del paisatge i el segon terme, paisatgisme és considerat un terme relacionat a gestió del paisatge, tal i com figura en el tesaurus AGROVOC (*landscape conservation*): http://aims.fao.org/en/agrovoc-term-info?mytermcode=4186&mylang_interface=en&myLanguage=EN

1.24 Alineament 24: manure versus organic fertilizers



- **Avaluació:** tal i com apareix en la definició dels conceptes, són termes que es poden considerar quasi sinònims. Ja que un ferm és una substància orgànica, i dintre de la jerarquia (GMET) penja del concepte superior “fertilitzant” per tant, s’està considerant com un fertilitzant orgànic, tot i que els conceptes per separat tenen definicions diferents.

1.25 Alineament 25: Arable farming versus Tillage



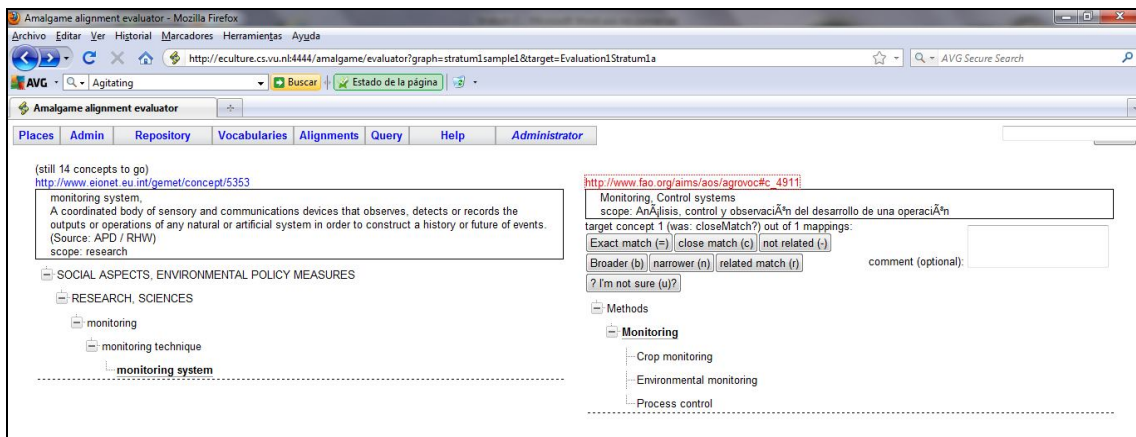
- **Comentaris:** El primer concepte, “arable farming”, fa referència a un tipus de tècnica i mètode agrari aplicat als cultius (traducció gemet: “cultiu del camp”) i en el segon terme, “tillage”, fa referència a l’acció de cultivar el camp. Ambdós termes són termes relacionats (RT), tot i que podrien ser fàcilment confosos per termes quasi sinònims perquè “tillage” potser interpretat també com una tècnica de cultiu.

1.26 Alineament 26: Mixing versus Agitating



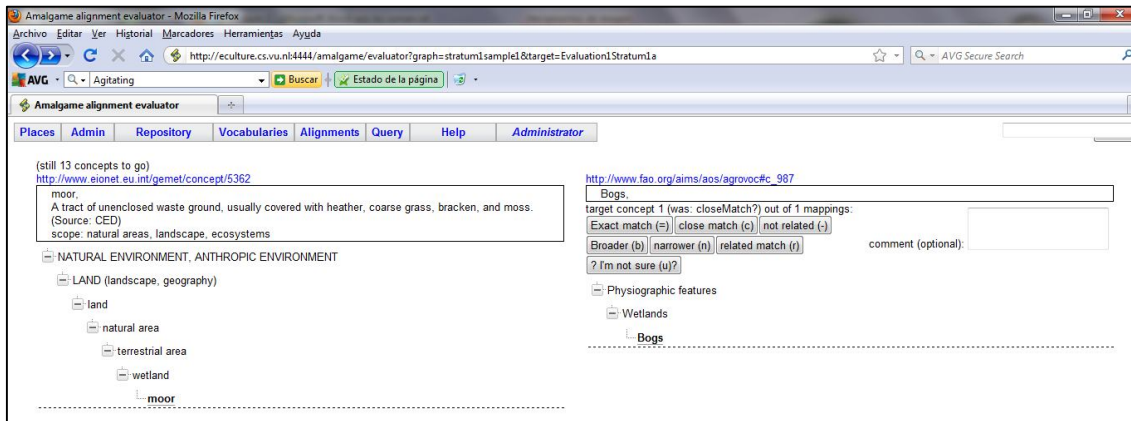
- **Avaluació:** Ambdós termes tenen un terme superior “sinònim” i inclús en AGROVOC, ambdós termes comparteixen el mateix terme genèric “Processing”, per tant, són termes relacionats (RT).

1.27 Alineament 27: Monitoring system *versus* Monitoring



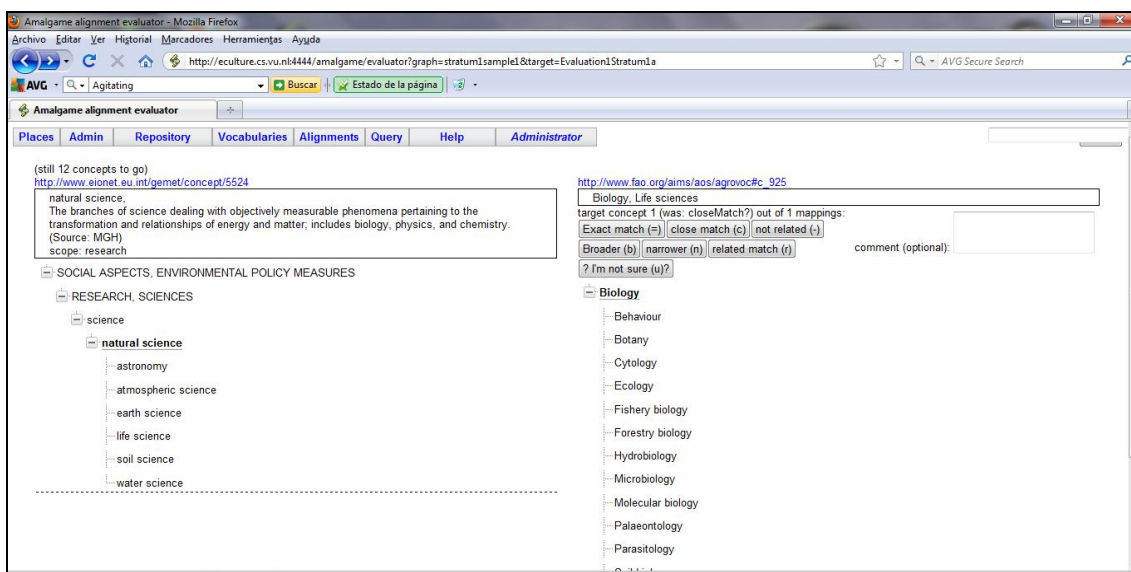
Avaluació: Ambdós termes segons criteris institucionals podrien ser presos com a termes equivalents, perquè el concepte en si de “Monitoring” ja té implícit en la seva definició que és un tipus de sistema, com apareix a Agrocvoc, és un tipus de sistema de control. No obstant això, si seguim, l’equivalència dels termes dintre dels vocabularis que s’estan alineant, doncs en la jerarquia de GEMET es veu clarament que “Monitoring” és un terme superior (BT) a “Monitoring system”.

1.28 Alineament 28: moor *versus* bogs



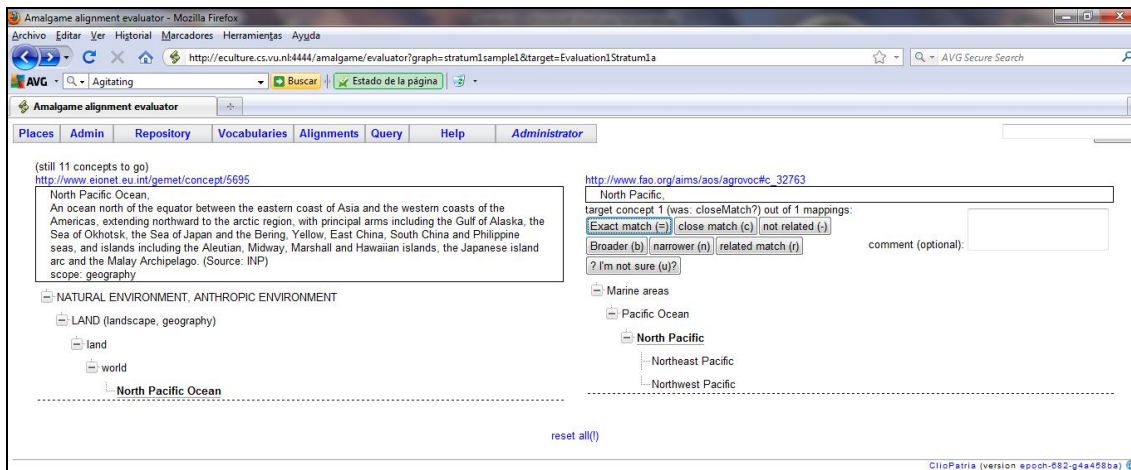
Avaluació: Són termes relacionats (RT), en aquest cas el sistema ho identifica clarament perquè el terme superior d'ambdós termes és el mateix. Fet que identifica que els dos termes són Zones Humides (o ecosistemes aquàtics). A més a més, en la versió actualitzada d'AGROVOC ambdós termes estan al mateix nivell jeràrquic i són identificats com a termes específics (NT) del BT *Wetlands*.

1.29 Alineament 29: Natural Sciences versus Biology



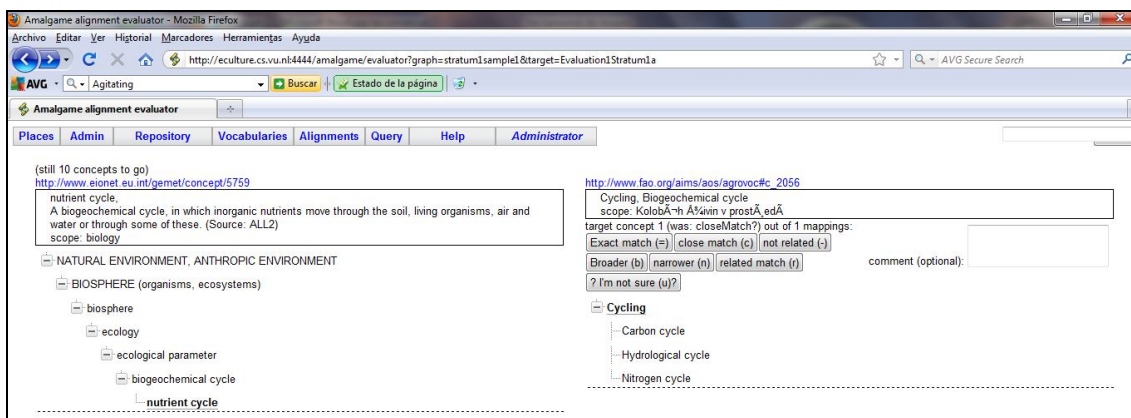
- **Avaluació:** En la nota d'abast de GEMET del concept "Natural Science", ja es defineix que "Biology" és una branca que forma part del propi concepte. El terme superior de Biologia són "life science" que ahora és un terme específic de "Natural Science". Lògicament "Natural Science" és un terme superior a *Biology*, però no es pot considerar que sigui una relació de "Broader Term", així doncs, es tipificarien com a un tipus de relació associativa o RT.

1.30. Alineament 30: North Pacific Ocean versus North Pacific



- **Avaluació:** ambdós termes estan representats amb una jerarquia diferent dintre dels corresponents vocabularis, no obstant, són termes equivalents o tipus de relació “Exact Match”, perquè ambdós representen el mateix concepte, l'àmbit geogràfic Oceà Pacífic, subdivisió Pacífic Nord.

1.31. Alineament 31: “Nutrient Cycle” versus Cycling (Use for Biogeochemical cycle)

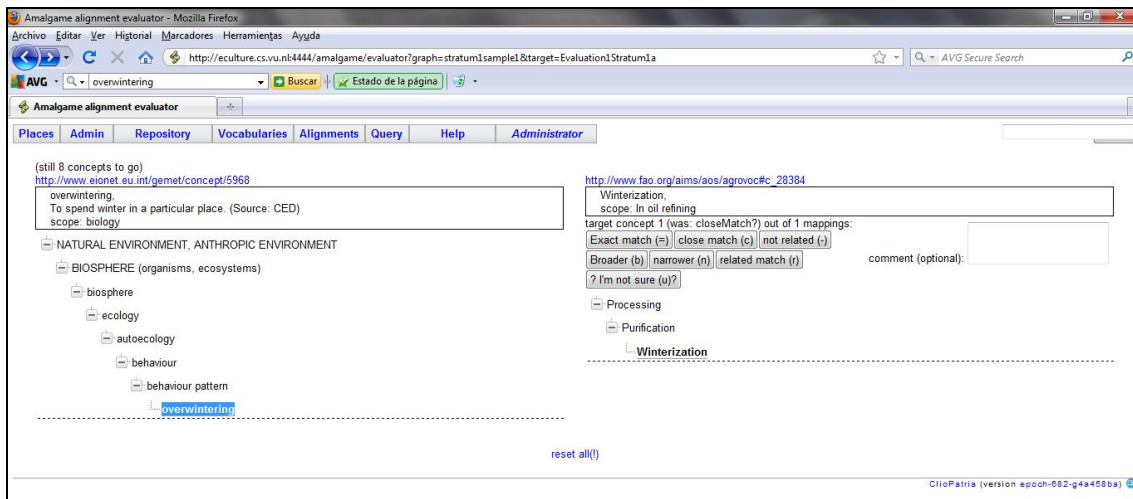


- **Avaluació:** Tipus de relació “Broader Term” (genèrica o BT). El sistema ha identificat que el concepte genèric de “Nutrient Cycle” (“biogeochemical cycle”), és equivalent al concepte recuperat pel sistema d'AGROVOC “Cycling”. Aquesta relació l'ha fet a través de la identificació del terme alternatiu o no acceptat del tesaurus AGROVOC, “Biochemical cycle”. Cada tesaurus, doncs, identifica termes preferits diferents, però en aquest sentit el sistema de mapeig funciona correctament i els identifica igualment.

1.32. Alineament 32: Organic Substance versus Organic matter

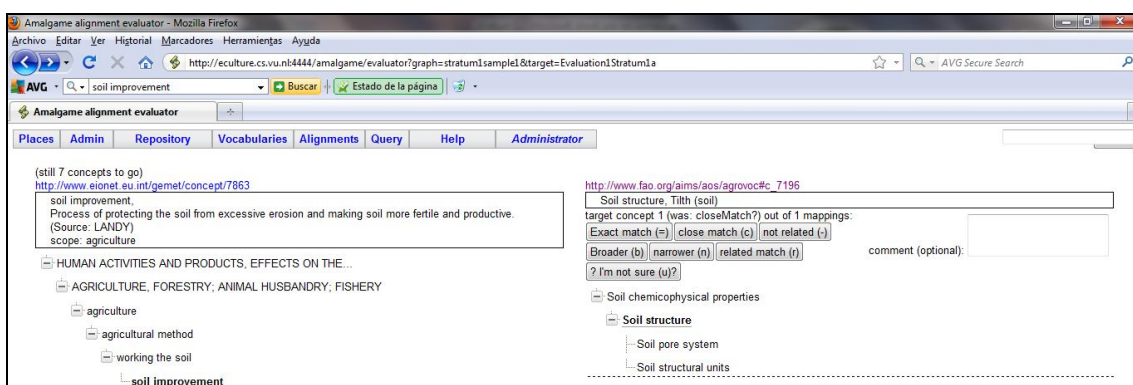
- **Avaluació:** “Matèria orgànica” és un concepte més genèric que “substància orgànica”, ja que la matèria orgànica està formada per substàncies i compostos orgànics. Però en aquest cas, AGROVOC utilitza el terme “organic matter” com a un terme sinònim, contemplant el concepte des del punt de vista químic, ja que el seu terme superior és “Substàncies” i substàncies és sinònim de compostos que és el terme que defineix GEMET com a substància orgànica. Segons Wikipedia, “La matèria orgànica i els compostos orgànics se'ls designa sota el terme de « substàncies » orgàniques” (Química Orgànica. Wikipedia, actualització 30/06/2013). Per tant, la decisió d'agafar els conceptes com a termes sinònims (o quasi-sinònims) o com a termes jeràrquics (BT-NT), dependrà de la cobertura i especificitat del vocabulari controlat a emprar. D'altra banda, aquesta alineació és un exemple també de diferències significatives a nivell de cobertura dels tesaurus. En aquest cas i/o disciplina concreta, GEMET inclou un gran nombre de descriptors específics (NT), i AGROVOC no n'inclou cap (en la versió actualitzada a data 2013, només inclou 2 NT's).

1.33. Alineament 33: Winterization versus Overwinter



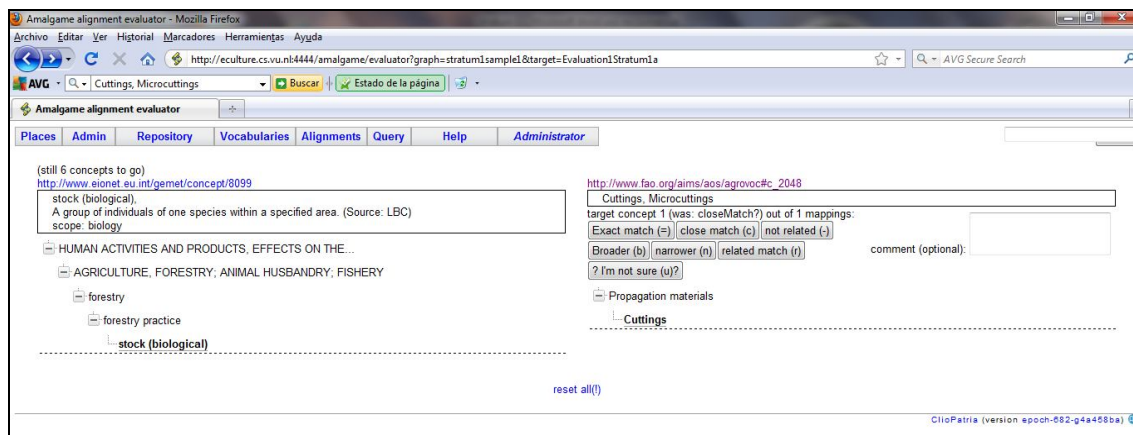
- **Avaluació:** Tal i com s'observa d'entrada en la jerarquia dels dos tesaurus que es comparen, són termes no relacionats (*non related*). Però a més a més, en la seva definició són també conceptes diferents. **Winterization** es refereix al procés de preparar alguna cosa per a un proper hivern. També s'utilitza comunament en relació a característiques aquàtiques, com fonts o similars, les quals han de ser drenades o segellades per a que l'aigua del interior no es congeli o que no es causi trencament en les canonades. O com es defineix en AGROVOC, com un procés centrat en el refinament d'oli. En canvi, **Overwinter** es refereix a passar o esperar durant el transcurs de hivern, fent les activitats normals o amb una supervivència complicada o impossible degut a les condicions atmosfèriques severes (baixes temperatures, neu, gel, subministrament d'aliments limitat, etc.).

1.34. Alineament 34: Soil improvement versus Soil structure



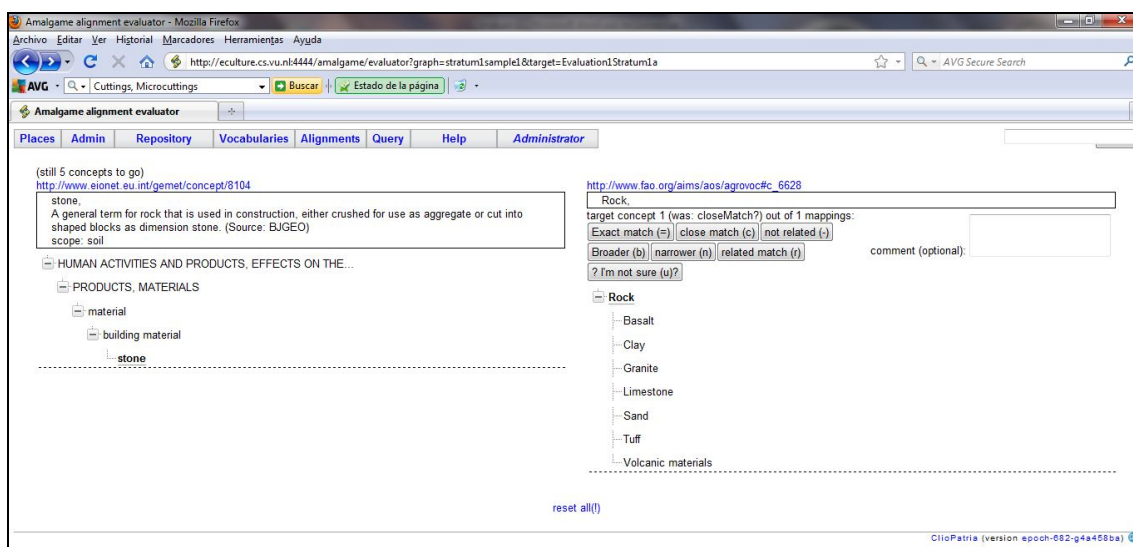
- **Avaluació:** Tal i com defineixen cada tesaurus, són termes no relacionats (Not related). "Soil improvement" fa referència al procés de protegir el sòl de l'erosió excessiva (definició extreta de GEMET), i "soil structure" fa referència a la disposició de les partícules del sòl en grups o individualment, donant una textura fluixa o amb fermesa definició extreta d'AGROVOC).

1.35. Alineament 35: Stock (biological) versus Cuttings



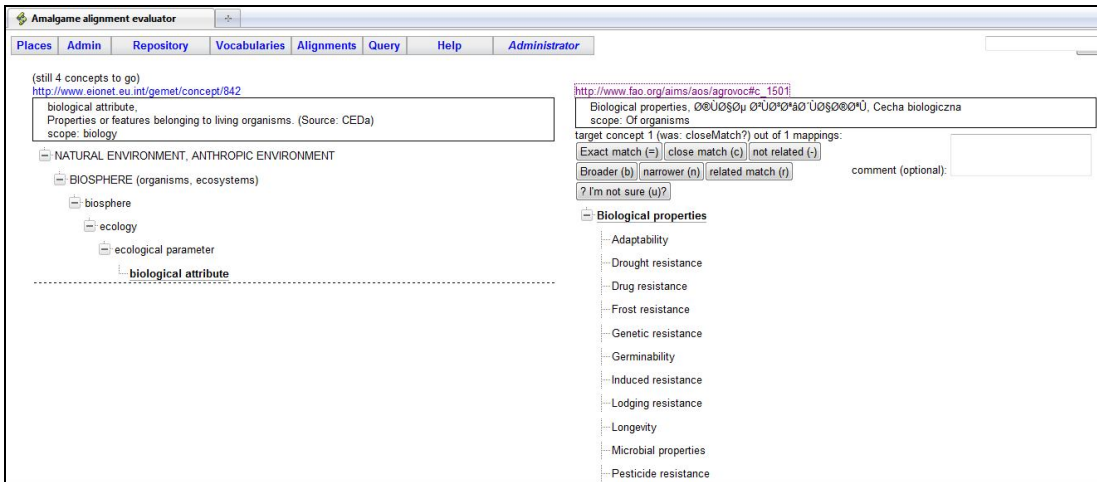
- **Avaluació:** Termes no relacionats. Tant l'estructura com les definicions dels termes són completament diferents. Cuttings o esqueixos (AGROVOC) són fragments de plantes separats amb una finalitat reproductor, i estocs (biològics) fa referència a un grup d'individus dintre d'una espècie (GEMET).

1.36. Alineament 36: Stone versus Rock



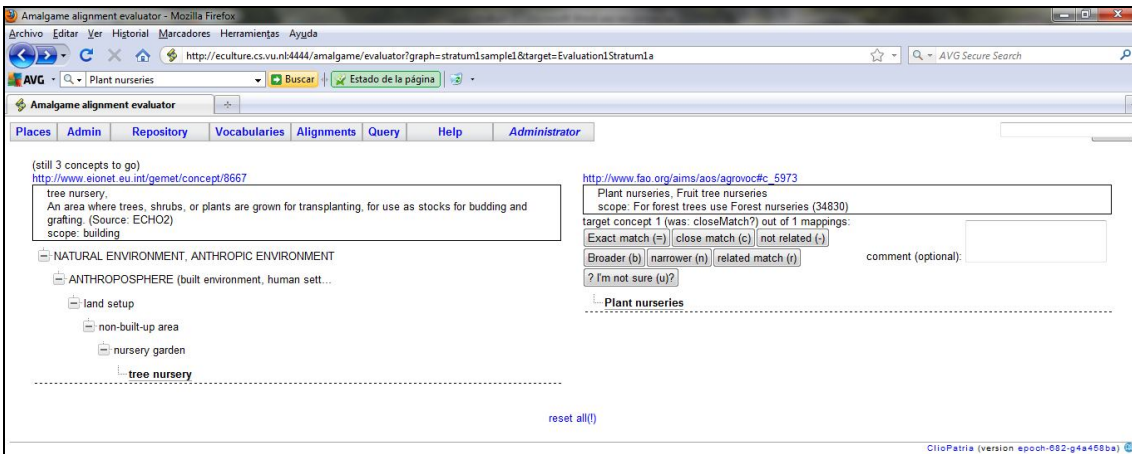
- **Avaluació:** Ambdós termes són termes relacionats perquè fan referència a tipus de material o substància. En AGROVOC, de fet, el terme "Stone" és un terme relacionat del concepte "Rock". No obstant això, el concepte "rock" a vegades és definit també com a "Stone".

1.37. Alineament 37: Biological attribute versus Biological properties



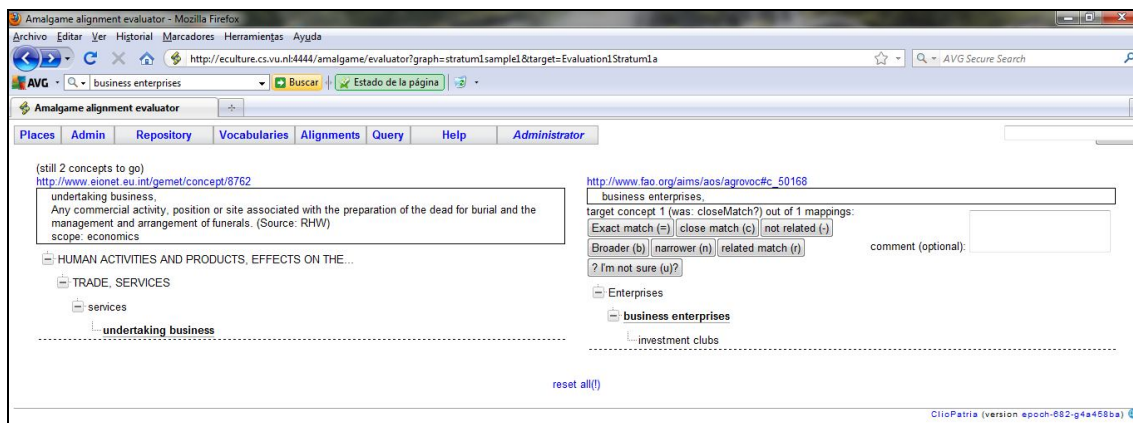
- **Avaluació:** Segons les definicions i notes d'abast dels propis tesaurus, els dos conceptes presentats a dalt són relacions de tipus "exact match", ja que ambdós fan referència a característiques dels organismes vius.

1.38 Alineament 28: Tree nursery versus "plant nurseries"



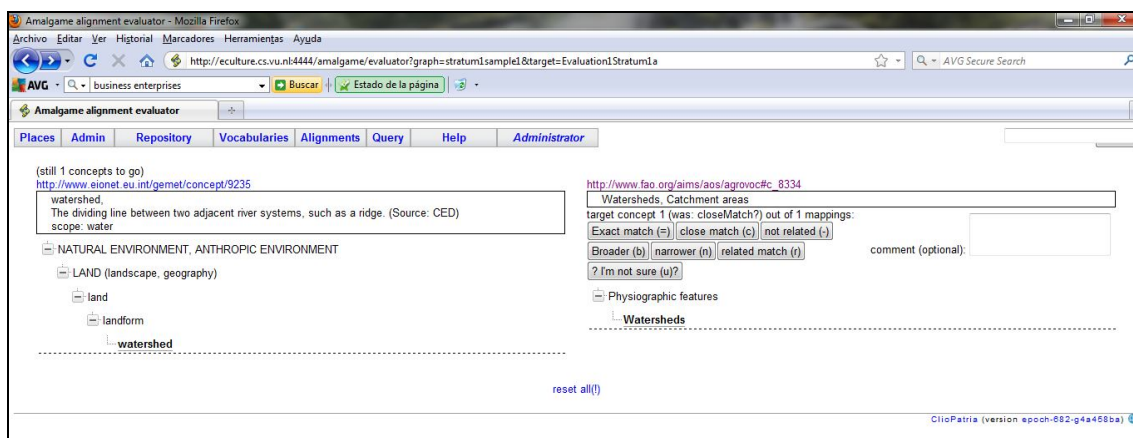
- **Avaluació:** Ambdós termes no tenen equivalències exactes en els seus vocabularis respectius. No, en AGROVOC si existeix els terme específics "fruit tree nurseries" i "forest nurseries", així doncs s'enten que "tree nursery" podria ser considerat un terme específic o NT (*Narrower Term*).

1.39 Alineament 39: undertaking business versus business enterprises



- **Avaluació:** Termes no relacionats. El primer concepte, “undertaking business” fa referència a qualsevol activitat (o servei) comercial, i en la nota d’abast concreta amb activitats relacionades amb la gestió de funerals. En canvi el segon concepte fa referència a empreses de tipus comercial.

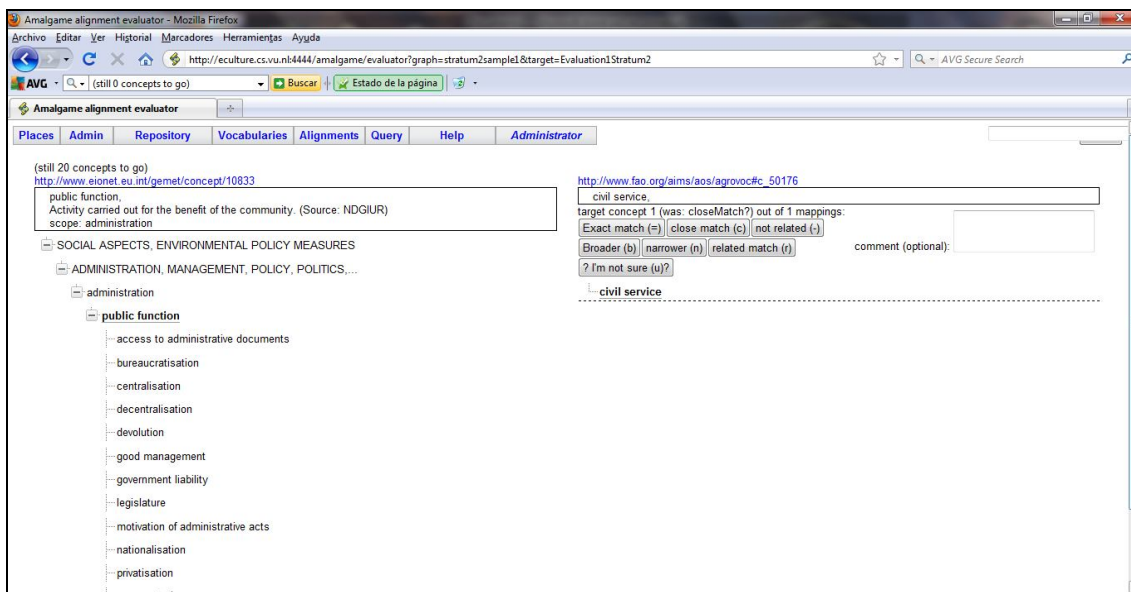
1.40 Alineament 40: Watershed versus Watersheds



- **Avaluació:** Relació de tipus exacte o “Exact Match”. La única diferència entre els termes és que GEMET utilitza el terme en singular i AGROVOC amb plural.

Stratum 2: Mostra aleatòria de 20 alineacions

2.1 Alineament 1: public service versus civil service



- **Avaluació:** Tipus de relació “Narrower Term”. En els tesaurus que es presenten, “Civil service” d’ AGROVOC, és el terme específic de “Public services” (AGROVOC), que apareix a GEMET com a terme equivalent de “Public function”. Però són termes que segons la traducció poden arribar a ser termes equivalents, per exemple, en AGROVOC, la traducció de “Civil service” a l’espanyol és el terme “Funció Pública”. Veure exemple:

<p>Public function (GEMET). Url: http://www.eionet.europa.eu/gemet/concept?cp=6817&langcode=en&ns=1</p>	<p>Civil service (AGROVOC). Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=50176&mylang_interface=en&myLanguage=EN</p>
<p>Broader terms: Services</p> <p>Narrower terms:</p> <ul style="list-style-type: none"> - air conditioning - energy supply - gas supply - heat supply - library service - lighting - municipal cleansing - on-line service - telematics - transportation business - water supply <p>Scope note: not available</p> <p>Groups: TRADE, SERVICES</p>	<p>BT 6353 - Public services (EN)</p> <p>RT 11230 - Government (EN)</p> <p>RT 28938 - Public administration (EN)</p> <p>RT 50336 - bureaucracy (EN)</p> <hr/> <p>Función pública (ES - AGROVOC). Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=50176&mylang_interface=en&myLanguage=ES</p>

<p>Themes: social aspects, population trade, services</p> <p>Other relations:</p> <p>Has exact match</p> <ul style="list-style-type: none"> - AGROVOC: Public services <p>Has close match</p> <ul style="list-style-type: none"> - <u>UMTHES: Öffentliche Dienste</u> <p>Has narrower match</p> <ul style="list-style-type: none"> - <u>EuroVoc: urban infrastructure</u> 	<p>BT 6353 - Servicios públicos (ES)</p> <p>RT 11230 - Gobierno (ES)</p> <p>RT 28938 - Administración pública (ES)</p> <p>RT 50336 - burocracia (ES)</p>
---	--

2.2 Alineament 2: Devolution *versus* Devolution

<p>Devolution (GEMET). Url: http://www.eionet.europa.eu/gemet/concept?cp=11632&langcode=en&ns=1</p>	<p>Devolution (AGROVOC). Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=9000032&mylang_interface=en&myLanguage=EN</p>
<p>Definition: The act of assigning or entrusting authority, powers or functions to another as deputy or agent, typically to a subordinate in the administrative structure of an organization or institution. (Source: RHW)</p> <p>Broader terms: public function</p> <p>Scope note: not available</p> <p>Groups: ADMINISTRATION, MANAGEMENT, POLICY, POLITICS, INSTITUTIONS, PLANNING</p> <p>Themes: administration</p> <p>Other relations</p> <p>Has exact match:</p> <ul style="list-style-type: none"> - AGROVOC: Devolution - EuroVov: devolution 	<p>BT 130 - administration (EN)</p> <p>BT 7141 - socioeconomic organization (EN)</p> <p>RT 2143 - decentralization (EN)</p>

Has a close match:	
- UMTHES: Devolution	

- **Avaluació:** Tipus de relació d'equivalència exacta o "Exact Match". Ambdós conceptes són idèntics, tant en definició com en tipus de relacions que indiquen que ambdós termes fan referència a la temàtica d'Administració o funció pública. Però a més, en la jerarquia del terme presentada per GEMET, s'estableix que hi ha una relació exacta amb el mateix terme en AGROVOC.

2.3 Alineament 3: cultivation system versus cropping system

Cultivation system (GEMET). Url: http://www.eionet.europa.eu/gemet/concept?cp=13207&langcode=en&ns=1	Cropping system (AGROVOC). Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=1971&mylang_interface=en&myLanguage=EN
Definition: Any overall structure or set-up used to organize the activity of preparing land or soil for the growth of new crops, or the activity of promoting or improving the growth of existing crops. (Source: RHW / AGP)	UF 10201 - cultivation systems (EN)
Broader terms: cultivation of agricultural land	BT 330985 - systems (EN)
Narrower terms:	NT 4986 - Multiple cropping (EN)
- Agrosystem	NT 4915 - Monoculture (EN)
- Forage crop	NT 4871 - Mixed cropping (EN)
- Slash and burn culture	NT 3910 - Intercropping (EN)
Scope note: not available	NT 6662 - Crop rotation (EN)
Groups: AGRICULTURE, FORESTRY; ANIMAL HUSBANDRY; FISHERY	NT 6908 - Seasonal cropping (EN)
Themes: Agriculture	NT 7038 - Shifting cultivation (EN)
Other relations:	NT 6977 - Sequential cropping (EN)
Has exact match:	NT 24935 - Off season cultivation (EN)
- EuroVoc: cultivation system	RT 6209 - production structure (EN)
	RT 7170 - soil fertility (EN)
	RT 36555 - Alternative methods (EN)
	RT 29762 - cultural methods (EN)
	RT 2807 - farming systems (EN)

	<p>RT 207 - agroforestry (EN)</p> <p>RT 2018 - Cultivation (EN)</p> <p>RT 1936 - Cover plants (EN)</p> <p>RT 16094 - Crop management (EN)</p> <p>RT 8075 - Upland crops (EN)</p>
--	--

- **Avaluació:** Relació de tipus “exact match”. La diferencia entre els dos tesaurus és que cadascun ha escollit un terme preferit diferent (vegeu concepte UF, en AGROVOC, i *Other relations*)

2.4 Alineament 4: agricultural exploitation versus farms

<p>Agricultural exploitation (GEMET). Url: http://www.eionet.europa.eu/gemet/concept?cp=199&langcode=en&ns=1</p>	<p>Farms (AGROVOC). Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=2809&mylang_interface=en&myLanguage=EN</p>
<p>Definition: no definition needed</p> <p>Broader terms: agricultural production</p> <p>Scope note: scope note is not available</p> <p>Groups: AGRICULTURE, FORESTRY; ANIMAL HUSBANDRY; FISHERY</p> <p>Themes: agriculture</p> <p>Other relations:</p> <p>Has exact match:</p> <ul style="list-style-type: none"> - EuroVoc: agricultural holding 	<p>BT 10626 - enterprises (EN)</p> <p>NT 36678 - Fish farms (EN)</p> <p>NT 34762 - Dairy farms (EN)</p> <p>NT 4616 - Market gardens (EN)</p> <p>NT 5879 - Pilot farms (EN)</p> <p>NT 7113 - small farms (EN)</p> <p>NT 28845 - collective farms (EN)</p> <p>NT 28732 - Medium size farms (EN)</p> <p>NT 28726 - Marginal farms (EN)</p> <p>NT 28720 - large farms (EN)</p> <p>NT 13581 - Private farms (EN)</p> <p>NT 7376 - State farms (EN)</p> <p>NT 25247 - Livestock farms (EN)</p> <p>NT 2787 - Family farms (EN)</p> <p>RT 2894 - Fields (EN)</p>

	<p>RT 34761 - Dairies (EN)</p> <p>RT 35102 - Farm transfers (EN)</p> <p>RT 28639 - Agroindustrial complexes (EN)</p> <p>RT 10749 - Farm models (EN)</p> <p>RT 10747 - Farm leases (EN)</p> <p>RT 10743 - Farm buildings (EN)</p> <p>RT 202 - agricultural structure (EN)</p>
--	--

- **Avaluació:** Les granges són definides per explotacions agropecuàries. I al mateix temps poden ser un tipus d'explotació agrària. Així doncs que segons el punt de vista en que s'apliqui el descriptor es pot afirmar que "Farms" té una relació de "Narrower term" amb "Agricultural exploitation", i al mateix temps es pot afirmar que ambdós són termes equivalents. En aquest cas, però, el thesaurus d'AGROVOC facilita també la designació de tipus de relació d'equivalència exacta, ja que la traducció de "farms" al espanyol és "explotaciones agrarias"¹²¹. Davant el coneixement de la definició del terme "farms" com a un tipus d'explotació agrària, i la resta de casuístiques, l'avaluador entén que estem al davant d'un tipus de relació de "quasi-sinònims" o "close match".

2.5 Alineament 5: Digester versus Digesters

Digester (GEMET). Url:	Digesters (AGROVOC). Url:
http://www.eionet.europa.eu/gemet/concept?cp=2188&langcode=en&ns=1	http://aims.fao.org/en/agrovoc-term-info?mytermcode=16120&mylang_interface=en&myLanguage=EN
<p>Definition: Machine which takes refuse and produces gas such as methane from it. (Source: PHC)</p> <p>Broader term: purification facility</p> <p>Scope note: not available</p> <p>Groups:</p>	<p>BT 2631 - Equipment (EN)</p> <p>RT 23990 - Anaerobiosis (EN)</p> <p>RT 25193 - Gasifiers (EN)</p> <p>RT 27469 - Bioreactors (EN)</p> <p>RT 34990 - Anaerobic treatment (EN)</p>

¹²¹ Explotaciones Agrarias. AGROVOC. Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=2809&mylang_interface=en&myLanguage=ES (Darrera consulta: 01/11/2013).

INDUSTRY, CRAFTS; TECHNOLOGY; EQUIPMENTS Themes: industry pollution water	
---	--

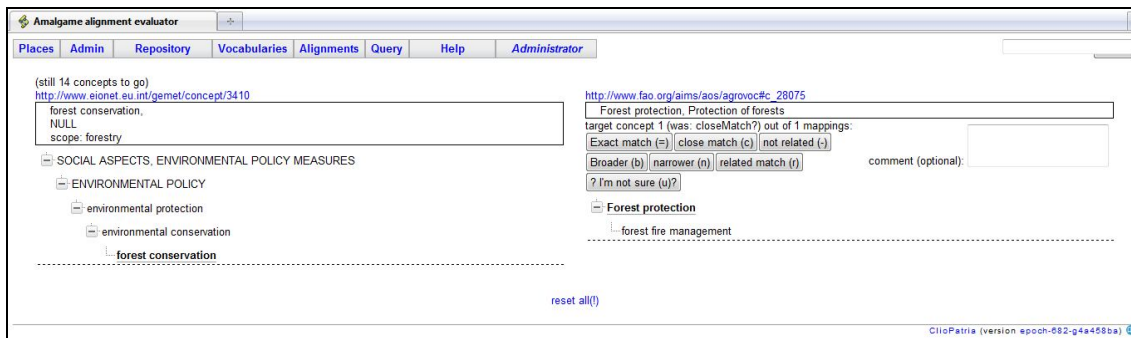
- **Comentaris:** relació de mapeig de tipus exacte o “exact match”. Ambdós tesaurus presenten la mateixa definició i traducció en diversos idiomes. Com en d’altres casos, la única diferencia és l’ús de singular del terme per part de GEMET i ús de plural en AGROVOC.

2.6 Alineament 6: Ecotype *versus* Ecotypes

Ecotype (GEMET). Url:	Ecotypes (AGROVOC). Url:
http://www.eionet.europa.eu/gemet/concept?cp=2531&langcode=en&ns=1	http://aims.fao.org/en/agrovoc-term-info?mytermcode=27505&mylang_interface=en&myLanguage=EN
Definition: Species that has special characteristics which allow it to live in a certain habitat. (Source: PHC) Broader terms: synecology Related terms: - Ecological adaptation - Ecosystem type Scope not: not available Groups: BIOSPHERE (organisms, ecosystems) Themes: biology, natural areas , landscape, ecosystems	UF 1326275534213 - this term does not exist in EN UF 66959 - this term does not exist in EN BT 7280 - Species (EN) RT 117 - Adaptation (EN) RT 16166 - Biotypes (EN) RT 32886 - Land varieties (EN) RT 4177 - Land races (EN)

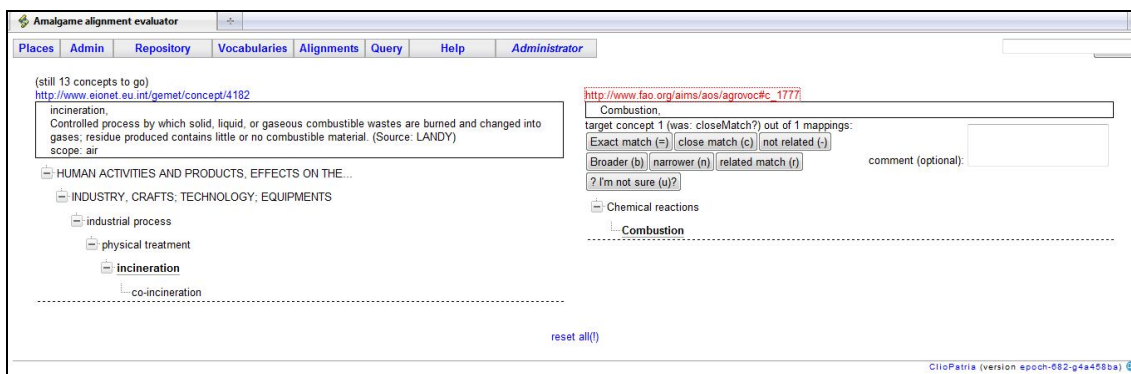
- **Avaluació:** relació de tipus exacte o “exact match”. La única diferencia és l’ús de singular del terme per part de GEMET i ús de plural en AGROVOC.

2.7 Alineament 7: forest conservation *versus* forest protection



- **Avaluació:** La conservació de boscos implica també la seva protecció, no obstant molts cops s'empren com a termes quasi sinònims. I en aquest cas, en ambdós tesaurus no es fa aquesta diferència, i només existeix un sol terme. A nivell de sistema, es pot dir que s'estableix un tipus de relació d'equivalència exacta o "exact match" que seria correcta pel context dels tesaurus, no obstant, la relació més adequada seria "close match", ja que les definicions de conservació i protecció tenen connotacions diferents, tot i que aquí s'empren per definir-les com accions que formen part de la gestió de boscos i protecció o conservació mediambiental.

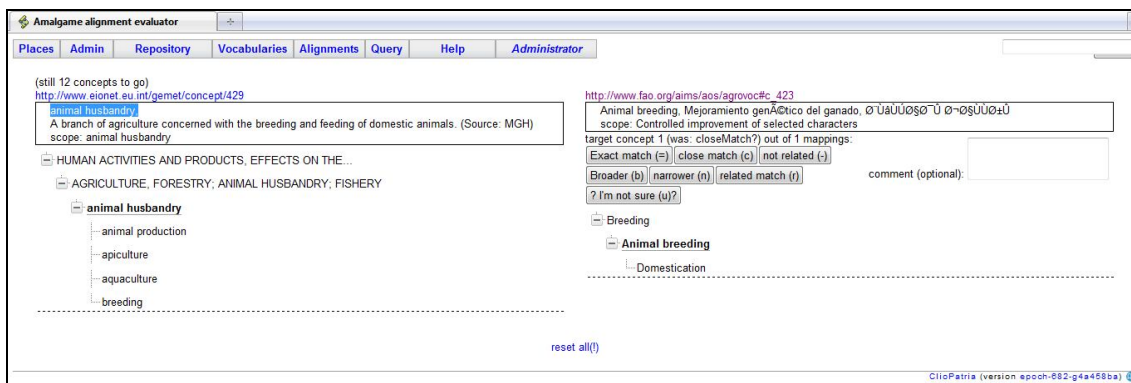
2.8 Alineament 8: incineration versus combustion



- **Avaluació:** incineració (GEMET) i combustió són termes no relacionats. Combustió fa referència a una reacció química que és la que produeix la combustió (procés d'oxidació amb despreniment de llum i calor, amb flama o sense), i incineració, és el procés d'incinerar o reduir alguna cosa a cendres¹²², GEMET ho descriu com un procés industrial de crema de residus (líquids, sòlids o gasosos) que són convertits a gas.

2.9 Alineament 9: Animal husbandry versus Animal breeding

¹²² Definicions extretes del Diccionari de la Llengua Catalana (Online). Grec.cat. Accessible a: <http://www.grec.cat/cqibin/mlt00.pgm>



- **Avaluació:** RELATED TERMS: Podrien ser tractar com a quasi sinònims o “close match”, però són termes relacionats (Related Terms), ja que “Animal husbandry” és referix a la producció animal; i “Animal breeding” fa referència a la a la cria d’animals “per a la millora animal”, de fet, AGROVOC tradueix el terme en castellà com “mejoramiento animal”. Ambdós tesaurus diferencien “Animal husbandry” i “Animal Breeding”.

2.10 Alineament 10: information service versus information services

Information service (GEMET). Url: http://www.eionet.europa.eu/gemet/concept?cp=4309&langcode=en&ns=1	Information services (AGROVOC). URL: http://aims.fao.org/en/agrovoc-term-info?mytermcode=3865&mylang_interface=en&myLanguage=EN
<p>Definition: An organized system of providing assistance or aid to individuals who are seeking information, such as by using databases and other information sources to communicate or supply knowledge or factual data.</p> <p>Broader terms: communications</p> <p>Scope note: not available</p> <p>Groups: INFORMATION, EDUCATION, CULTURE, ENVIRONMENTAL AWARENESS</p> <p>Themes: information</p>	<p>BT 6989 - Services (EN)</p> <p>NT 22565 - Agrinter (EN)</p> <p>NT 36525 - Market information services (EN)</p> <p>NT 36661 - Internet (EN)</p> <p>NT 3860 - diffusion of information (EN)</p> <p>RT 4311 - libraries (EN)</p> <p>RT 3861 - information needs (EN)</p> <p>RT 37829 – Access to information (EN)</p> <p>RT 36660 - information technology (EN)</p> <p>RT 24833 - Databases (EN)</p>

<p>Other relations:</p> <p>Has exact match</p> <ul style="list-style-type: none"> - AGROVOC: information services <p>Wikipedia article</p> <ul style="list-style-type: none"> - Information broker 	<p>RT 11769 - information systems (EN)</p>
---	--

- **Comentaris:** termes amb equivalència exacta. Tal i com indica el tesaurus GEMET en la descripció del terme, AGROVOC utilitza el mateix concepte però en plural, i s'indica que hi ha un tipus de relació "exact match".

2.11 Alineament 11: Low flow versus Low water

- **Avaluació:** termes que tot i que poden ser emprats amb diferents usos, en aquest cas, tal i com indica la jerarquia dels termes, ambdós són emprats per dir que hi ha un cabal baix de l'aigua. De fet GEMET, defineix "low flow" o "flujo bajo" com "phase of the lowest of a water course". I segons AGROVOC "Low water" és "ESTIAJE" o caudal baix o mínim d'un ecosistema aquàtic (riu o llacuna). Així doncs, en aquest context, estem davant d'una relació de tipus "close match" o de quasi sinònims.

2.12 Alineament 12: phomone versus phomones

- **Avaluació:** tipus d'equivalència exacta o “exact match”, la única diferència que s'aprecia és entre l'ús de singulars (GEMET) i plurals (AGROVOC). No obstant això, ambdós tesaurus tenen punts de vista diferents, i en la jerarquia superior d'aquests termes és veu clarament aquesta diferència a nivel de (Biologia i Química).

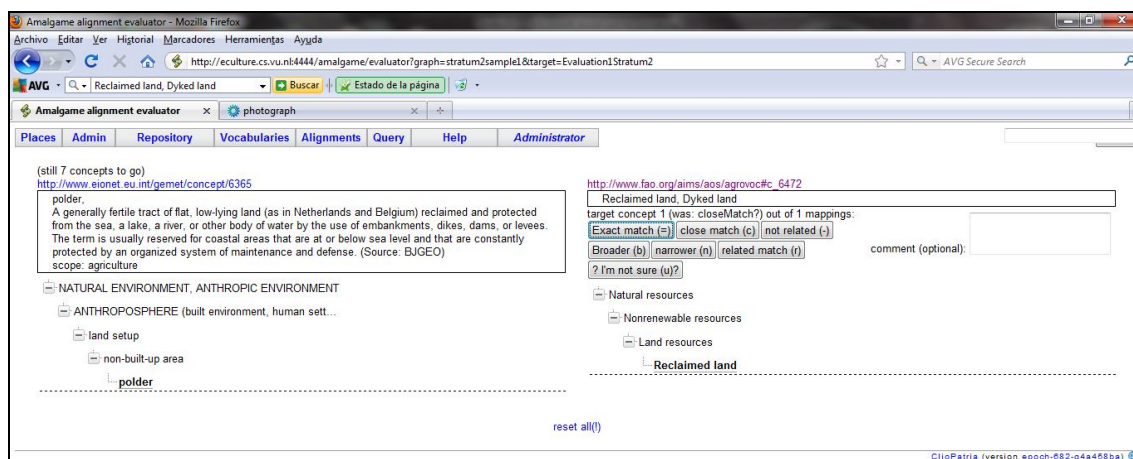
2.13 Alineament 13: Photograph *versus* Photography

Photograph (GEMET). Url: http://www.eionet.europa.eu/gemet/concept?cp=6205&langcode=en&ns=1	Photography (AGROVOC). Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=5808&mylang_interface=en&myLanguage=EN
<p>Definition: An image captured by a camera or some other device and reproduced as a picture, usually on a sensitized surface and formed by the chemical action of light or of radiant energy. (Source: MHD / RHW)</p> <p>Broader terms: documentation</p> <p>Narrower terms:</p> <ul style="list-style-type: none"> - Aerial photograph - Mosaic - Satellite image <p>Scope note:</p> <p>scope note is not available</p> <p>Groups:</p> <p>INFORMATION, EDUCATION, CULTURE, ENVIRONMENTAL AWARNESS</p> <p>Themes:</p> <p>information</p>	<p>BT 36760 - Imagery (EN)</p> <p>NT 8634 – Aerial photograhy</p> <p>NT 13319 - Photomicrography (EN)</p> <p>NT 5807 - Photogrammetry (EN)</p> <p>RT 37359 - Image Processing (EN)</p> <p>RT 6428 - Radiography (EN)</p> <p>RT 709 - audiovisual aids (EN)</p> <p>RT 2899 - films (EN)</p> <p>RT 16375 - Photointerpretation (EN)</p> <p>RT 9861 - cinematography (EN)</p>

- **Avaluació:** En GEMET contempla ambdós termes, utilitzant “photography.” per a designar el procés, és a dir “fotografiar” i “photograph” com al producte resultant, la imatge. En canvi, AGROVOC només utilitza un dels termes, “photography” i l'empra com a el producte o imatge. És a dir, per AGROVOC són termes sinònims, tot i que en realitat són termes relacionats. Així doncs, com no només la definició del terme per part d'AGROVOC (vegeu traducció del tesaurus del terme en

castellà), com l'estructura i representació jeràrquica del terme (termes específics similars), l'avaluador identifica aquí “photography” i “photograph” com a un tipus de relació de quasi sinònims o “close match”.

2.14. Alineament 14: Polder versus Reclaimed land



- **Avaluació:** Polder, és un terme emprat en Holanda i Bèlgica, que significa “terres recuperades del mar, rius o llacunes” que són aprofitades generalment per activitats agrícoles (wikipedia). El propi tesaurus GEMET identifica que té un tipus de relació d'equivalència exacta amb el terme “Reclaimed land” que és el concepte que ha alineat el sistema de mapeig. Per tant, podem dir que hi ha un tipus de relació “exact match”.

2.15 Alineament 15: Pollution versus Contamination

Pollution (GEMET). Url:	Contamination (AGROVOC). Url:
http://www.eionet.europa.eu/gemet/concept?cp=6445&langcode=en&ns=1	http://aims.fao.org/en/agrovoc-term-info?mytermcode=28317&mylang_interface=en&myLanguage=EN
Definition: The indirect or direct alteration of the biological, thermal, physical, or radioactive properties of any medium in such a way as to create a hazard or potential hazard to human health or to the health, safety or welfare of any living species. (Source: ALL)	BT6400 - Quality (EN) BT6077 - Pollution (EN) NT32389 - Bioaccumulation (EN) NT28321 - Radioactive contamination (EN) NT28319 – Chemical contamination (EN) NT28318 - Biological contamination (EN)
Narrower terms:	
<ul style="list-style-type: none"> - Contamination - Deposition - Emission 	

<ul style="list-style-type: none"> - Environmental pollution - Immission - Pollution abatement - Pollution type - Source of pollution <p>Scope note: not available</p> <p>Groups:</p> <p>WASTES, POLLUTANTS, POLLUTION</p> <p>Themes:</p> <p>pollution</p> <p>Other relations:</p> <p>Has exact match:</p> <ul style="list-style-type: none"> - AGROVOC: Pollution - EuroVoc: pollution <p>Wikipedia article: Pollution</p>	<p>NT10962 - Food contamination (EN)</p> <p>RT6076 - pollutants (EN)</p> <p>RT3739 - hygiene (EN)</p> <p>RT33995 - Food safety (EN)</p> <p>RT3019 – Food hygiene (EN)</p> <p>RT28320 - Foreign bodies (EN)</p> <p>RT138 - Adulteration (EN)</p> <p>RT10350 - Deterioration (EN)</p>
--	---

- **Avaluació:** Tal i com mostren la jerarquia d'ambdós tesaurus, 'Contaminació' és un terme específic de 'Pol·lució', per tant, el tipus de relació de l'alineament és de tipus *Narrower*.

12.4. **Alineament 16:** Slaughtering of animals *versus* Slaughtering

<p>Slaughtering of animals. Url: http://www.eionet.europa.eu/gemet/concept?cp=7726&langcode=en&ns=1</p>	<p>Slaughtering. Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=7107&mylang_interface=en&myLanguage=EN</p>
<p>Definition: Killing of animals for food. (Source: PHC)</p> <p>Broader terms: agriculture and cattle industry</p> <p>Related terms:</p> <ul style="list-style-type: none"> - Animal for slaughter - Livestock breeding - Slaughterhouse <p>Scope note: not available</p> <p>Groups:</p>	<p>UF 11976 - Killing of livestock (EN)</p> <p>UF 9507 - Butchering (EN)</p> <p>BT 6195 - processing (EN)</p> <p>NT 29002 - Bleeding (slaughtering) (EN)</p> <p>RT 2744 - Evisceration (EN)</p> <p>RT 28500 - Plucking (EN)</p> <p>RT 4675 – Meat industry (EN)</p>

<p>INDUSTRY, CRAFTS; TECHNOLOGY; EQUIPMENTS</p> <p>Themes:</p> <p>animal husbandry, food, drinking water industry</p> <p>Other relations:</p> <p>Has exact match:</p> <p>- EuroVoc: slaughter of animals</p>	<p>RT 7475 - stunning (EN)</p>
--	--

- **Avaluació:** Ambdós termes fan referència a 'sacrifici d'animals', l'únic que AGROVOC no ho especifica així, si que a través de la jerarquia s'observa que els dos conceptes tenen el mateix significat, per tant el tipus de relació de l'alineament és de tipus "exact match".

2.17 Alineament 17: Poriferan *versus* Sponges

<p>Poriferan (GEMET). Url: http://www.eionet.europa.eu/gemet/concept?cp=8006&langcode=en&ns=1</p>	<p>Sponges (AGROVOC). Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=7321&mylang_interface=en&myLanguage=EN</p>
<p>Definition: The sponges, a phylum of the animal kingdom characterized by the presence of canal systems and chambers through which water is drawn in and released; tissues and organs are absent. (Source: MGH)</p> <p>Broader terms: invertebrate</p> <p>Scope note: not available</p> <p>Groups:</p> <p>BIOSPHERE (organisms, ecosystems)</p> <p>Themes:</p> <p>biology</p> <p>Other relations:</p> <p>Has close match: UMTHES: Schwämme</p>	<p>BT 2941 - Fishery products (EN)</p> <p>BT 430 - Animal fibres (EN)</p> <p>RT 7320 - Sponge fisheries (EN)</p> <p>RT 7322 - Spongia (EN)</p>

Wikipedia article: Sponge	
---	--

- **Avaluació:** El tipus de relació entre els termes és d'exact match". Ja que la definició d'ambdós és esponges (GEMET usa el singular i AGROVOC el plural), a més la traducció del terme amb espanyol és la mateixa.

2.18 Alineament 18: Tanker (ship) versus Tanker ships

Tanker (ship). Url: http://www.eionet.europa.eu/gemet/concept?cp=8285&langcode=en&ns=1	Tanker ships. Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=7603&mylang_interface=en&myLanguage=EN
<p>Broader terms: Ship</p> <p>Narrower terms:</p> <ul style="list-style-type: none"> - Oil tanker - Tween-deck tanker <p>Scope note:</p> <p>scope note is not available</p> <p>Groups:</p> <p>TRAFFIC, TRANSPORTATION</p> <p>Themes:</p> <p>transport</p> <p>water</p> <p>Other relations:</p> <p>Has close match: UMTHES: Tankschiff</p> <p>Wikipedia article: Tanker (ship)</p>	<p>UF 12914 – oil tankers (EN)</p> <p>BT 7040 - Ships (EN)</p>

- **Avaluació:** Ambdós termes tenen una relació d'equivalència exacta o "exact match". Tant en la seva definició com jerarquia superior (BT) coincideixen.

2.19 Alineament 19: Biosphere reserve versus Nature reserves

Biosphere reserve (GEMET). Url: http://www.eionet.europa.eu/gemet/c	Nature reserves (AGROVOC). Url: http://aims.fao.org/en/agrovoc-term-
--	--

concept?cp=894&langcode=en&ns=1	info?mytermcode=16141&mylang_interface=en&myLanguage=EN
<p>Protected land and coastal areas that are approved under the Man and Biosphere programme (MAB) in conjunction with the Convention on International Trade in Endangered Species (CITES). Each reserve has to have an ecosystem that is recognized for its diversity and usefulness as a conservation unit. The reserves have at least one core area where there can be no interference with the natural ecosystem. A transition zone surrounds this and within it scientific research is allowed. Beyond this is a buffer zone which protects the whole reserve from agricultural, industrial and urban development. Biosphere reserves and buffer zones were regarded as examples of a new generation of conservation techniques. (Source: WRIGHT)</p> <p>Broader terms: protected area</p> <p>Related terms: biological reserve</p> <p>Scope note: not available</p> <p>Groups:</p> <p>LAND (landscape, geography)</p> <p>Themes:</p> <p>Biology, environmental policy, natural areas, landscape, ecosystems</p> <p>Other relations:</p> <p>Has close match: UMTHES: Biosphärenreservat</p>	<p>UF 28016 - Parks (reserves) (EN)</p> <p>UF 37232 - Wildlife parks (EN)</p> <p>UF 37213 – Biosphere reserves (EN)</p> <p>BT 37952 - protected areas (EN)</p> <p>NT 36402 - Marine parks (EN)</p> <p>NT 5079 - National parks (EN)</p> <p>NT 3184 - Game reserves (EN)</p> <p>RT 28126 - Protected forests (EN)</p> <p>RT 37967 – Natural heritage (EN)</p> <p>RT 4182 - land use (EN)</p> <p>RT 5091 - Natural resources (EN)</p> <p>RT 5092 - Nature conservation (EN)</p> <p>RT 8257 - Virgin land (EN)</p> <p>RT 8389 - Wildlife (EN)</p>

Wikipedia article: [Man and the Biosphere Programme](#)

- **Avaluació:** Estem al davant de dos termes que són equivalents, o sigui, que tenen una relació de tipus “exact match”. Tal i com apareix en la jerarquia del terme d'AGROVOC, aquest considera que “Biosphere reserve” és un terme “no preferit” o el que és el mateix una altra forma d'anomenar “Nature reserves”, que en canvi en GEMET és el seu terme preferit. A més a més, ambdós tesaurus tenen el mateix punt de vista en aquest concepte ja que el terme superior o BT, és el mateix (*Protected areas*)

2.20. Alineament 20: Wood versus forests

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'wood' is shown with its hierarchy: LAND (landscape, geography) > land > natural area > terrestrial area > forest > wood. The wood hierarchy includes: Mediterranean wood, coniferous wood, deciduous wood, mixed woodland, and temperate woodland. On the right, the AGROVOC concept 'Forests' is shown with its hierarchy: Land cover > Vegetation > Forests. The Forests hierarchy includes: Broadleaved forests, Coniferous forests, Microphyllous forests, Mixed forests, Rain forests, Secondary forests, Temperate forests, Tropical forests, and Virgin forests. A comparison table is visible, showing 'Exact match' and 'Broader' relationships between the two concepts.

- **Avaluació:** Bosc/os (forest) o “Forests”, tal i com es pot veure en la jerarquia de GEMET, té un tipus de relació genèrica (BT), sobre el terme wood, ja que interpreta “wood” com a una densa vegetació d’arbres, però que no arriben a formar un bosc. En canvi, si es cerca el terme “wood” a AGROVOC, el concepte representa a un tipus de material (la fusta) o producte que prové dels boscos, i és per això que el seu terme superior és “Forest products”.

Stratum 3: Mostra aleatòria de 20 alineacions

3.1 Alineament 1: Biological cycle versus Life cycle

The screenshot shows the Amalgame alignment evaluator interface. On the left, a concept box for 'biological cycle' is displayed, including its definition and a hierarchical tree structure: NATURAL ENVIRONMENT, ANTHROPIC ENVIRONMENT > BIOSPHERE (organisms, ecosystems) > biosphere > biological process > life cycle > biological cycle. On the right, a target concept box for 'Life cycle, Life history' is shown, with a 'target concept 1 (was: closeMatch?) out of 1 mappings:' section. Below this, there are buttons for 'Exact match (=)', 'close match (c)', and 'not related (-)'. The 'close match (c)' button is highlighted. Other buttons include 'Broader (b)', 'narrower (n)', 'related match (r)', and '? I'm not sure (u)?'. A 'comment (optional):' field is also present. At the bottom, there is a 'reset all(!)' button and the version number 'ClíoPatría (version epoc)'. The top navigation bar includes 'Places', 'Admin', 'Repository', 'Vocabularies', 'Alignments', 'Query', 'Help', and 'Administrator'.

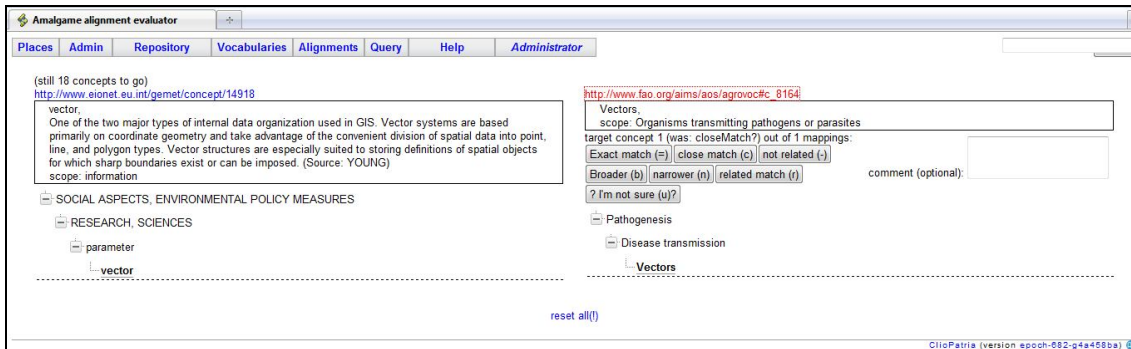
- **Avaluació:** Ambdós termes estan relacionats però no tenen un tipus d'equivalència exacta. Tal i com exposa GEMET, "Biological cycle" és un terme específic de "life cycle", és a dir, aquest darrer té una relació de tipus "broader". Però tot i que GEMET contempla aquesta diferència i relaciona els dos termes, AGROVOC no ho fa, i considera que "biological cycle" és un terme equivalent però no acceptat (USE FOR) com a terme d'indexació en el tesaurus. Així doncs s'ha d'establir que el concepte que apareix en AGROVOC, "life cycle", té una relació de quasi sinònim o "close match" amb "biological cycle".

3.2 Alineament 2: Chemical pollution versus Chemical Contamination

The screenshot shows the Amalgame alignment evaluator interface. On the left, a concept box for 'chemical pollution' is displayed, including its definition and a hierarchical tree structure: HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE... > WASTES, POLLUTANTS, POLLUTION > pollution > pollution type > chemical pollution > photochemical pollution. On the right, a target concept box for 'Chemical contamination' is shown, with a 'target concept 1 (was: closeMatch?) out of 1 mappings:' section. Below this, there are buttons for 'Exact match (=)', 'close match (c)', and 'not related (-)'. The 'Exact match (=)' button is highlighted. Other buttons include 'Broader (b)', 'narrower (n)', 'related match (r)', and '? I'm not sure (u)?'. A 'comment (optional):' field is also present. At the bottom, there is a 'reset all(!)' button and the version number 'ClíoPatría (version epoc)'. The top navigation bar includes 'Places', 'Admin', 'Repository', 'Vocabularies', 'Alignments', 'Query', 'Help', and 'Administrator'.

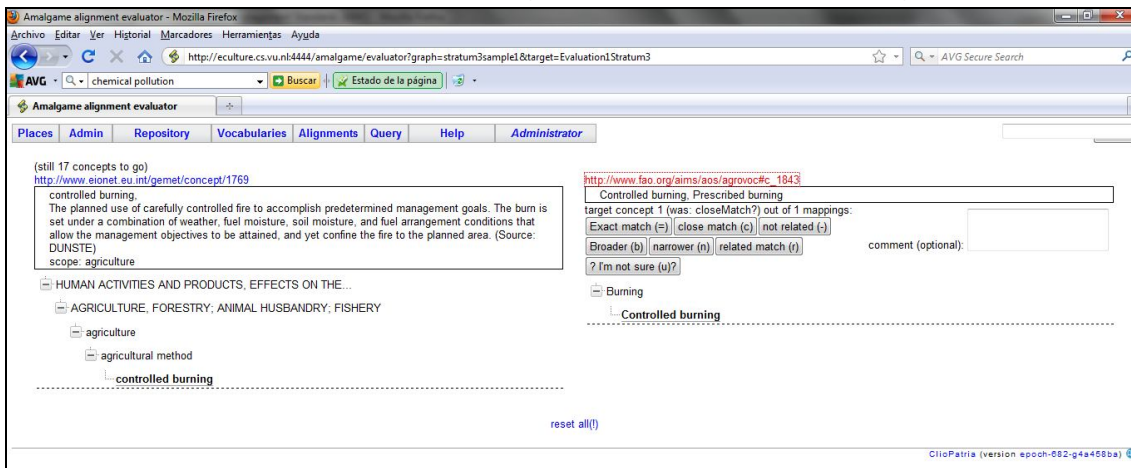
- **Avaluació:** Contaminació i Pol·lució són conceptes que es consideren sinònims. Aquest és un exemple clar, de com l'ús del mapeig multilingüe facilita la identificació i recuperació de termes equivalents o relacions de tipus "Exact Match". Si s'accedeix als tesaurus online, s'observa que la definició i traducció del terme en espanyol és la mateixa.

3.3 Alineament 3: vector versus vectors



- **Avaluació:** Els conceptes semblen idèntics, però en realitat no són termes equivalents perquè tenen un abast i significat diferent. S'estableix doncs una relació de termes "no relacionats". Vector en GEMET, està relacionat amb Sistemes d'Informació Geogràfica (GIS), i el terme Vectors en AGROVOC, està relacionat amb els organismes (vectors com a organismes transmissors de patògens i paràsits)

3.4 Alineament 4: Controlled burning versus Controlled burning



- **Avaluació:** Relació de tipus d'equivalència exacta ("exact match"). Tot i que la jerarquia no és equivalent en ambdós tesaurus, l'enfocament dels conceptes té el mateix abast, l'agricultura, a més de tenir el mateix significat, foc controlat o crema controlada, que és de fet, una pràctica agrícola habitual.

3.5 Alineament 5: Agricultural equipment versus Farm equipment



- **Avaluació:** Relació de tipus "close match" o de quasi sinònims. AGROVOC els considera termes que poden ser emprats com a termes equivalents, ja que considera que "Agricultural equipment" és un no-descriptor de "Farm

equipment". No obstant això, "Farm equipment" podria ser considerat, en un altre tesaurus, com un terme específic (NT) d'"Agricultural equipment". Però en aquest context, on s'avaluen els alineaments dels dos tesaurus citats, és consideren termes quasi sinònims.

3.6 Alineament 6: finances versus finance

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'finances' is displayed with its definition: 'The monetary resources or revenue of a government, company, organization or individual. (Source: RHW) scope: economics'. Below the definition is a hierarchical tree structure under 'SOCIAL ASPECTS, ENVIRONMENTAL, POLICY MEASURES' and 'ECONOMICS, FINANCE', listing various sub-concepts like 'budget', 'cost', 'credit', 'financial fund', etc. On the right, the AGROVOC concept 'Finance' is displayed with its definition: 'target concept 1 (was: closeMatch?) out of 1 mappings: Exact match (=) close match (c) not related (-)'. Below the definition is a hierarchical tree structure under 'Finance', listing sub-concepts like 'Accounting', 'Budgets', 'Capital', 'Cooperative finance', etc. The interface includes a navigation menu at the top with options like 'Places', 'Admin', 'Repository', 'Vocabularies', 'Alignments', 'Query', 'Help', and 'Administrator'.

- **Avaluació:** Clar exemple de tipus de relació "exact match" o d'equivalència exacta entre els termes, ja que tant en la jerarquia com en la definició dels termes en ambdós tesaurus utilitzen el descriptor dintre el mateix abast econòmic. No obstant aquest és un exemple també de diferències de cobertura temàtica i especificitat, ja que GEMET desplega un major nombre de termes específics, de fet, si s'observa la jerarquia dels tesaurus només coincideixen en 3 conceptes. També, de forma excepcional, GEMET empra el plural del concepte i AGROVOC el singular, quan normalment és a la inversa.

3.7 Alineament 7: Food colourant versus Food colourants

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'food colourant' is displayed with its definition: 'Any digestible substance, usually a synthetic dye, which manufacturers add to food to give it color and enhance its appearance. (Source: INP) scope: human health'. Below the definition is a hierarchical tree structure under 'HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE...' and 'PRODUCTS, MATERIALS', listing sub-concepts like 'product', 'chemical product', 'dye', and 'food colourant'. On the right, the AGROVOC concept 'Food colourants' is displayed with its definition: 'target concept 1 (was: closeMatch?) out of 1 mappings: Exact match (=) close match (c) not related (-)'. Below the definition is a hierarchical tree structure under 'Additives' and 'Food additives', listing sub-concepts like 'Erythrosine' and 'Tartrazine'. The interface includes a navigation menu at the top with options like 'Places', 'Admin', 'Repository', 'Vocabularies', 'Alignments', 'Query', 'Help', and 'Administrator'.

- **Avaluació:** els termes tenen una relació d'equivalència exacta o "Exact match". Hi ha però petites diferències en cada tesaurus, ús de singular del

terme per part de GEMET i ús de plural en AGROVOC. Tot i que ambdós termes tenen el mateix significat, a través de la jerarquia s'observen diferències amb l'abast dels conceptes. GEMET engloba el terme coma una substància química, un tipus de tint alimentari que aporta color entre d'altres propietats als aliments. En canvi, AGROVOC inclou el concepte dintre un tipus d'additiu alimentari. Amb l'elaboració d'un tercer tesaurus, es podrien afegir ambdós punts de vista ja que no són incompatibles (un additiu alimentari potser un tint o colorant alimentari químic, però també hi ha additius naturals).

3.8. Alineament 8: Alloy *versus* Alloys

- **Avaluació:** tipus de relació d'equivalència exacta ("Exact match"). Cada tesaurus té una jerarquia diferent (tant en termes superiors com inferiors), també existeix la diferència de singular i plural, però l'abast i definició d'ambdós conceptes és la mateixa. Ambdós tesaurus interpreten el concepte "alloy/s" com un tipus de material metàl·lic (GEMET parla de substàncies metal·liques, i un material¹²³ potser definit també com una substància).

3.9 Alineament 9: garden waste *versus* Garden wastes

¹²³ Material: En enginyeria, un material és una substància (element o compost químic) amb alguna propietat útil, mecànica, elèctrica, òptica, tèrmica o magnètica. Font: Wikipedia. Url:

<http://es.wikipedia.org/wiki/Material>

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'garden waste' is displayed with its scope and a hierarchical tree structure. On the right, the AGROVOC concept 'Garden wastes, Yard trimmings' is displayed with its scope and a hierarchical tree structure. The interface indicates an 'Exact match' between the two concepts.

- **Avaluació:** Relació d'equivalència exacta ("Exact match"). Ambdós termes tenen el mateix abast temàtic, els residus orgànics, concretament fan referència a residus de jardins com per exemple fulles, gespa, mala herba, etc. També com s'observa en la imatge, ambdós pegen del concepte de Residus, tot i que el seu terme genèric immediatament superior canvia i no fan referència al mateix punt de vista, però això no determina el tipus de relació de l'alineament. No obstant si que mostra una diferència a nivell de cobertura dels tesaurus, de fet GEMET considera que "organic waste" és un terme específic de "compostable waste" (BT de "garden waste"), és a dir el considera més un terme relacionat que no pas un terme superior com si ho fa AGROVOC.

3.10 Alineament 10: water power *versus* water power

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'water power' is displayed with its scope and a hierarchical tree structure. On the right, the AGROVOC concept 'Water power, Hydropower' is displayed with its scope and a hierarchical tree structure. The interface indicates an 'Exact match' between the two concepts.

- **Avaluació:** Alineament amb tipus de relació d'equivalència exacta ("Exact match"). No obstant això, s'observa una diferència amb els termes genèric (BT) que mostra un cop més la diferència de cobertura dels tesaurus. En el cas de GEMET el BT té un abast més ampli "conventional energy", i en canvi AGROVOC presenta un BT més específic "Renewable energy".

3.11 Alineament 11: "Land reclamation" vs "reclamation"

- **Avaluació:** Ambdós termes per separat tenen un dignificat diferent, però dintre dels tesaurus avaluats si són representats amb el mateix sentit. És a dir, *Reclamation* té una relació semàntica d'*exact match*, amb *Land Reclamation*, ja que en AGROVOC el terme “Reclamation” només fa referència a “Land Reclamation”, que l’identifica com a un no-descriptor (USE FOR).

3.12 Alineament 12: mill versus mills

- **Avaluació:** Relació de tipus “exact match”. Mill i Mills tenen la mateixa definició, però GEMET usa el singular del terme i AGROVOC el plural. Si s’observa la jerarquia inferior dels termes, coincideixen en nombre i tipus de descriptors específics (NT).

3.13 Alineament 13: Mycorrhiza versus Mycorrhizae (singular versus plural)

A **mycorrhiza** (Gk.,: *fungus roots*,^[1] pl **mycorrhizae**, **mycorrhizas**) is a **symbiotic** (generally **mutualistic**, but occasionally weakly **pathogenic**) association between a **fungus** and the roots of a vascular **plant**.^[2]

Amalgame alignment evaluator - Mozilla Firefox

http://eculture.cs.vu.nl:4444/amalgame/evaluator?graph=stratum3sample1&target=Evaluation1Stratum3a

AVG - mycorrhiza

Places Admin Repository Vocabularies Alignments Query Help Administrator

(still 8 concepts to go)

<http://www.eionet.eu.int/gemet/concept/5465>

mycorrhiza.
The symbiotic association of the root of a higher plant with a fungus. In an ectotrophic mycorrhiza (e.g. heath, pine trees) the fungal mycelium covers the outside of the roots; in an endotrophic mycorrhiza (e.g. orchids) the fungus grows inside the cells of the root cortex. (Source: ALL)
scope: biology

NATURAL ENVIRONMENT, ANTHROPIC ENVIRONMENT

- BIOSPHERE (organisms, ecosystems)
 - biosphere
 - organism
 - plant (biology)
 - mycorrhiza

http://www.fao.org/aims/aos/agrovoc#c_5023

Mycorrhizae, Ectomycorrhizae

target concept 1 (was: closeMatch?) out of 1 mappings:
Exact match (=) | close match (c) | not related (-)

Broader (b) | narrower (n) | related match (r) | comment (optional):

? I'm not sure (u)?

- Organisms
 - Symbionts
 - Mycorrhizae
 - Vesicular arbuscular mycorrhizae

- **Avaluació:** els termes de l'alineament tenen un tipus de relació d'equivalència exacta o "exact match", ja que fan referència organismes (relació entre un fong i les arrels d'una planta vascular). Però un altre cop, trobem que AGROVOC empra el plural i GEMET el singular. De fet, "Mycorrhiza" a part de "Mycorrhizae" també té un altre plural, "Mycorrhizas".

3.14 Alineament 14: Nitrosamine versus Nitrosamines

Amalgame alignment evaluator - Mozilla Firefox

http://eculture.cs.vu.nl:4444/amalgame/evaluator?graph=stratum3sample1&target=Evaluation1Stratum3a

AVG - Nitroso compounds

Places Admin Repository Vocabularies Alignments Query Help Administrator

(still 7 concepts to go)

<http://www.eionet.eu.int/gemet/concept/5616>

nitrosamine.
Any one of a class of neutral, usually yellow oily compounds containing the divalent group = NNO. (Source: CED)
scope: chemistry

HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE...

- CHEMISTRY, SUBSTANCES, PROCESSES
 - chemical
 - organic substance
 - organonitrogen compound
 - amine
 - nitrosamine

http://www.fao.org/aims/aos/agrovoc#c_12637

Nitrosamines, Diethylnitrosamine

target concept 1 (was: closeMatch?) out of 1 mappings:
Exact match (=) | close match (c) | not related (-)

Broader (b) | narrower (n) | related match (r) | comment (optional):

? I'm not sure (u)?

- Nitroso compounds
 - Nitrosamines

- **Avaluació:** Tipus de relació entre els termes de tipus "Exact match", com s'observa en la imatge superior, la única diferència és l'ús del singular per part de GEMET, i el plural per part d'AGROVOC.

3.15 Alineament 15: ornithology versus ornithology

Amalgame alignment evaluator - Mozilla Firefox

http://eculture.cs.vu.nl:4444/amalgame/evaluator?graph=stratum3sample1&target=Evaluation1Stratum3a

AVG - Nitroso compounds

Places Admin Repository Vocabularies Alignments Query Help Administrator

(still 6 concepts to go)

<http://www.eionet.eu.int/gemet/concept/5942>

ornithology.
The branch of zoology that deals with the study of birds, including their physiology, classification, ecology, and behaviour. (Source: AMHER / CED)
scope: biology

SOCIAL ASPECTS, ENVIRONMENTAL POLICY MEASURES

- RESEARCH, SCIENCES
 - science
 - natural science
 - life science
 - zoology
 - ornithology

http://www.fao.org/aims/aos/agrovoc#c_49861

ornithology.

target concept 1 (was: closeMatch?) out of 1 mappings:
Exact match (=) | close match (c) | not related (-)

Broader (b) | narrower (n) | related match (r) | comment (optional):

? I'm not sure (u)?

- Biology
 - Zoology
 - ornithology

- **Avaluació:** L'alineament mostra una relació d'equivalència exacta entre els termes ("exact match"). En aquest cas és evident que estem al davant del

mateix concepte, per la terminologia i definició, però també es comprova la jerarquia, i ambdós casos el concepte superior o terme genèric (BT), “Zoology”, també coincideix.

3.16 Alineament 16: Percolating water *versus* infiltration water

<p>Percolating water (GEMET). Url: http://www.eionet.europa.eu/gemet/concept?cp=6111&langcode=en&ns=1</p>	<p>Infiltration water (AGROVOC). Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=25321&mylang_interface=en&myLanguage=EN</p>
<p>Definition: Subsurface water that passes, under the force of gravity, through rocks or soil along the line of least resistance. (Source: WWC) Broader terms: waste water Scope note: not available Groups: WASTES, POLLUTANTS, POLLUTION Themes: soil, water Other relations: Has close match: - UMTHES: Sickerwasser</p>	<p>UF 25604 - Gravitational water (EN) UF 33446 - Seepage water (EN) BT 7205 – Soil water (EN) RT 26789 - Infiltration (EN) RT 3391 - groundwater (EN) RT 8318 – Water erosion (EN)</p>

- **Avaluació:** Aquest alineament presenta diferents qüestions a avaluar i s’ha identificat com un alineament dubtós o opció “I’m not sure”. Aquest és un cas que s’hauria de consensuar amb d’altres avaluadors i/o depenent de la col·lecció o la institució. D’una banda, la traducció d’ambdós termes podria portar a confusions, però la nota d’abast que proporciona GEMET, clarifica que els dos conceptes són diferents. D’una banda “infiltration water” es refereix a el moviment de l’aigua de pluja o rec en el sòl, mentre que “percolating water” es refereix al moviment descendent de l’aigua dintre de la terra, i en aquest cas segons GEMET, fa referència a aigües residuals. Es planteja que poden ser termes relacionats.

3.17 Alineament 17: Reflectometry *versus* Reflectometry

The screenshot shows the Amalgame alignment evaluator interface. On the left, the concept 'reflectometry' is defined as 'The study of the reflectance of light or other radiant energy. (Source: WEBSTE) scope: physics'. On the right, the concept 'Reflectometry, Tdr (time domain reflectometry)' is shown with a target concept 1 (was: closeMatch?) out of 1 mappings. The interface includes a navigation menu with 'Places', 'Admin', 'Repository', 'Vocabularies', 'Alignments', 'Query', 'Help', and 'Administrator'. Below the navigation menu, there are search bars and a comparison view with options for 'Exact match (=)', 'close match (c)', 'not related (-)', 'Broader (b)', 'narrower (n)', and 'related match (r)'. There is also a 'comment (optional):' field and a 'Methods' section with 'Analytical methods'.

- **Avaluació:** Alineament amb tipus de relació d’equivalència exacta o “Exact match”. L’abast temàtic està també en la mateixa línia, ja que el seus termes superiors tot i no ser idèntic són termes relacionats, “laboratory

tecnica i “analytical methods” (una tècnica de laboratori potser un mètode analític), i ambdós s’engloben en un concepte superior que fa referència a que “reflectometry” és un tipus de mètode o metodologia.

3.18 Alineament 18: Close match: for the scope but they are synonyms

The screenshot shows the Amalgame alignment evaluator interface. On the left, a concept from GEMET is displayed: 'seed (biology)' with the definition 'A mature fertilized plant ovule, consisting of an embryo and its food store surrounded by a protective seed coat (testa). (Source: CED)'. Its scope is 'biology'. The hierarchy includes 'NATURAL ENVIRONMENT, ANTHROPIC ENVIRONMENT' > 'BIOSPHERE (organisms, ecosystems)' > 'biosphere' > 'biological process' > 'embryogenesis' > 'seed (biology)'. On the right, a concept from AGROVOC is displayed: 'Seeds, Caryopses' with the definition 'scope: Botanically, for the propagation material use Seed (6927)'. The target concept is 'Seeds, Caryopses'. The interface shows a 'close match (c)' between the two concepts. Below the match, there are buttons for 'Exact match (e)', 'Broader (b)', 'narrower (n)', 'related match (r)', and '? I'm not sure (u)?'. A comment field is also present.

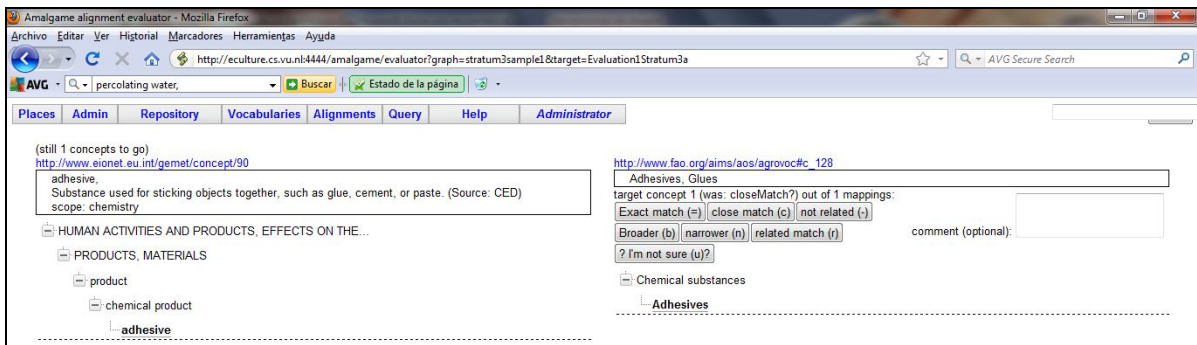
- **Avaluació:** Relació de tipus quasi sinònims o “close match”. Ambdós termes són sinònims, però GEMET li dóna un punt de vista diferent tal i com s’observa en la nota d’abast i en la jerarquia superior dels termes. Diguem que AGROVOC contempla només el concepte aplicat a organismes vegetals i GEMET o fa extensible a tot procés biològic (animal, vegetal o pluricel·lular).

3.19 Alineament 19: shipbuilding versus Shipbuilding

The screenshot shows the Amalgame alignment evaluator interface. On the left, a concept from GEMET is displayed: 'shipbuilding' with the definition 'The art or business of designing and constructing ships. (Source: AMHER)'. Its scope is 'water'. The hierarchy includes 'HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE...' > 'INDUSTRY, CRAFTS; TECHNOLOGY; EQUIPMENTS' > 'industry' > 'vehicle manufacturing industry' > 'shipbuilding'. On the right, a concept from AGROVOC is displayed: 'Shipbuilding, Naval engineering'. The target concept is 'Shipbuilding'. The interface shows an 'exact match (e)' between the two concepts. Below the match, there are buttons for 'Broader (b)', 'narrower (n)', 'related match (r)', and '? I'm not sure (u)?'. A comment field is also present.

- **Avaluació:** Alineament de tipus de relació “exact match” o equivalència exacta. Ambdós termes fan referència al concepte de “construcció naval” o acció de construir un vaixell. Tot i que no comparteixen el mateix terme genèric immediatament superior (BT), si que ambdós pengen de la mateixa disciplina, la enginyeria. No hi ha confusió doncs del tipus d’equivalència.

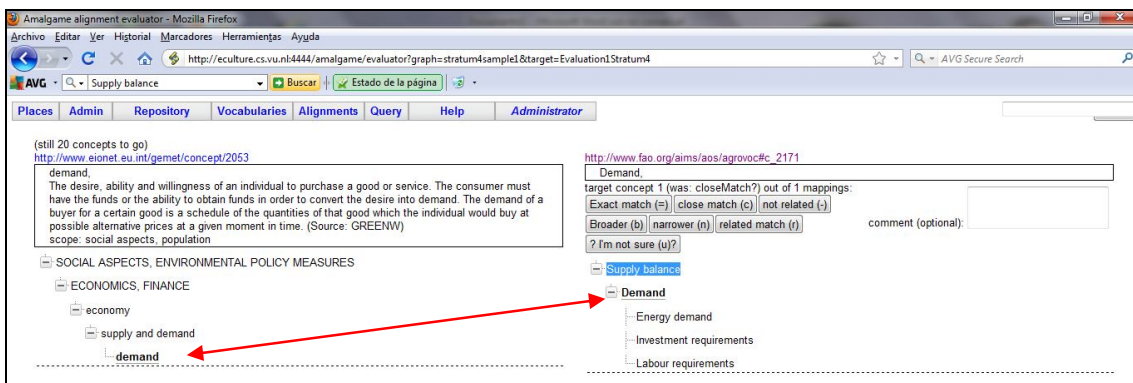
3.20 Alineament 20: adhesive versus Adhesives



- **Avaluació:** Alineament de tipus de relació “exact match” o equivalència exacta. Ambdós conceptes són sinònims, tot i que els termes són descrits amb singular en el tesaurus GEMET i s'utilitza el plural en AGROVOC. Els seus termes superiors o BT (*chemical product versus chemical substances*), indiquen també que l'abast del concepte coincideix en els dos tesaurus.

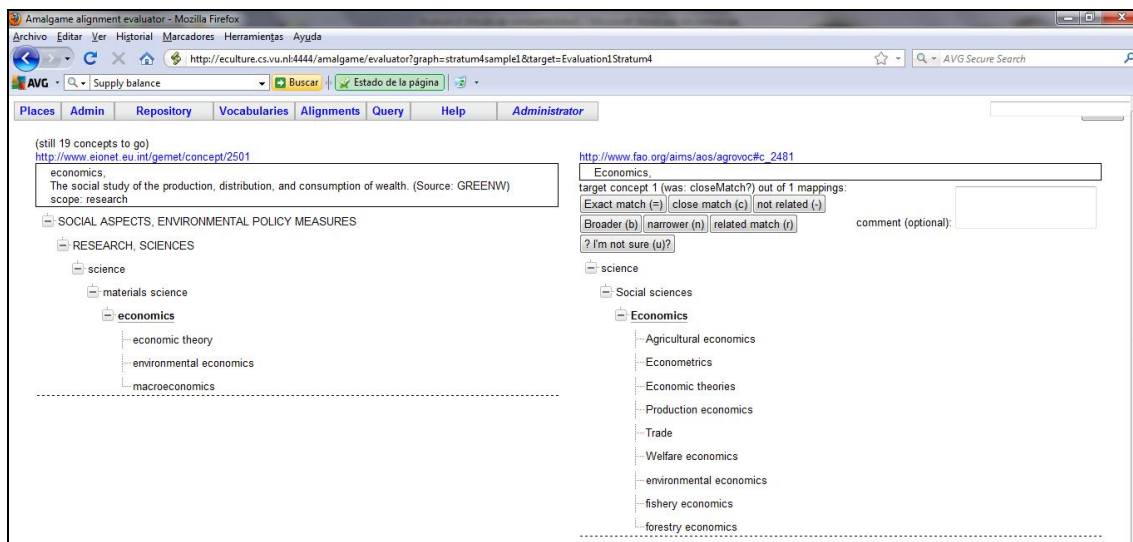
Stratum 4: Mostra aleatòria de 20 alineacions

4.1 Alineament 1: demand versus demand



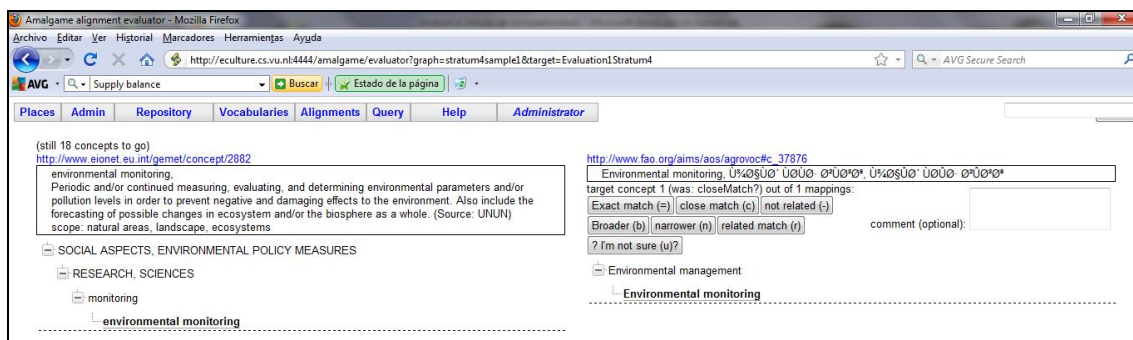
- **Avaluació:** Alineament de tipus de relació “exact match” o equivalència exacta. Ambdós conceptes són sinònims. No obstant els seus termes superiors varien però fan referència al mateix concepte, “Supply and Demand” versus “Supply balance”. Una altra diferència és a nivell d'estructura i especificitat, AGROVOC en aquest cas ofereix una major especificitat ja que inclou 3 termes específics, i GEMET en canvi no inclou cap terme específic.

4.2 Alineament 2: economics versus economics



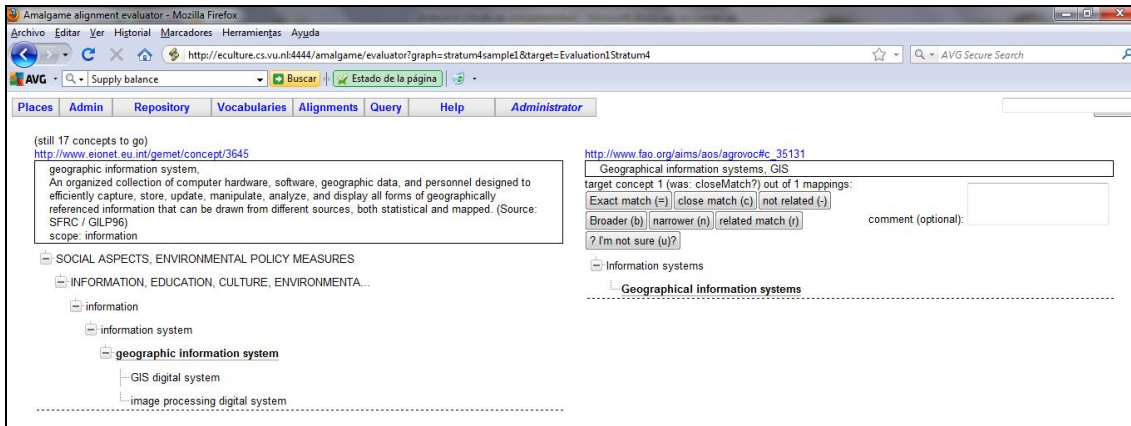
- **Avaluació:** Alineament de tipus de relació “exact match” o equivalència exacta. Ambdós conceptes són sinònims, no obstant, els seus termes superiors (BT) divergeixen tant en terminologia com en l’abast, “material science” *versus* “social science”. En el cas de GEMET, és evident que el terme “economics” fa referència a l’especialització temàtica del tesaurus i per això fa referència a economia aplicada a les ciències ambientals, però segons la nota d’abast el concepte també inclou economia des del punt de vista social.

4.3 Alineament 3: environmental monitoring *versus* environmental monitoring



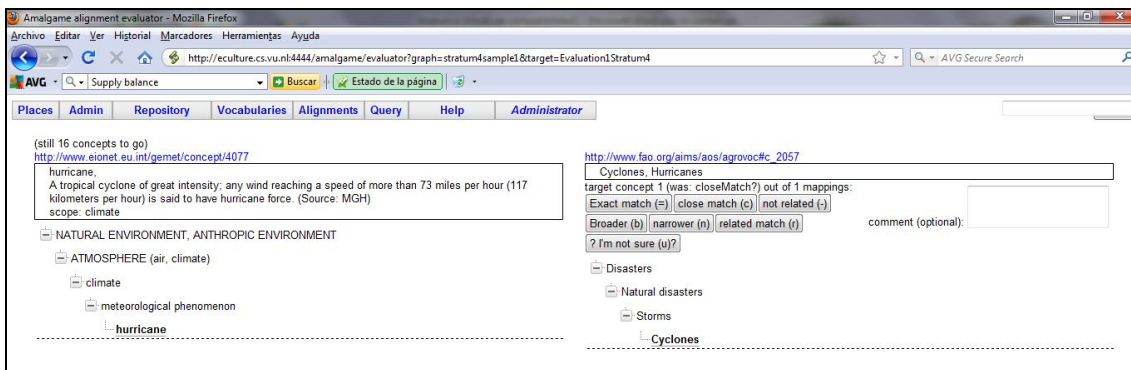
- **Avaluació:** Alineament de tipus de relació “exact match” o equivalència exacta. Ambdós conceptes són sinònims, tot i que els seus termes superiors varien, “Monitoring” *versus* “environmental management”. Independentment de les diferències dels seus termes superiors o BT, ambdós conceptes tenen el mateix punt de vista de gestió mediambiental a través de sistemes o mesures de monitoreig.

4.4 Alineament 4: geographical information systems *versus* geographical information systems



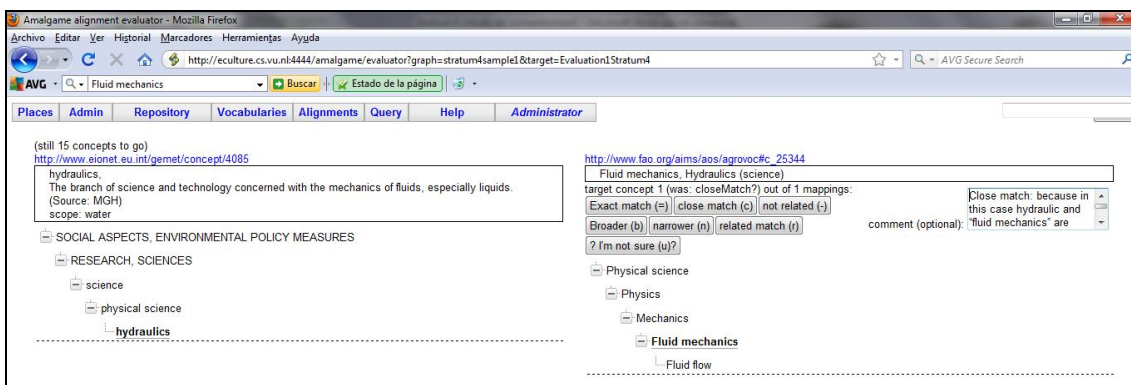
- **Avaluació:** Alineament de tipus de relació “exact match” o equivalència exacta. De fet, es recupera el mateix concepte, amb el mateix punt de vista i a més a més, comparteixen el mateix terme superior, “information systems”.

4.5 Alineament 5: hurricane versus Cyclones



- **Avaluació:** relació d'equivalència de tipus “close match”. Un huracà és un tipus de cicló, per tant, es podrien categoritzar com un tipus de relació *Broader (Cyclone)*, però en aquest cas, es defineix una relació de “quasi-sinònims” perquè el propi tesaurus AGROVOC considera ambdós termes com a sinònims.

4.6 Alineament 6: hydraulics versus fluid mechanics



- **Avaluació:** relació d'equivalència de tipus “close match”. En el tesaurus AGROVOC considera ambdós termes com a sinònims, “Hydraulics” és un concepte “emprat per” designar “Fluid mechanics”. Però a nivell de definició estricta i en un altre context, es podrien categoritzar com un tipus de relació *Broader (Fluid Mechanics)*.

4.7 Alineament 7: inland fisheries versus inland fisheries

The screenshot shows the Amalgame alignment evaluator interface. On the left, the source concept is 'inland fishery' from GEMET, with a description: 'Fishing grounds located in lakes, streams, etc. (Source: RRDA)'. Its hierarchy includes 'HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE...', 'AGRICULTURE, FORESTRY, ANIMAL HUSBANDRY, FISHERY', and 'fishery'. On the right, the target concept is 'Inland fisheries' from AGROVOC, with a description: 'Inland fisheries, Freshwater fisheries'. Its hierarchy includes 'Fisheries', 'Inland fisheries', 'Lagoon fisheries', 'Lake fisheries', and 'River fisheries'. The relationship is identified as 'close match (c)'.

- **Avaluació:** alineament amb tipus de relació “exact match” o equivalència exacta. Ambdós conceptes són sinònims, tot i que els termes són descrits amb singular en el tesaurus GEMET i amb plural en AGROVOC, al igual que els seus termes superiors o BT. Cal destacar també que AGROVOC en aquest cas té una major cobertura i especificitat, ja que incorpora descriptors específics (*Narrower Terms*) i GEMET no.

4.8 Alineament 8: liquid waste versus liquid wastes

The screenshot shows the Amalgame alignment evaluator interface. On the left, the source concept is 'liquid waste' from GEMET, with a description: 'Fluid wastes, consisting of sewage and domestic wastewater, or processed water, or other liquids, produced by industrial activity, particularly by such industries as pulp and paper production, food processing, and the manufacture of chemicals. (Source: ISEF)'. Its hierarchy includes 'HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE...', 'WASTES, POLLUTANTS, POLLUTION', 'waste', 'type of waste', and 'liquid waste'. On the right, the target concept is 'Liquid wastes' from AGROVOC, with a description: 'Liquid wastes'. Its hierarchy includes 'Wastes', 'Liquid wastes', 'Liquid manures', and 'Wastewater'. The relationship is identified as 'close match (c)'.

- **Avaluació:** alineament amb tipus de relació “exact match” o equivalència exacta. La única diferència entre els termes és l'ús del singular en el tesaurus GEMET i ús del plural del terme en AGROVOC, i el mateix passa amb els seus termes superiors o BT. Destacar també que si hi ha diferències a nivell de cobertura o especificitat, AGROVOC inclou dos termes específics (*Liquid manures* i *wastewater*) i GEMET inclou només un terme específic (whey), però no coincident amb la els termes específics d'AGROVOC.

4.9 Alineament 9: marginal land *versus* marginal land

The screenshot shows the Amalgame alignment evaluator interface. The left pane displays the GEMET concept 'marginal land' with its definition and hierarchical structure. The right pane displays the AGROVOC concept 'Marginal land, Terrenos marginales' with its definition and hierarchical structure. A comparison box between them shows 'Exact match (=)' selected, indicating an exact match between the two concepts.

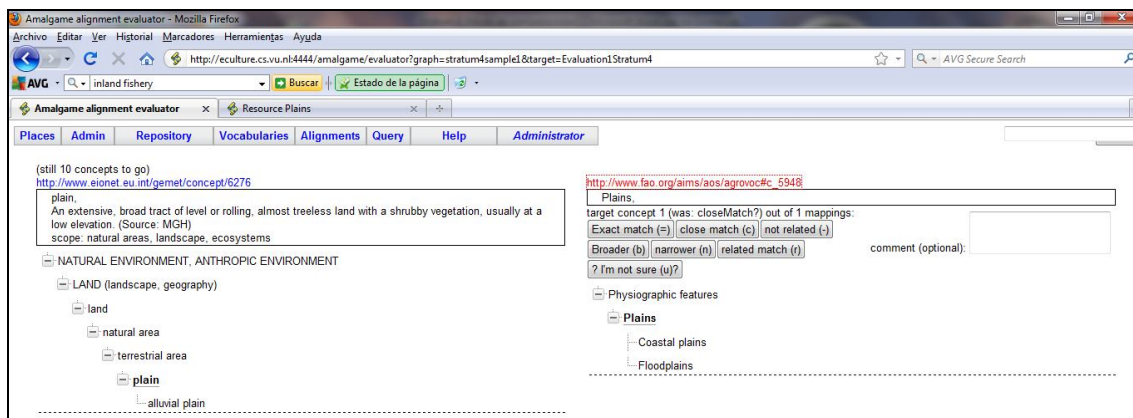
- **Avaluació:** alineament amb tipus de relació “exact match” o equivalència exacta. Ambdós termes tenen termes superiors diferents (*terrestrial area versus land resources*), fet que dóna una connotació diferent a ambdós termes superiors quant al seu abast, en GEMET el concepte té una cobertura més exhaustiva que AGROVOC, “àrea terrestre” *versus* “recursos de la terra”. Aquestes diferències no lleven que els termes alineats són equivalents.

4.10 Alineament 10: *microcomputer versus microcomputers*

The screenshot shows the Amalgame alignment evaluator interface. The left pane displays the GEMET concept 'microcomputer' with its definition and hierarchical structure. The right pane displays the AGROVOC concept 'Microcomputers, Personal computers' with its definition and hierarchical structure. A comparison box between them shows 'Exact match (=)' selected, indicating an exact match between the two concepts.

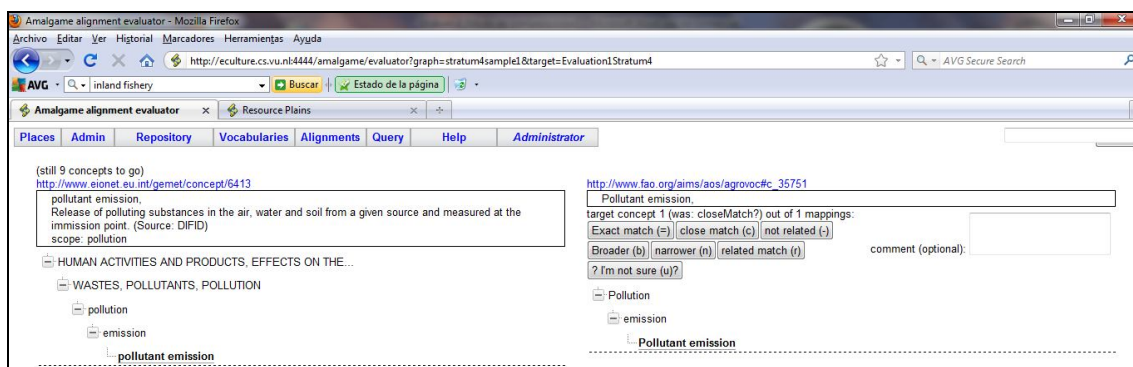
- **Avaluació:** Alineament de tipus de relació “exact match” o equivalència exacta. Ambdós conceptes són sinònims, tot i que els seus termes superiors varien, “Electrotechnical equipment” *versus* “computers”. Independentment de les diferències dels seus termes superiors o BT, ambdós conceptes tenen el mateix punt de vista i per tant, són equivalents.

4.11 Alineament 11: *plain versus plains*



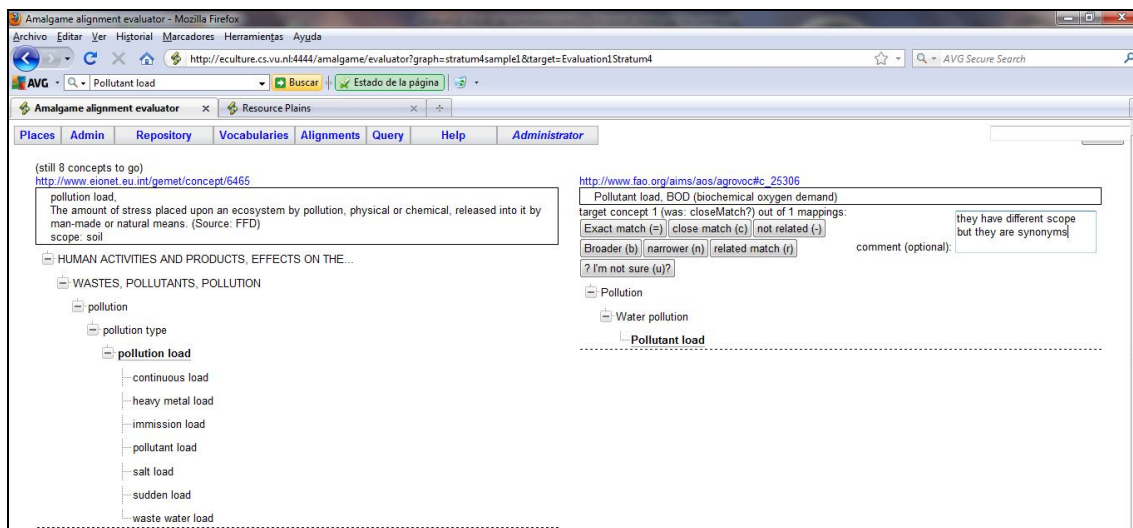
- **Avaluació:** alineament amb tipus de relació “exact match” o equivalència exacta. La única diferència entre els termes és l’ús del singular en el tesaurus GEMET i ús del plural del terme en AGROVOC. Si existeixen diferències a nivell estructural i de cobertura entre els tesaurus, i diferències també amb la designació dels termes superiors (àrea terrestre i característiques fisiogràfiques), però això no canvia el fet de que siguin termes amb equivalència exacta.

4.12 Alineament 12: pollutant emission versus pollutant emission



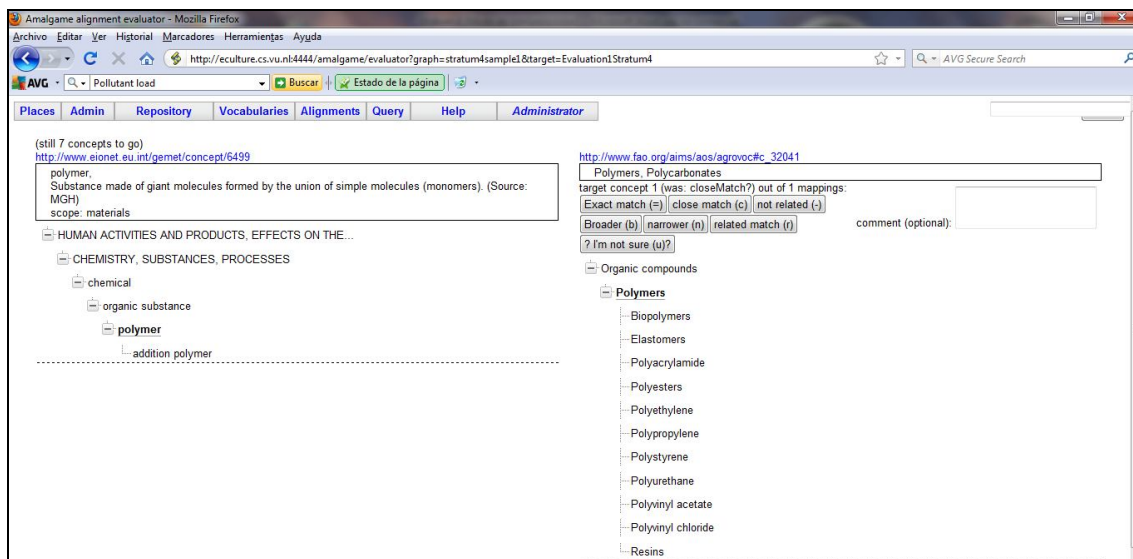
- **Avaluació:** alineament amb tipus de relació “exact match” o equivalència exacta. No existeixen diferències ni respecte a l’ús del plural/singular del concepte, ni tampoc en la seva jerarquia superior. Com s’ha explicat ja des del capítol 3.1, GEMET sempre té una jerarquia més àmplia, ja que a part del “top concept”, classifica els termes en grups i súper grups.

4.13 Alineament 13: pollution load versus pollutant load



- **Avaluació:** termes amb equivalència quasi exacta (relació de “close match”). El terme “pollutant load” (AGROVOC) és més específic que “pollution load” (GEMET), de fet el seu sinònim amb equivalència exacta podria ser “waste water load”, ja que en AGROVOC “pollutant load” fa referència a “water pollution”, tal i com indica el seu terme superior (“Water pollution”). L’abast del concepte és també diferent entre els conceptes superiors dels termes (terra i aigua), no obstant, es consideren com a quasi sinònims per dues raons, primerament la traducció d’ambdós conceptes en algunes llengües com per exemple el castellà, ambdós identifiquen ambdós termes amb el mateix significat ([GEMET](#) i [AGROVOC](#), tradueix el terme de la mateixa forma, “carga contaminante”). L’altre motiu és perquè en AGROVOC no existeix el terme “pollution load”, i per la jerarquia dels termes específics que apareix en GEMET, on es defineixen diferents tipus de càrregues, s’observa que són conceptes específics que també podrien estar associats al concepte “pollutant load” com a terme genèric. Dit d’una altra manera, la informació/documentació relacionada amb “pollution load” es podria indexar també amb “pollutant load”. Tot judici de valor al respecte ha de ser consensuat per la comunitat bibliotecària corresponent que faci ús d’aquests tesaurus i/o la construcció d’un tercer a partir d’aquests. Aquest cas, és també un exemple de que és necessari tenir una col·lecció definida per prendre una decisió més consistent.

4.14 Alineament 14: polymer versus polymers



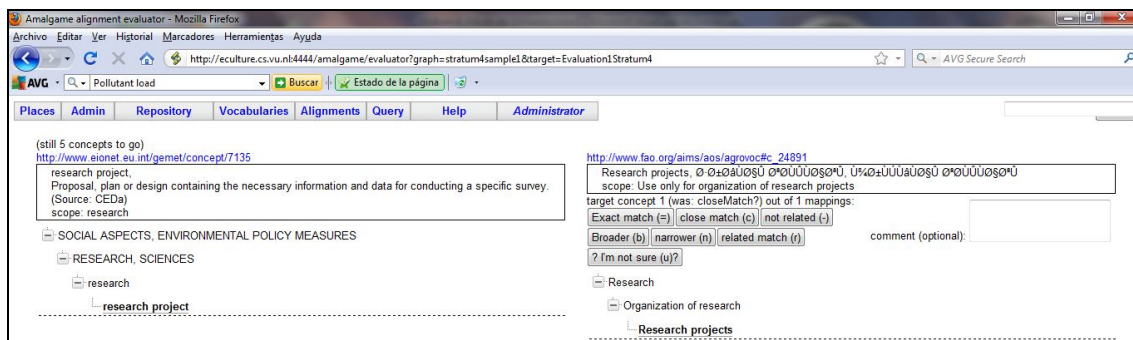
- **Avaluació:** alineament amb tipus de relació “exact match” o equivalència exacta. Diferències respecte l’ús del plural/singular del concepte, però són conceptes que tenen el mateix abast o punt de vista temàtic. Si hi ha diferències terminològiques a nivell de BT, però també continuen tenint el mateix enfocament. La diferència més gran entre els conceptes és la seva jerarquia, ja que està molt més desenvolupada en AGROVOC que en GEMET tal i com s’observa en la imatge superior.

4.15 Alineament 15: avalanche versus avalanches



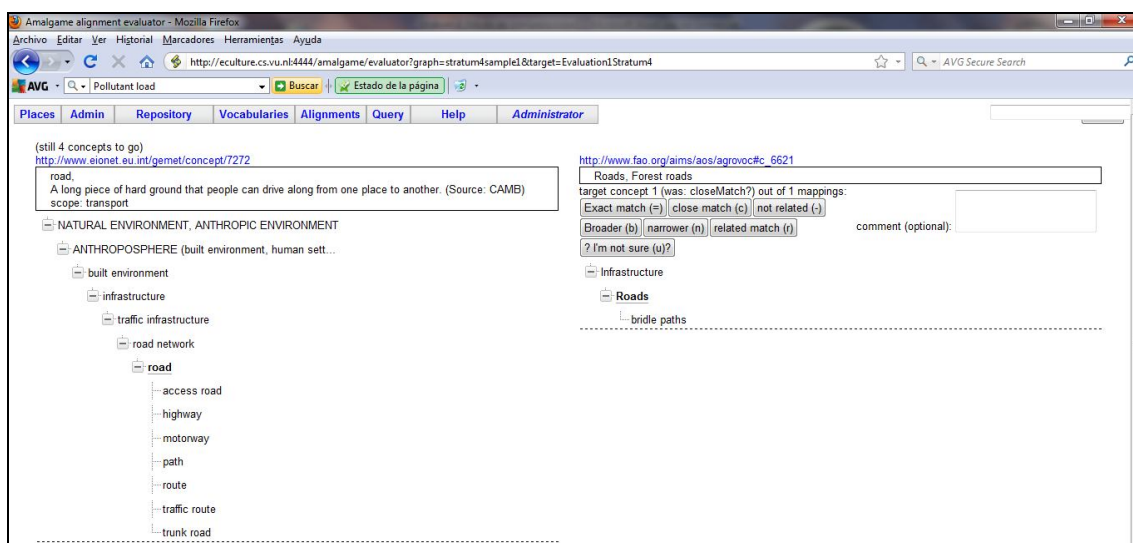
- **Avaluació:** alineament amb tipus de relació “exact match” o equivalència exacta, amb diferència de l’ús del singular (GEMET) i plural (AGROVOC). Com s’observa a nivell dels termes superiors (BT), hi ha diferència també en l’abast temàtic, GEMET engloba el concepte “avalanche” dintre d’un procés geomòrfic de la terra i AGROVOC ho contempla com un desastre natural.

4.16 Alineament 16: research projects versus



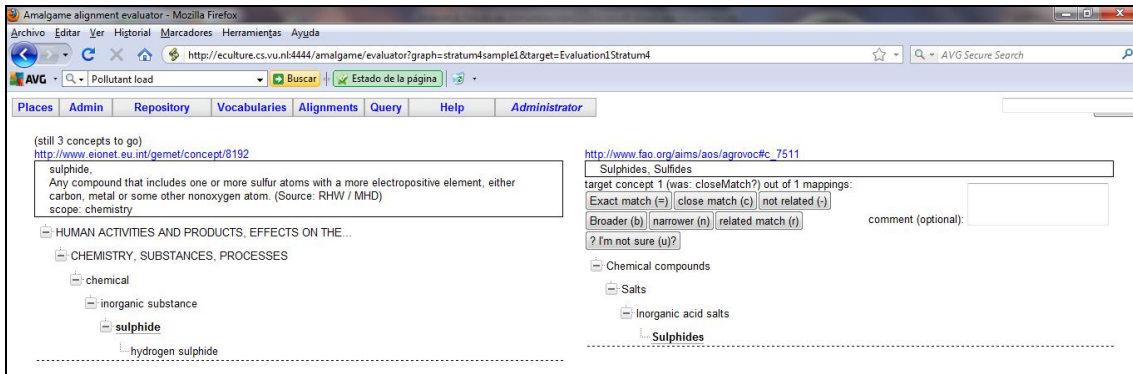
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta, amb diferència de l’ús del singular (GEMET) i plural (AGROVOC). En aquest cas, si es troba diferència estructural quant a termes superiors. AGROVOC ofereix un nivell més de descripció jeràrquic que GEMET, i tot i que ambdós termes tenen el mateix abast temàtic (recerca), el “Broader term” canvia, “research” (GEMET) *versus* “organization of research” que al mateix tems és un NT de “Research” (AGROVOC).

4.17 Alineament 17: road *versus* roads



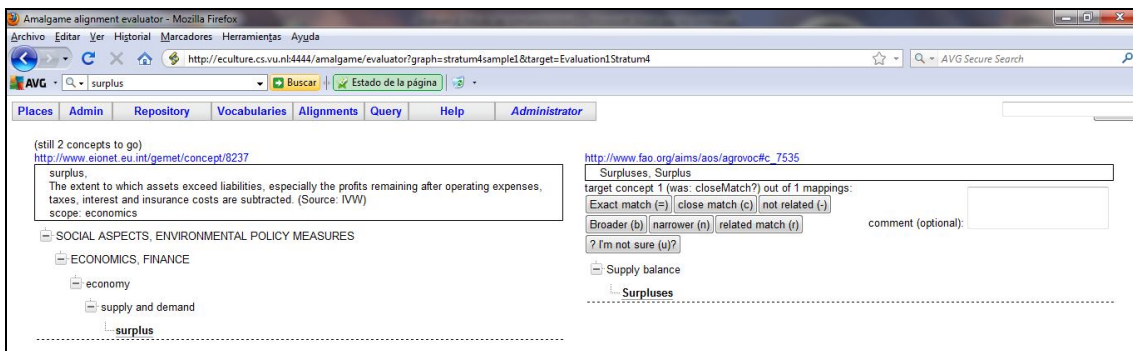
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta, amb diferència de l’ús del singular (GEMET) i plural (AGROVOC). Ambdós conceptes tenen el mateix abast temàtic, perquè es refereixen a “roads” com a un tipus d’infraestructura. No obstant això, la jerarquia de GEMET és completament diferent a la d’AGROVOC i també els seus termes superiors (*Road network versus infrastructure*).

4.18 Alineament 18: sulphide *versus* sulphides



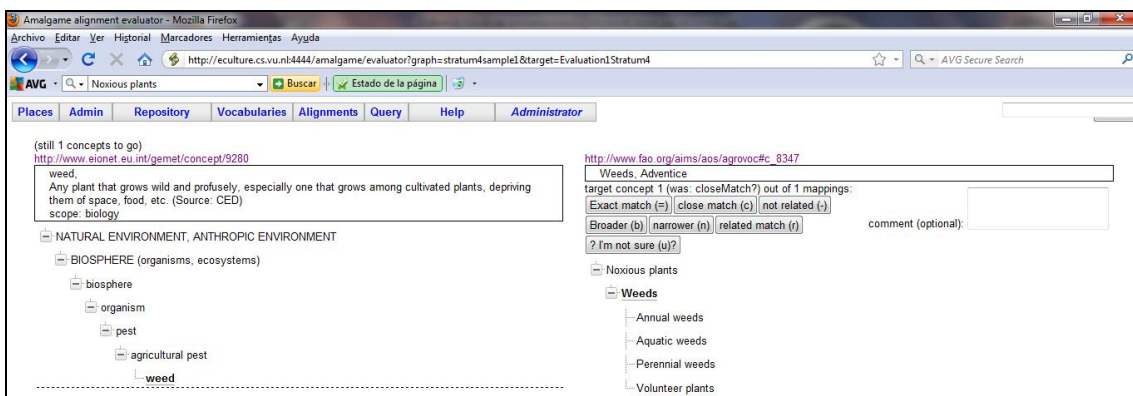
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta, amb diferència de l’ús del singular (GEMET) i plural (AGROVOC). Diferències amb jerarquia i termes superiors però ambdós conceptes fan referència al mateix punt de vista, la sal com a un compost químic.

4.19 Alineament 19: surplus versus surpluses



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta, amb diferència de l’ús del singular (GEMET) i plural (AGROVOC). Ambdós termes tenen el mateix abast (economia). Els termes superiors són diferents però també equivalents i amb el mateix punt de vista.

4.20 Alineament 20: weed versus weeds



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta, amb diferència de l’ús del singular (GEMET) i plural (AGROVOC). Ambdós termes tenen un mateix punt de vista, no però a nivell de termes superiors que si existeix una cobertura diferent (agricultural pest *versus* noxious plants).

Stratum 5: Mostra aleatòria de 20 alineacions

5.1 Alineament 1: arboretum *versus* arboreta

- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta, amb diferència de l’ús del singular (GEMET) i plural (AGROVOC). Ambdós conceptes tenen el mateix punt de vista, tot i que el seu terme superior és diferent (*vegetation versus botanical gardens*).

5.2 Alineament 2: pruning *versus* pruning

- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós conceptes tenen el mateix punt de vista,

tot i que el seu terme superior és diferent (*agricultural method versus cultivation*).

5.3 Alineament 3:

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'vitriification' is displayed with its definition and hierarchical structure. On the right, the AGROVOC concept 'Vitrification' is shown with its definition and hierarchical structure. The alignment is marked as 'not related'.

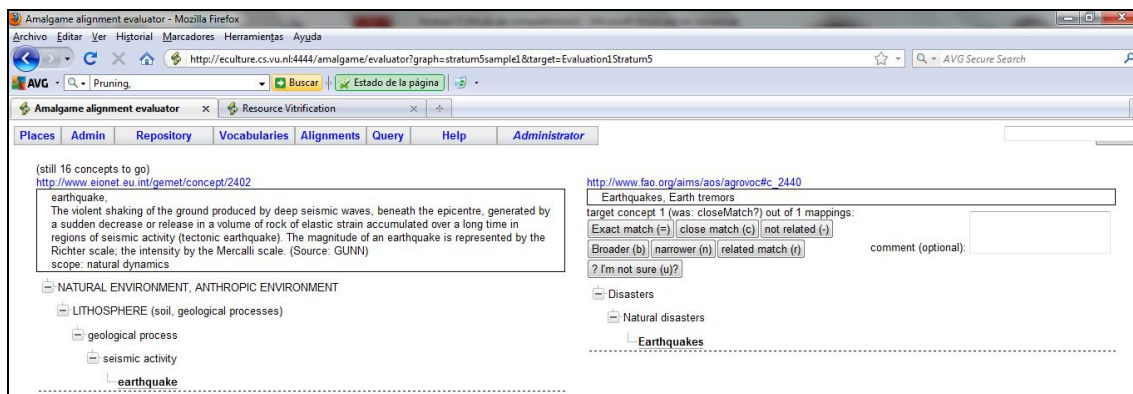
- **Avaluació:** alineament amb tipus de relació “not related”, són termes no equivalents. “Vitrification” en GEMET fa referència a un procés industrial, i en AGROVOC el terme fa referència a malalties de plantes.

5.4 Alineament 4: cultural heritage versus cultural heritage

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'cultural heritage' is displayed with its definition and hierarchical structure. On the right, the AGROVOC concept 'Cultural heritage' is shown with its definition and hierarchical structure. The alignment is marked as 'exact match'.

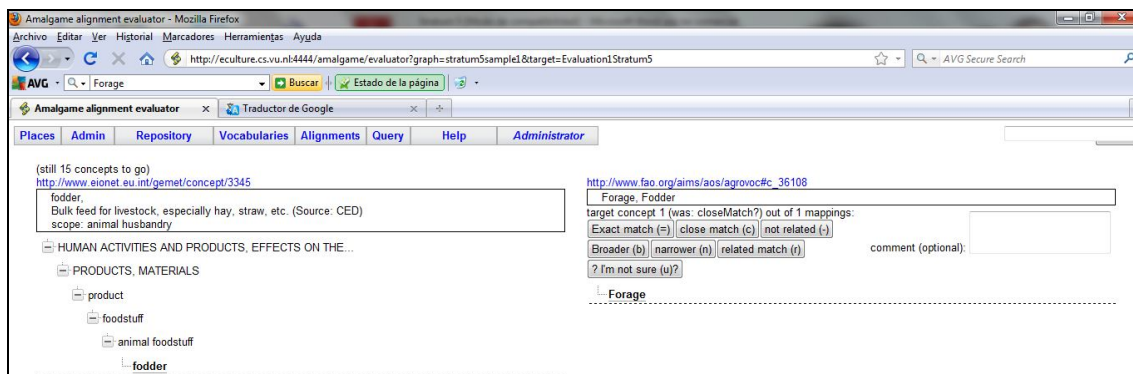
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix punt de vista.

5.5 Alineament 5:



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. GEMET usa el terme en singular i AGROVOC amb plural. Els termes tenen termes superiors diferents, el BT en GEMET és “seismic activity” i en AGROVOC “natural disasters”, com s’observa el punt de vista entre els tesaurus és diferent, GEMET es refereix al concepte “earthquake” com un procés geològic i AGROVOC com un desastre natural. No obstant això, són termes equivalents i s’estableix un tipus de relació d’equivalència exacta.

5.6 Alineament 6: fodder versus forage



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Tot i que són termes que aparentment no pareixen equivalents, són termes sinònims, de fet, AGROVOC contempla el terme “fodder” (GEMET) com un concepte “no acceptat” però que sol ser un terme utilitzat per definir “forage”, els considera termes sinònims.

5.7 Alineament 7: amine versus amines

- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes alineats tenen el mateix punt de vista ja que els dos tesaurus engloben el terme dintre dels “compostos de nitrogen orgànic”. Hi ha dues diferències entre els dos tesaurus, d’una banda l’ús del terme en singular (GEMET) i ús del plural (AGROVOC), i d’altra banda, l’estructura jeràrquica, tant a nivell superior com a nivell inferior del terme, essent AGROVOC el tesaurus amb major termes específics.

5.8 Alineament 8: hybridisation versus hybridization

- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Els dos termes tenen una diferència ortogràfica, però tant per la definició en aquest cas com pels termes superiors dels conceptes, ajuden a identificar que estem al davant de dos termes d’equivalència exacta. Els termes superiors són diferents, “reproduction (biology)” versus “breeding methods”.

5.9 Alineament 9: hydrocarbon versus hydrocarbons

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'hydrocarbon' is selected, with its definition: 'A very large group of chemical compounds composed only of carbon and hydrogen. (Source: MGH)'. The hierarchy shows 'hydrocarbon' under 'organic substance' and 'chemical'. On the right, the AGROVOC concept 'Hydrocarbons, Alkanes' is selected, with its definition: 'Hydrocarbons, Alkanes'. The hierarchy shows 'Hydrocarbons, Alkanes' under 'Hydrocarbons'. The interface includes a search bar, navigation tabs, and a 'comment (optional):' field.

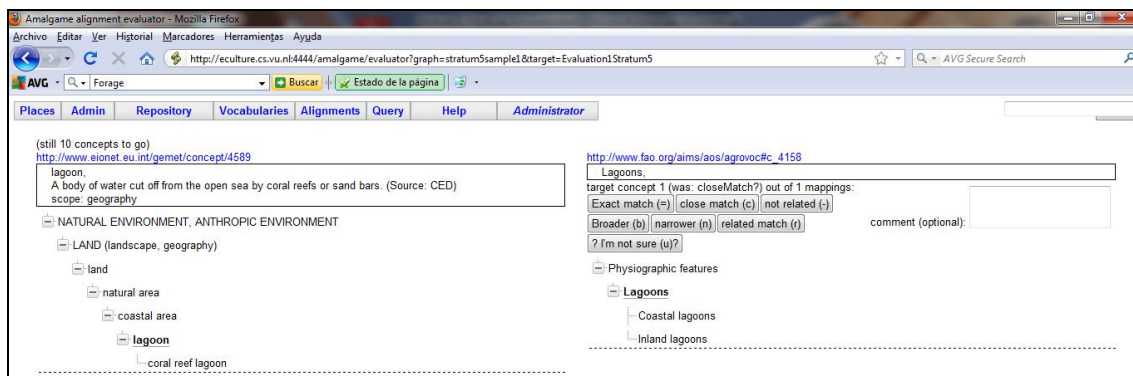
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix punt de vista i els seus termes superiors són considerats sinònims (*organic substance versus organic compounds*). Diferències entre els termes a nivell de jerarquies, tant en termes superiors com en termes específics (AGROVOC inclou molts més termes específics).

5.10 Alineament 10: kerosene versus paraffin

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'kerosene' is selected, with its definition: 'A thin oil distilled from petroleum or shale oil, used as a fuel for heating and cooking, in lamps, and as a denaturant for alcohol. (Source: AMHER)'. The hierarchy shows 'kerosene' under 'motor fuel', 'fuel', and 'product'. On the right, the AGROVOC concept 'Paraffin, Kerosene' is selected, with its definition: 'Paraffin, Kerosene'. The hierarchy shows 'Paraffin, Kerosene' under 'Paraffin'. The interface includes a search bar, navigation tabs, and a 'comment (optional):' field.

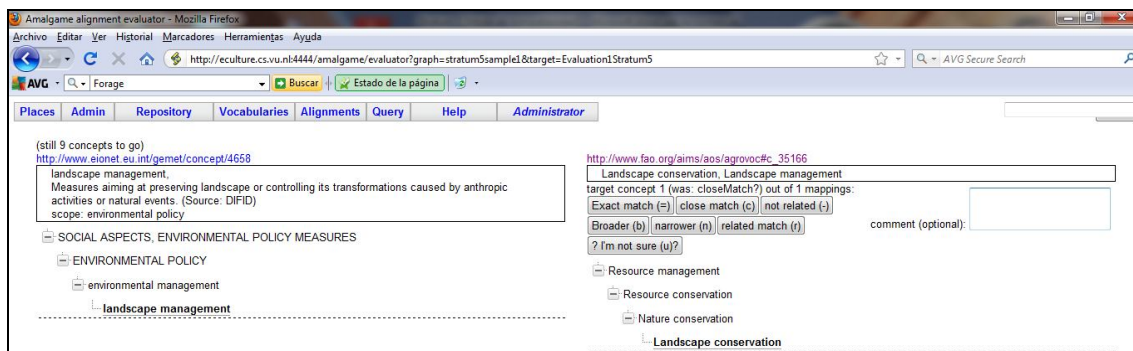
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Com s’observa en la imatge en l’encapçalament del terme apareix al costat i seguit d’una coma (,) el terme “Kerosene”, això indica que és un terme no acceptat però que s’utilitza com a equivalent de “Paraffin”. El punt de vista d’ambdós conceptes dintre del tesaurus és diferent (GEMET: carburants, combustible; AGROVOC ho considera com un component orgànic, químic). Així doncs, aquest fet fa que es puguin establir dos tipus de relacions, “close match” o “exact match”, la decisió dependrà de les polítiques i/o la col·lecció que tingui una institució determinada, en aquest cas el criteri establert és el del propi tesaurus d’AGROVOC, ja que és un dels vocabularis de referència en l’àmbit d’estudi, les ciències aquàtiques i marines.

5.11 Alineament 11: lagoon versus lagoons



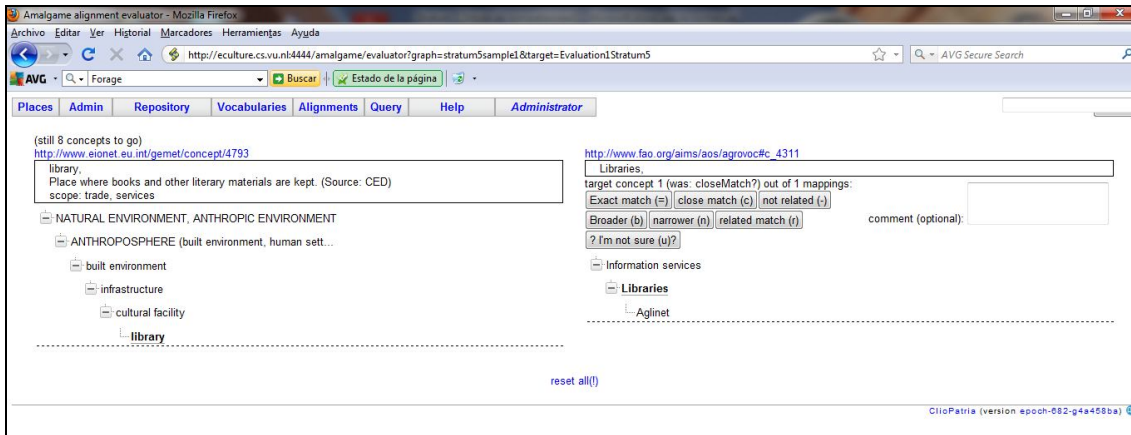
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. La única diferència és l’ús del singular en GEMET i el plural en AGROVOC. Els termes superiors indiquen que els tesaurus tenen punts de vista diferent, GEMET fa referència al terme “llacuna” com a part d’una àrea costanera, i AGROVOC ho assigna a característiques fisiogràfiques. La jerarquia i cobertura dels tesaurus també és diferent.

5.12 Alineament 12: landscape *versus* landscape



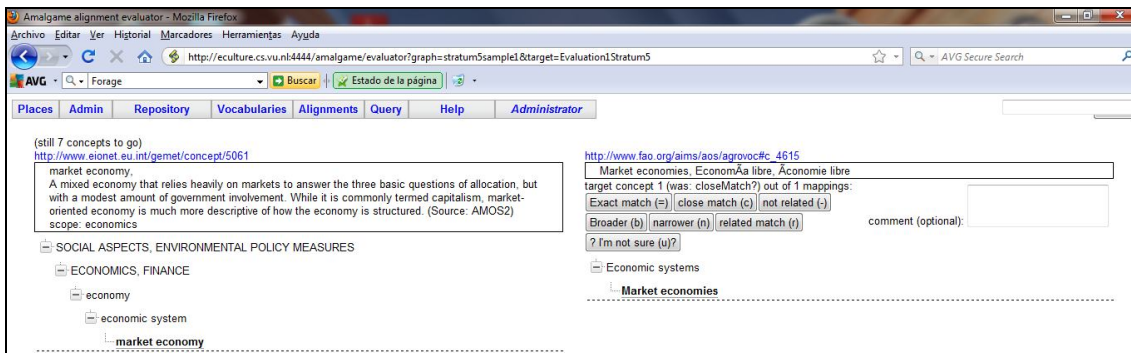
- **Avaluació:** relació de tipus “Close Match” o equivalència de quasi-sinònims. Tot i que “landscape management” pot ser intercanviable per “landscape conservation” segons AGROVOC. En aquest cas es considera que estem al davant d’una relació de “close match” per les diferències lèxiques, conceptuals i jeràrquiques. Els termes superiors fan referència a punts de vista diferents, GEMET considera que “landscape management” forma part de la gestió mediambiental i “landscape conservation” en AGROVOC fa referència a la conservació de la natura. De fet el concepte “conservació” està inclòs en la disciplina de “gestió mediambiental”.

5.13 Alineament 13: library *versus* libraries



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. La única diferència és l’ús del singular en GEMET i el plural en AGROVOC. Els termes superiors indiquen que els tesaurus tenen punts de vista diferents. GEMET ho inclou com una infraestructura cultural i AGROVOC com un servei d’informació.

5.14 Alineament 14: market economy *versus* market economies



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes alineats tenen el mateix punt de vista. La única diferència és l’ús del singular en GEMET i el plural en AGROVOC. Els termes superiors també tenen equivalència exacta i també es diferencien per l’ús del singular i plural respectivament.

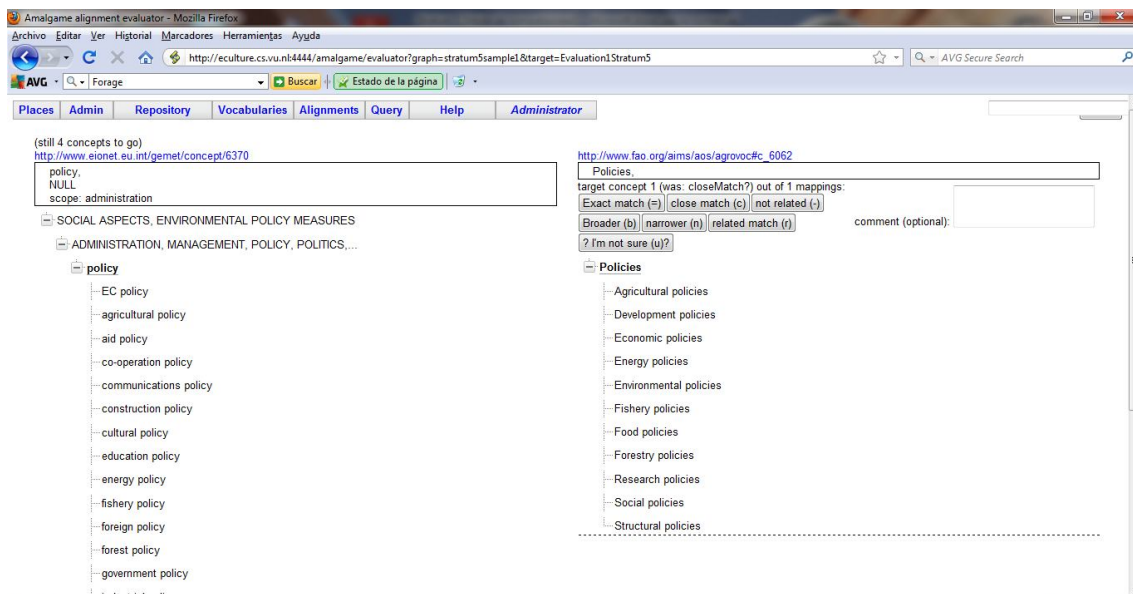
5.15 Alineament 15: natural heritage *versus* natural heritage

- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes alineats tenen el mateix punt de vista. També destacar la diferència d’estructura jeràrquica entre els dos conceptes, concretament en aquest cas és GEMET qui té una major cobertura, ja que inclou termes específics i AGROVOC no els inclou.

5.16 Alineament 16: parasite versus parasites

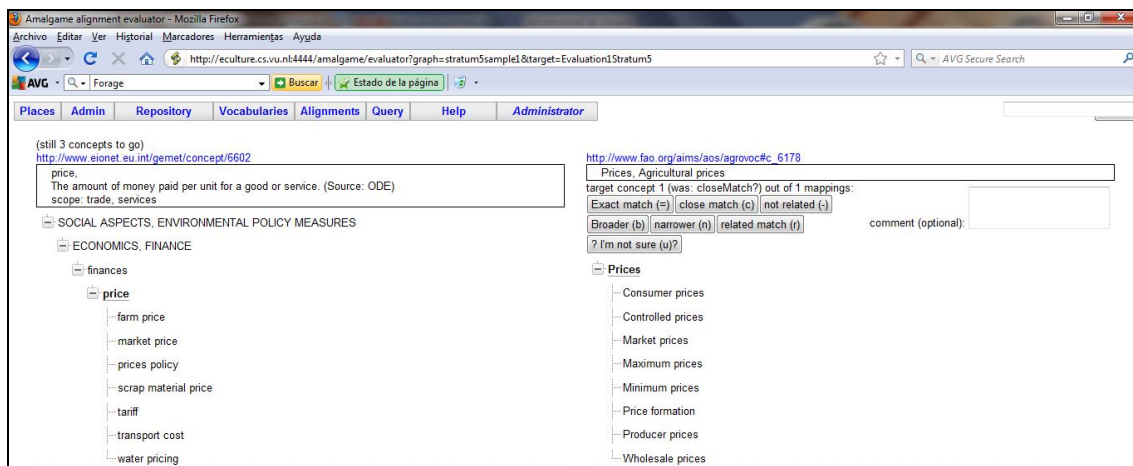
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes alineats tenen el mateix punt de vista ja que els dos tesaurus identifiquen el terme com un tipus d’organisme. GEMET utilitza el terme en singular i AGROVOC el plural. Les estructures jeràrquiques tenen diferències tant a nivell superior com a nivell inferior del terme, essent AGROVOC el tesaurus amb major termes específics.

5.17 Alineament 17: policies versus policies



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. GEMET utilitza el terme en singular i AGROVOC el plural. Tot i que els conceptes tenen el mateix significat, i ambdós s’utilitzen en el mateix sentit (informació relacionada amb polítiques), AGROVOC centra el terme en l’àmbit agrari, i els descriptors específics són també diferents de GEMET, també inclou el punt de vista agrari tot i que inclou una visió més ampla del terme.

5.18 Alineament 18: price versus prices



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. GEMET utilitza el terme en singular i AGROVOC el plural. Tot i que els conceptes tenen el mateix significat, i ambdós s’utilitzen en el mateix sentit (informació relacionada amb preus), els seus descriptors específics són diferents però no incompatibles. Aquest cas és un exemple, de com l’alineament de tesaurus és útil per ampliar la cobertura d’un tesaurus.

5.19 Alineament 19: rural population versus rural population

The screenshot shows the Amalgame alignment evaluator interface in a Mozilla Firefox browser. The URL is <http://eculture.cs.vu.nl:4444/amalgame/evaluator?graph=stratum5sample1&target=Evaluation1Stratum5>. The interface has a navigation menu with 'Places', 'Admin', 'Repository', 'Vocabularies', 'Alignments', 'Query', 'Help', and 'Administrator'. The main content area is split into two columns. The left column shows a GEMET concept: 'rural population, The total number of persons inhabiting an agricultural or pastoral region. (Source: RHW) scope: agriculture'. Below it is a tree structure: 'SOCIAL ASPECTS, ENVIRONMENTAL POLICY MEASURES' -> 'SOCIETY' -> 'demography' -> 'human population' -> 'rural population'. The right column shows an AGROVOC concept: 'Rural population, Społeczność wiejska'. Below it is a tree structure: 'Human population' -> 'Rural population' -> 'Agricultural population'. Between the two columns, there is a 'target concept 1 (was: closeMatch?) out of 1 mappings:' section with buttons for 'Exact match (=)', 'close match (c)', and 'not related (-)'. Below that are buttons for 'Broader (b)', 'narrower (n)', and 'related match (r)'. There is also a 'comment (optional):' field and a '? I'm not sure (u)?' button.

- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix punt de vista i el mateix terme superior, “human population”. Una altra diferència significativa és que AGROVOC inclou un terme específic (“Agricultural population”), i GEMET no inclou termes específics.

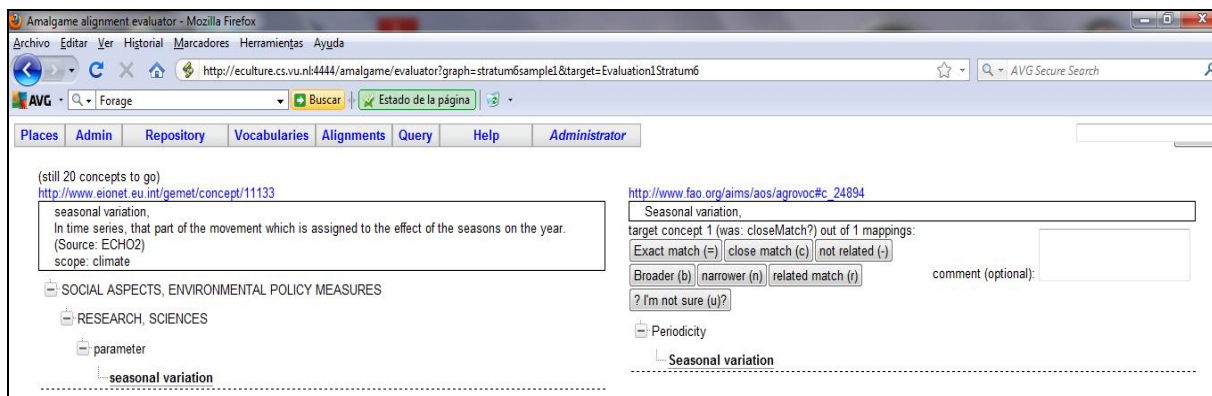
5.20 Alineament 20: waste reduction versus waste reduction

The screenshot shows the Amalgame alignment evaluator interface in a Mozilla Firefox browser. The URL is <http://eculture.cs.vu.nl:4444/amalgame/evaluator?graph=stratum5sample1&target=Evaluation1Stratum5>. The interface has a navigation menu with 'Places', 'Admin', 'Repository', 'Vocabularies', 'Alignments', 'Query', 'Help', and 'Administrator'. The main content area is split into two columns. The left column shows a GEMET concept: 'waste reduction, Practices that reduce the amount of waste generated by a specific source through the redesigning of products or patterns of production or consumption. (Source: TOE) scope: waste'. Below it is a tree structure: 'HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE...' -> 'WASTES, POLLUTANTS, POLLUTION' -> 'overburden' -> 'waste income' -> 'waste reduction'. The right column shows an AGROVOC concept: 'Waste reduction, Waste minimization'. Below it is a tree structure: 'Waste management' -> 'Waste reduction'. Between the two columns, there is a 'target concept 1 (was: closeMatch?) out of 1 mappings:' section with buttons for 'Exact match (=)', 'close match (c)', and 'not related (-)'. Below that are buttons for 'Broader (b)', 'narrower (n)', and 'related match (r)'. There is also a 'comment (optional):' field and a '? I'm not sure (u)?' button.

- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix punt de vista, tot i que no comparteixen el mateix terme superior (*waste income versus waste management*).

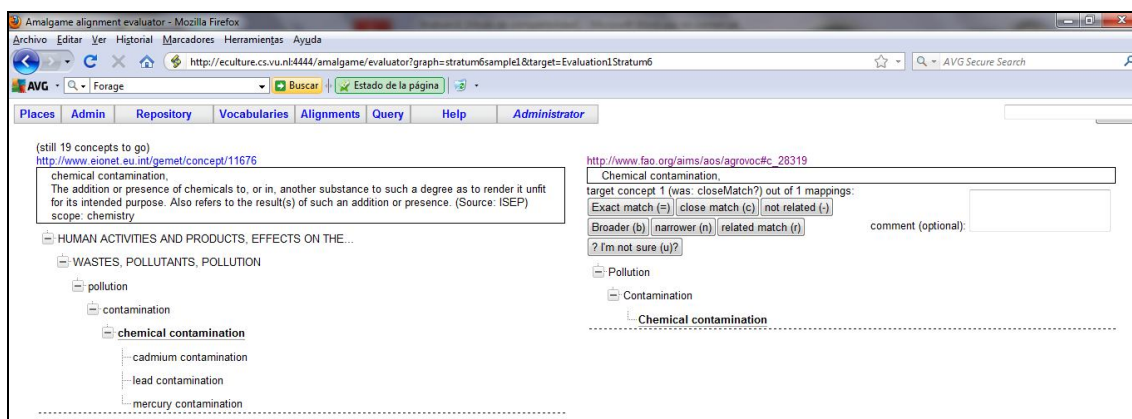
Stratum 6: Mostra aleatòria de 20 alineacions

6.1 Alineament 1: seasonal variation versus seasonal variation



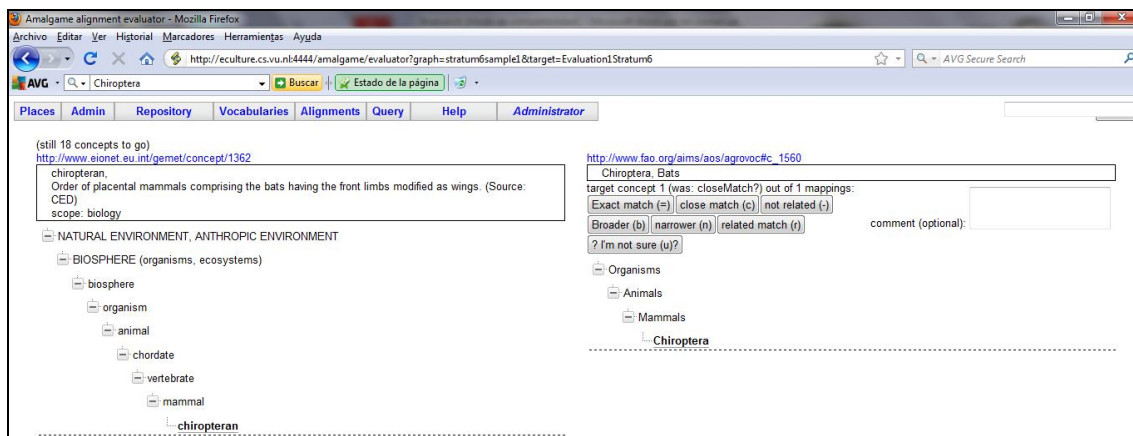
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. No obstant això, els seus termes superiors divergeixen (*parameter versus periodicity*), però en la definició del terme que apareix en GEMET, s’observa que els termes fan referència al mateix punt de vista, variació periòdica a nivell climàtic.

6.2 Alineament 2: chemical contamination versus chemical contamination



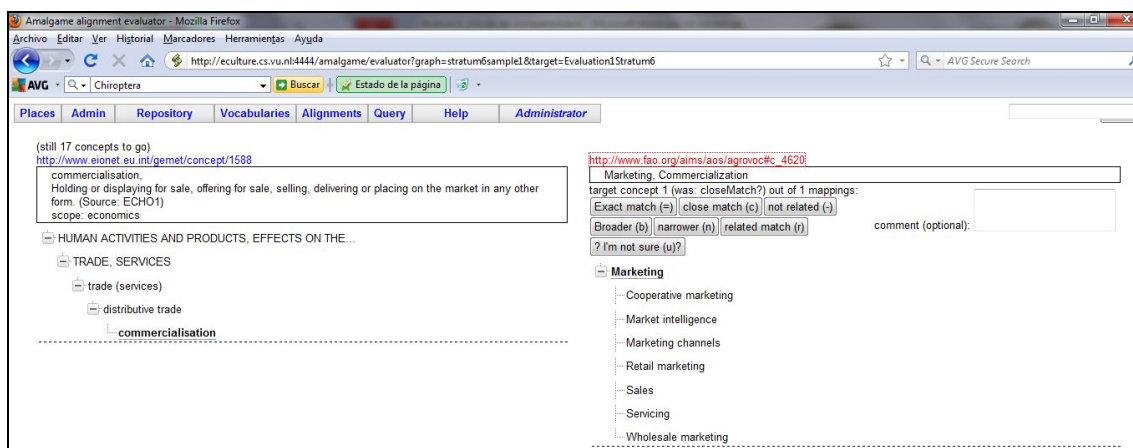
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Els termes superiors coincideixen fins a dos nivells, així com el mateix punt de vista (pol·lució). En aquest cas el tesaurus que ofereix una major especificitat d’aquest tema és GEMET que inclou fins a tres termes específics del concepte d’estudi.

6.3 Alineament 3: Chiropteran / Chiroptera = Exact Match



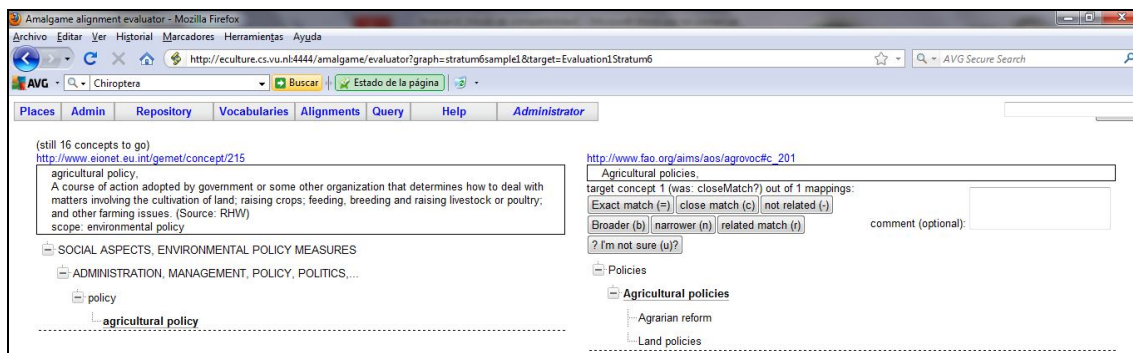
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Els termes superiors o *broader term* coincideixen i també el seu abast (biologia). L’estructura jeràrquica entre els dos tesaurus respecte el terme són completament diferents, tal i com s’observa en la imatge, *chiroptera* en GEMET té fins a 8 nivells superiors, a diferència d’AGROVOC que en té 3, fet que indica que GEMET en aquest aspecte ofereix una major especificitat.

6.4 Alineament 4: comercialization versus marketing



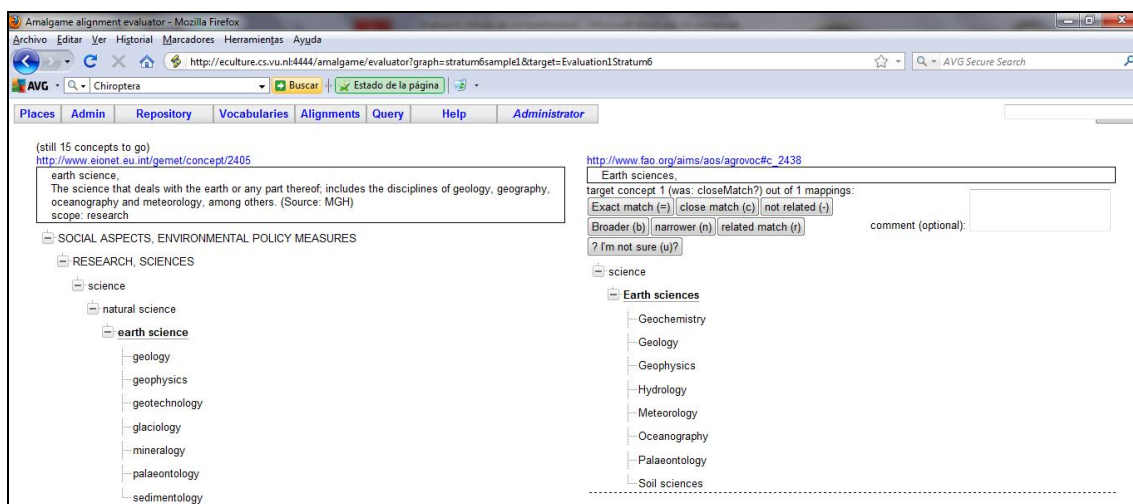
- **Avaluació:** alineament amb tipus de relació “close match” o quasi-sinònims. Ambdós termes tenen el mateix punt de vista econòmic, però no són termes estrictament sinònims, no obstant, en aquest cas es consideren termes quasi-sinònims perquè el tesaurus AGROVOC així els considera.

6.5 Alineament 5: agricultural policy versus agricultural policies



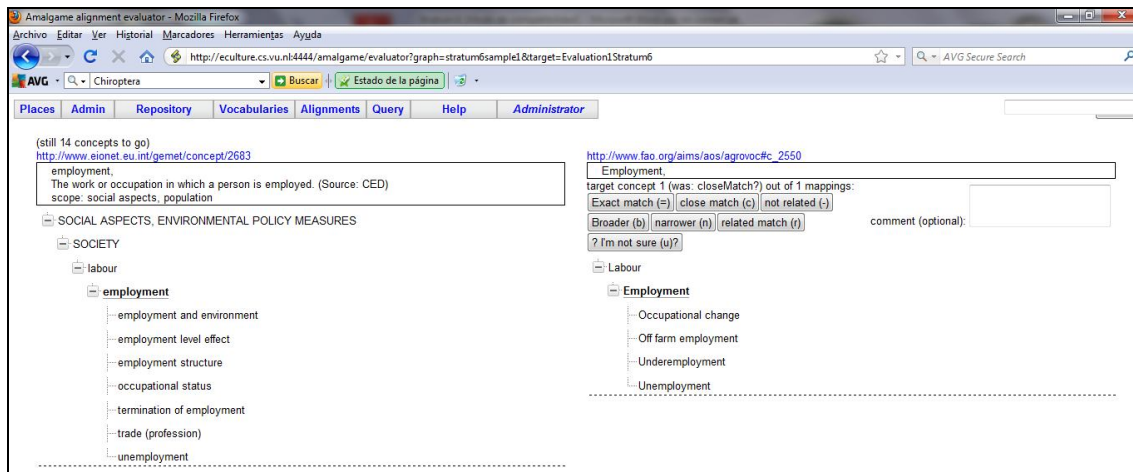
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. GEMET utilitza el terme en singular i AGROVOC el plural, ambdós termes tenen el mateix punt de vista i tenen també el mateix terme superior o *broader term*. A nivell jeràrquic AGROVOC inclou termes específics i GEMET no.

6.6 Alineament 6: earth science versus earth science



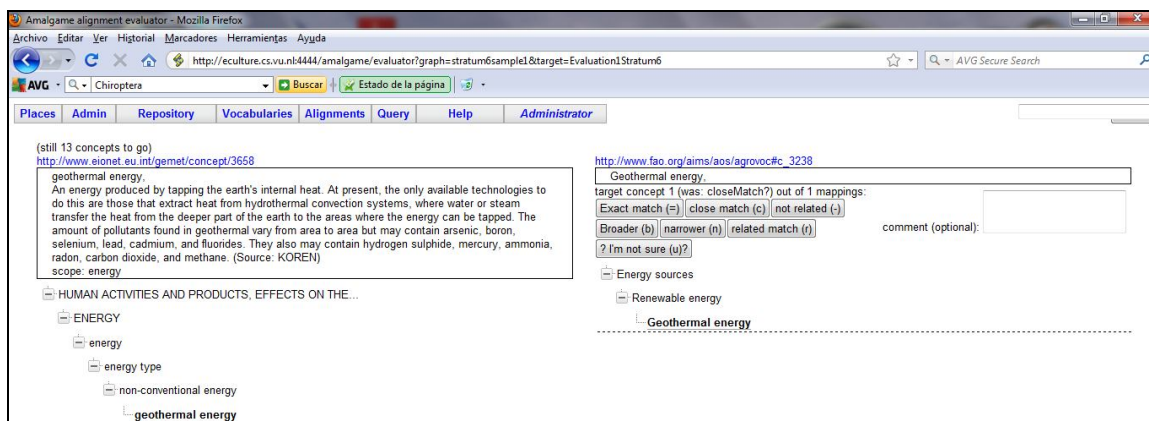
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. GEMET utilitza el terme en singular i AGROVOC el plural, ambdós termes tenen el mateix punt de vista (ciència). La seva estructura jeràrquica és diferent tant a nivell superior com a nivell inferior, però és bastant similar. Aquest cas és un exemple, de com l’alineament de tesaurus és útil per ampliar la cobertura d’un tesaurus.

6.7 Alineament 7: employment versus employment



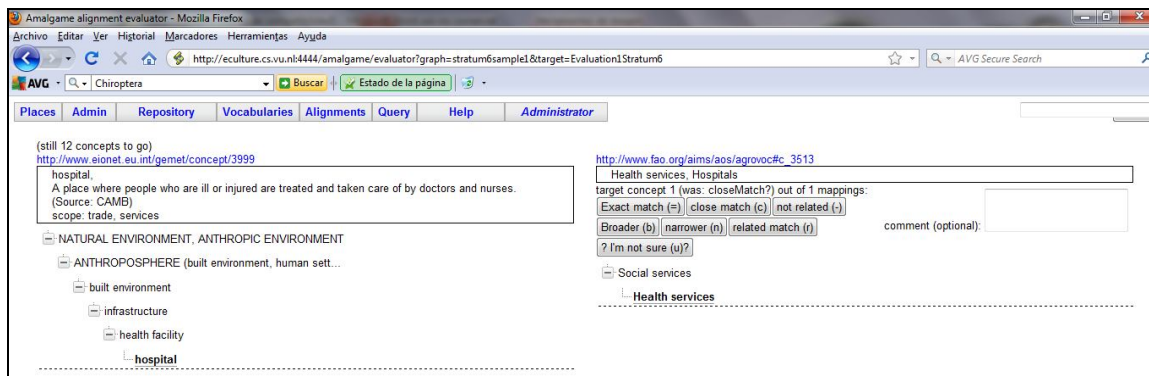
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, i el mateix terme superior o *broader term* (*labour*).

6.8 Alineament 8: geothermal energy versus geothermal energy



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic però no comparteixen el mateix terme superior (*geothermal energy versus geothermal energy*).

6.9 Alineament 9: hospitals versus health services



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Cada tesaurus utilitza un terme diferents per identificar el terme “hospitals”, i l’estructura jeràrquica entre els tesaurus també és diferent però ambdós termes són sinònims pel que fa al tesaurus AGROVOC. A més a més de compartir el mateix punt de vista, per tant, s’estableix una relació de tipus “exact match”.

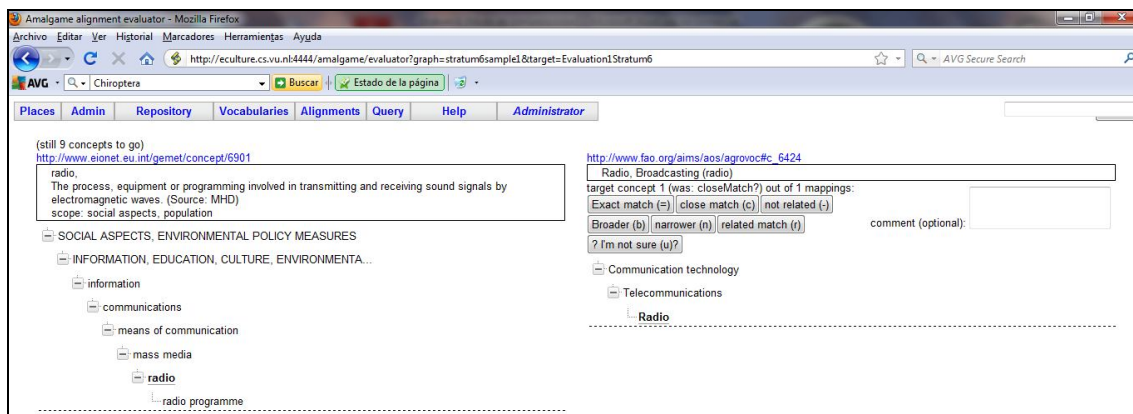
6.10 Alineament 10: monopoly versus monopolies

- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, però no comparteixen el mateix terme superior (*restriction on competition versus market structure*), ni tampoc tenen la mateixa estructura jeràrquica. Recalcar la diferència en l’ús de singular (GEMET) i plural (AGROVOC).

6.11 Alineament 11: ASEAN versus ASEAN

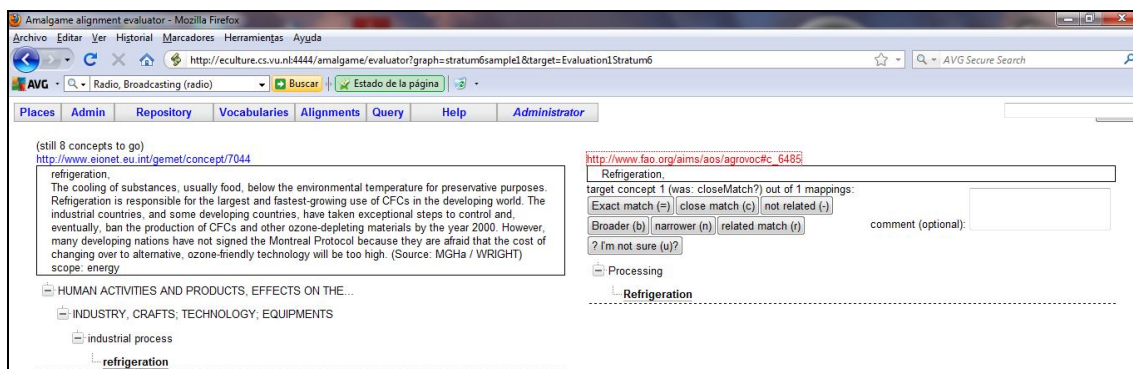
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, i comparteixen el mateix terme superior (*international organizations*).

6.12 Alineament 12: radio versus radio



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, però no comparteixen el mateix terme superior (*mass media versus telecommunications*), ni tampoc tenen la mateixa estructura jeràrquica. El tesaurus GEMET ofereix una cobertura més amplia i especificitat, tant a nivell de termes superiors com en el terme específic del concepte d’estudi. AGROVOC no inclou cap terme específic del concepte “Radio”.

6.13 Alineament 13: refrigeration versus refrigeration



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, però no comparteixen el mateix terme superior (*industrial process versus processing*), tot i que tenen el mateix enfocament, ja que identifiquen el terme “refrigeració” com un procés.

6.14 Alineament 14: activated sludge versus activated sludge

- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic (procés biològic per al tractament d’aigües), però no comparteixen el mateix terme superior (*sludge treatment versus adsorbents*), i aquests no tenen el mateix abast.

6.15 Alineament 15: social development versus social change

- **Avaluació:** aquest alineament s’avalua com una relació de tipus “close match” o de quasi sinònims. D’una banda els termes alineats en cada tesaurus tenen connotacions diferents (desenvolupament social versus canvi social), d’altra banda poden ser considerats sinònims, tal i com apunta AGROVOC. En d’altres casos s’han assignat aquest tipus de relació “exact match”, però en aquest cas essent conscients que fa falta una avaluació més acurada (avaluant col·leccions, interessos institucionals, etc.), s’assigna un valor de quasi sinònims, ja que el terme superior de “social change” en AGROVOC és “socioeconomic development” fet que indica que la cobertura de “desenvolupament social” estigui en el terme superior i no pas en l’específic. Les jerarquies i punts de vista entre els tesaurus són diferents (comportament social versus socioeconòmic).

6.16 Alineament 16: soil conservation *versus* soil consevation

The screenshot shows the Amalgame alignment evaluator interface in Mozilla Firefox. The browser address bar displays the URL: <http://eculture.cs.vu.nl:4444/amalgame/evaluator?graph=stratum6sample1&target=Evaluation1Stratum6>. The interface includes a navigation menu with options: Places, Admin, Repository, Vocabularies, Alignments, Query, Help, and Administrator. The main content area is divided into two columns. The left column displays a concept from GEMET: 'soil conservation' with its definition and a hierarchical tree structure under 'ENVIRONMENTAL POLICY'. The right column displays a concept from FAO: 'Soil conservation' with its definition and a hierarchical tree structure under 'Resource management'. The alignment is marked as an 'Exact match'.

- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic (conservació i preservació del sòl), però no comparteixen el mateix terme superior (*environmental conservation versus resource conservation*), i aquests no tenen el mateix abast. Cal destacar que GEMET està especialitzat amb medi ambient.

6.17 Alineament 17: sustainable development *versus*

The screenshot shows the Amalgame alignment evaluator interface in Mozilla Firefox. The browser address bar displays the URL: <http://eculture.cs.vu.nl:4444/amalgame/evaluator?graph=stratum6sample1&target=Evaluation1Stratum6>. The interface includes a navigation menu with options: Places, Admin, Repository, Vocabularies, Alignments, Query, Help, and Administrator. The main content area is divided into two columns. The left column displays a concept from GEMET: 'sustainable development' with its definition and a hierarchical tree structure under 'ENVIRONMENTAL POLICY'. The right column displays a concept from FAO: 'Sustainable development' with its definition and a hierarchical tree structure under 'Socioeconomic development'. The alignment is marked as an 'Exact match'.

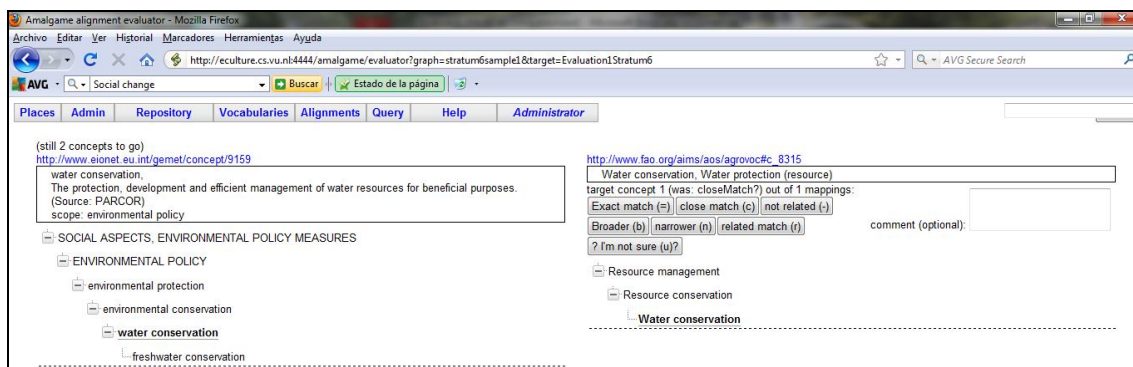
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, però no comparteixen el mateix terme superior (desenvolupament medi ambiental *versus* desenvolupament socioeconòmic), i cal tenir en compte que no tenen el mateix abast. Aquest és un altre cas que l’alineament és considera “exact match” però segons el criteri bibliotecari o institucional podrien ser assignats com a termes quasi sinònims o amb equivalència quasi exacta (*close match*).

6.18 Alineament 18: telecommunications *versus* telecommunications



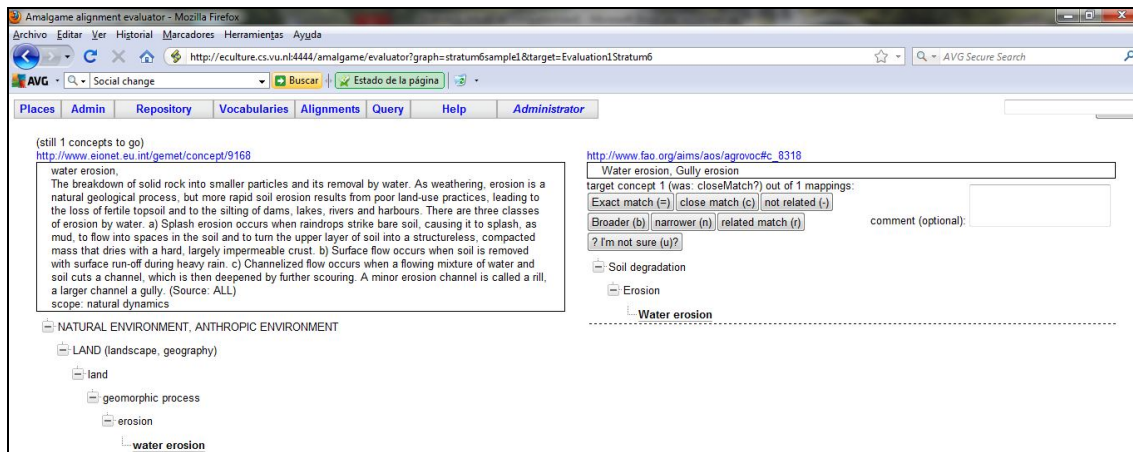
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, però no comparteixen el mateix terme superior (communications *versus communication technology*), no obstant aquests *Broader terms*, si tenen el mateix abast temàtic i tenen un tipus de relació d’equivalència exacta.

6.19 Alineament 19: water conservation *versus* water conservation



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta, perquè ambdós termes tenen el mateix abast temàtic (conservació, protecció i ús eficient de l’aigua com a recurs). Com en el cas d’alineament 6.16, no comparteixen el mateix terme superior (*environmental conservation versus resource conservation*), aquests no tenen el mateix abast (protecció mediambiental *versus* gestió de recursos, el darrer terme pot està inclòs dintre del primer, relació de *Broader Term*).

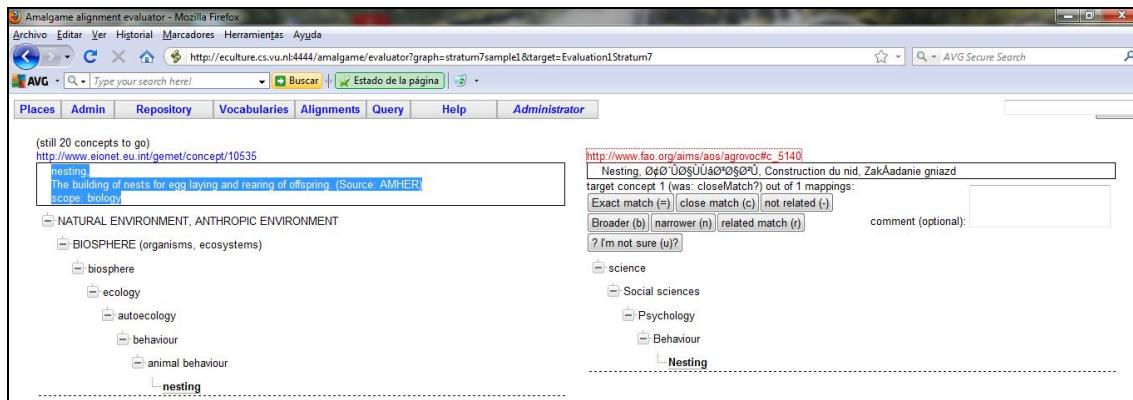
6.20 Alineament 20: water erosion *versus* water erosion



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta, perquè ambdós termes tenen el mateix lèxic i abast temàtic i també comparteixen el mateix terme superior.

Stratum 7: Mostra aleatòria de 20 alineacions

7.1 Alineament 1: nesting versus nesting



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta, perquè ambdós termes tenen el mateix lèxic i abast temàtic. Ambdós termes no comparteixen el mateix terme superior (behaviour and animal behaviour), però aquests si tenen el mateix punt de vista (comportament animal), i podrien ser considerats com a quasi sinònims (segons el tipus de col·lecció existent, etc.). L’estructura jeràrquica superior del concepte objecte d’estudi, és també diferent i mostra també una cobertura diferent (ecologia versus ciències socials), els termes no tenen una correspondència exacta ni relacionada a excepció dels conceptes ja citats.

7.2 Alineament 2: continuing education versus continuing education

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'continuing education' is displayed with its definition and a hierarchical tree structure. On the right, the FAO concept 'Continuing education, Further education' is displayed with its definition and a 'closeMatch?' status. The interface includes a search bar, navigation tabs, and a detailed comparison of the two concepts.

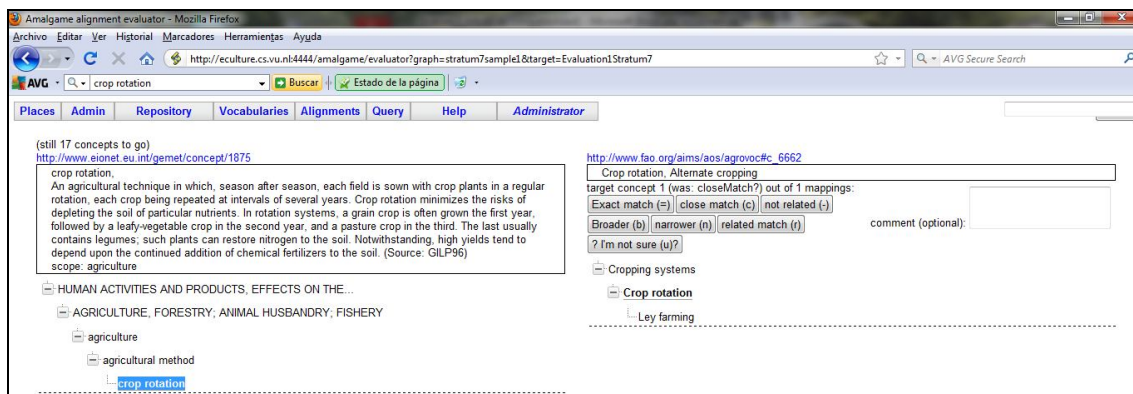
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta, perquè ambdós termes tenen el mateixa correspondència quant a lèxic i abast temàtic. Ambdós termes comparteixen un dels termes superiors (educació), però en el cas de GEMET, no és el terme immediatament superior. La jerarquia de GEMET com es dona en la majoria dels casos contempla més nivells jeràrquics, sobretot a nivell superior.

7.3 Alineament 3: comercial law versus comercial law

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'commercial law' is displayed with its definition and a hierarchical tree structure. On the right, the FAO concept 'Commercial law' is displayed with its definition and a 'closeMatch?' status. The interface includes a search bar, navigation tabs, and a detailed comparison of the two concepts.

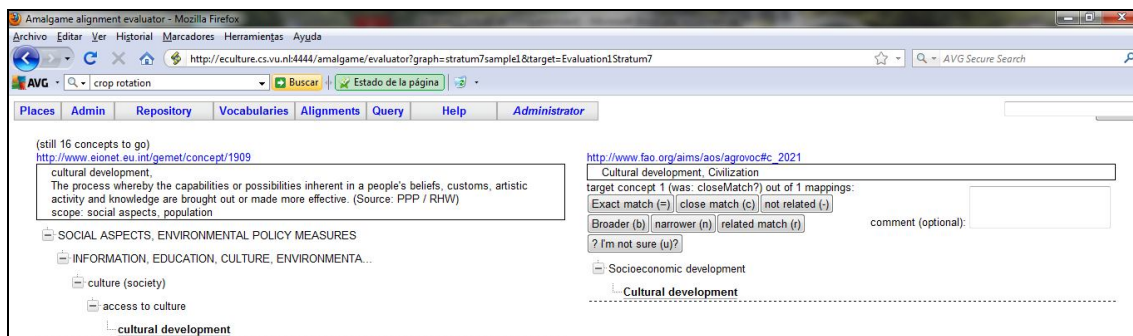
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta, perquè ambdós termes tenen el mateixa correspondència quant a lèxic i abast temàtic. Ambdós termes comparteixen un dels termes superiors (*Law*), però en el cas de GEMET, no és el terme immediatament superior. La jerarquia de GEMET com es dona en la majoria dels casos contempla més nivells jeràrquics, en aquest cas tant a nivell superior com inferior.

7.4 Alineament 4: crop rotation versus crop rotation



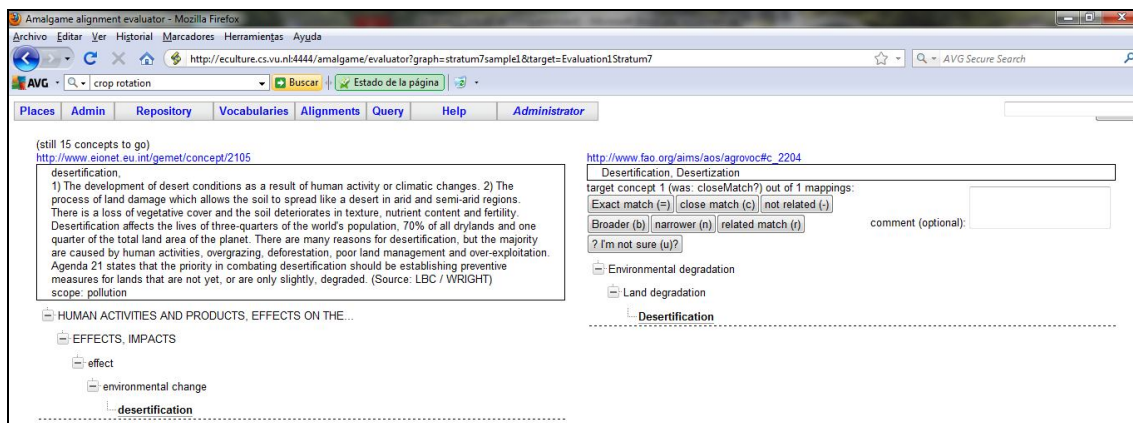
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix significat i punt de vista, tot i que no comparteixen el mateix terme superior o BT (*agricultural method versus cropping systems*). Els termes superiors tenen un punt de vista similar, però el concepte definit per AGROVOC (*cropping systems*), potser un terme específic de GEMET sota el concepte “agricultural method”. En aquest exemple, el que es demostra és l’especificitat del tesaurus AGROVOC amb la temàtica “agricultura”.

7.5 Alineament 5: cultural development versus cultural development



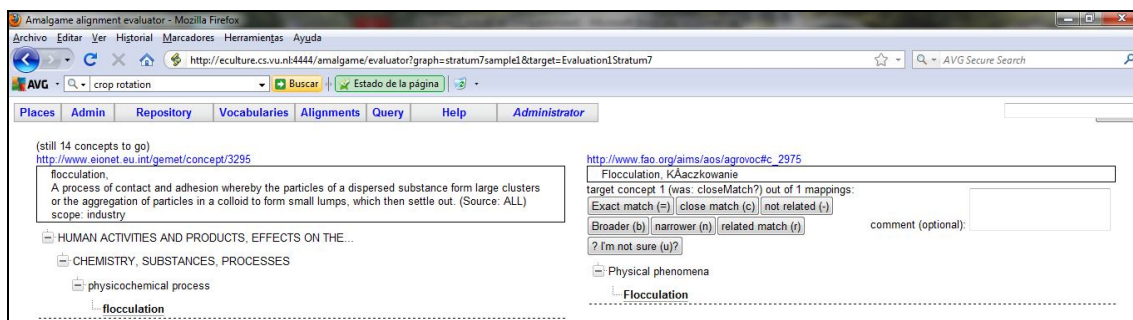
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, però no comparteixen el mateix terme superior (*access to culture versus socioeconomic development*), i cal tenir en compte que no tenen el mateix abast.

7.6 Alineament 6: desertification versus desertification



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, i tot i que no comparteixen el mateix terme superior (*environmental change versus land degradation*), ambdós termes superiors tenen el mateix punt de vista (la “degradació” com a fenomen mediambiental). No obstant això, segons la col·lecció a la que vagi dirigida el tipus de relació podria ser de “close match”.

7.7 Alineament 7: flocculation versus flocculation



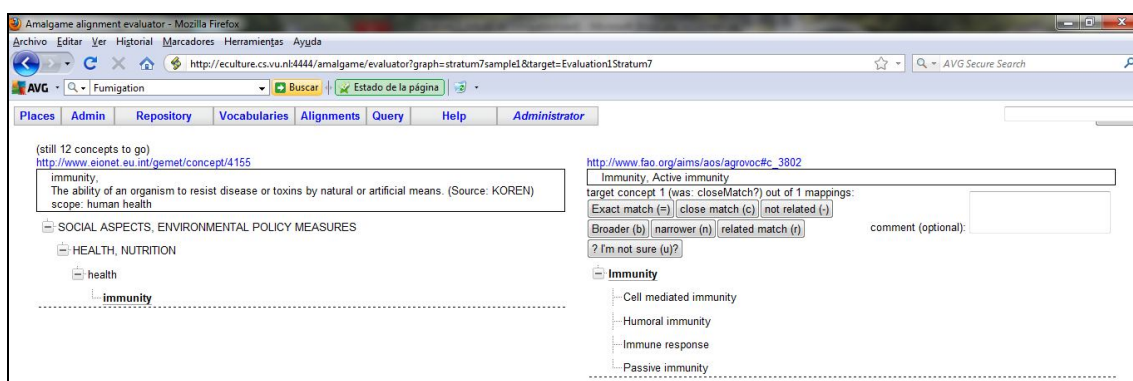
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, però els seus termes superiors són diferents (procés fisicoquímic *versus* fenomen físic), no obstant les connotacions que tenen de forma separada, en aquest context es podrien considerar d’equivalència exacta o quasi equivalència, perquè es refereixen al mateix enfocament: “flocculation”, és un procés químic que forma part del procés de potabilització d’aigües d’origen superficial i/o tractament d’aigües (domèstiques, industrials i mineres).

7.8 Alineament 8: fumigation versus fumigation



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, però tot i que tenen termes superiors diferents, aquests inclouen el mateix punt de vista (“fumigació” com un tractament o mètode), i amb un punt de vista genèric. AGROVOC especifica en la nota d’abast, que en alguns casos d’ús del terme de forma específica, s’ha d’emprar el terme que correspongui, per exemple, “smoking” com a mètode de preservació d’aliments o “soil fumigation” per a la fumigació del sòl.

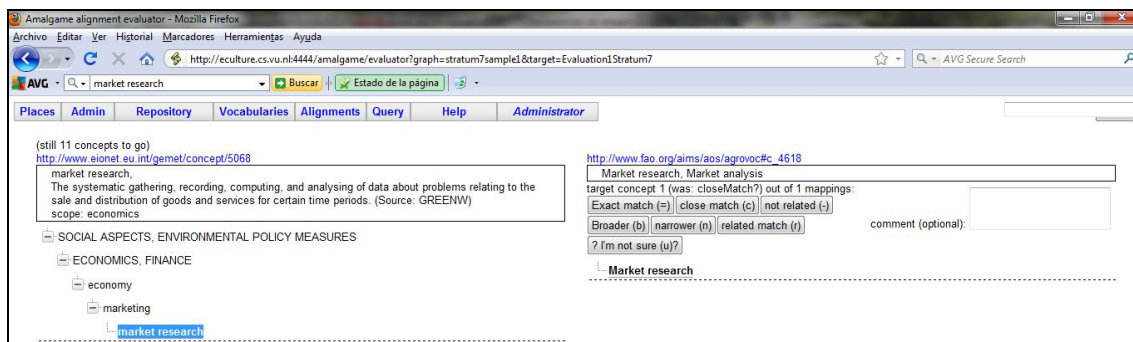
7.9 Alineament 9: immunity versus immunity



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. No obstant això, existeixen diferències entre les jerarquies dels tesaurus (AGROVOC inclou termes específics i GEMET no), i entre els termes superiors (*health versus biological phenomena*¹²⁴), fet que indica que els punts de vista són diferents però relacionats. En el cas de GEMET, l’abast temàtic del concepte “immunitat” és salut humana, en canvi per AGROVOC el terme té una visió més ampla i ho descriu com un fenomen biològic (potser humà o animal).

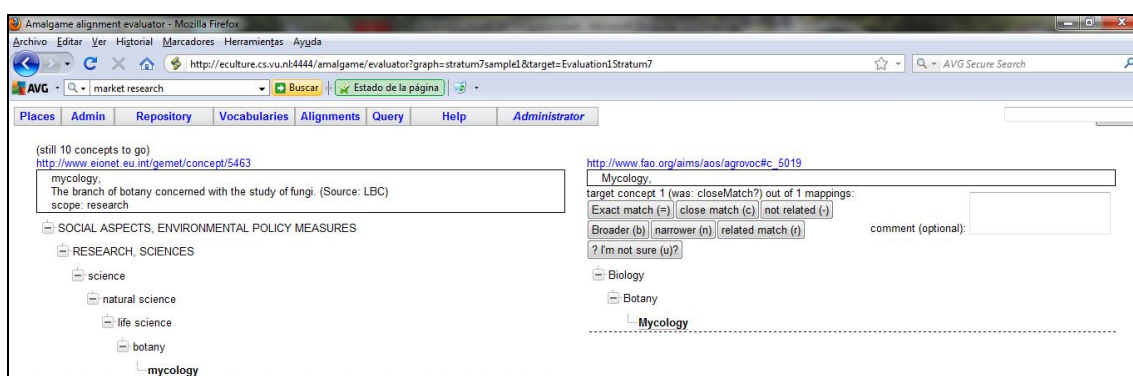
7.10 Alineament 10: market research versus market research

¹²⁴Veure AGROVOC Term Code 3802. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=3802&mylang_interface=en&myLanguage=EN [Última consulta: 01/09/2013]



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, però tenen termes superiors diferents, els termes superiors diferents (marketing *versus* activities¹²⁵), en AGROVOC el concepte “Marketing” és un terme relacionat (RT), de “Market research” (vegeu AGROVOC 4618 Code Term). Una altra diferència entre la representació d’aquests termes, és que GEMET no inclou relacions de tipus associatiu (vegeu 5068 GEMET concept¹²⁶).

7.11 Alineament 11: mycology *versus* mycology



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Com s’observa en la imatge superior, ambdós termes tenen el mateix lèxic i abast temàtic i també comparteixen el mateix terme superior. Com en d’altres casos, la jerarquia superior de GEMET és molt més àmplia que en AGROVOC.

7.12 Alineament 12: privatisation *versus* privatization

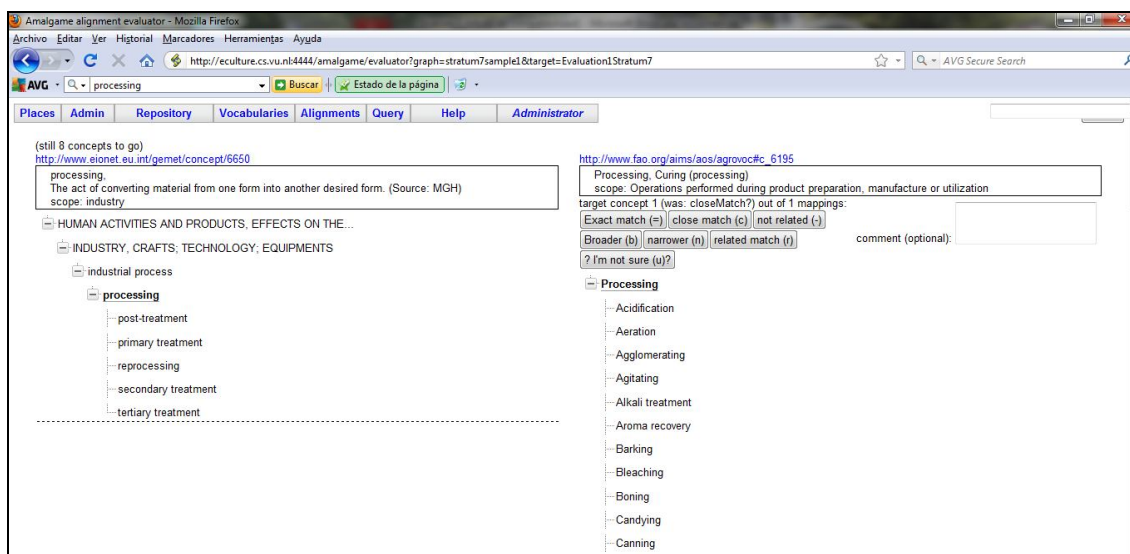
¹²⁵Veure AGROVOC Code Term 4618. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=4618&mylang_interface=en&myLanguage=EN [Última consulta: 01/09/2013]

¹²⁶Veure GEMET 5068 Concept. Accessible a: <http://www.eionet.europa.eu/gemet/concept/5068> [Últim accés 01/09/2012]



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Com s’observa en la imatge superior, ambdós termes tenen una escriptura ortogràfica diferent però són paraules idèntiques, de fet, canvia l’ús de la ‘s’ en anglès britànic per l’ús de la ‘z’ per anglès americà. Com s’ha observat al llarg d’aquest experiment, quan es tracta de termes econòmics cada tesaurus mostra un representació jeràrquica dels termes, i en aquest cas no canvia, els termes superiors són diferents, però l’abast temàtic en aquest cas és el mateix (transferència de poders públics al sector privat).

7.13 Alineament 13: processing versus processing



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta, ja que coincideixen amb lèxic, definició i tenen un abast quasi sinònim. La diferència principal quant al punt de vista temàtic, és que AGROVOC és més generalista en l’abast del terme (terme superior: “activitates”), i GEMET dóna un enfocament únicament com a “procés industrial”, tal i com expressa el seu terme superior “industrial process”. Les jerarquies entre els termes són bastant diferents

tant a nivell superior com inferior. En la versió online d'AGROVOC¹²⁷, hi ha 122 termes que estan relacionats amb el concepte “processing” (1 terme superior, 100 termes específics i 21 termes relacionats).

7.14 Alineament 14: activated carbon *versus* activated carbon

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept is displayed: 'activated carbon' with a definition: 'A powdered, granular or pelleted form of amorphous carbon characterized by a very large surface area per unit volume because of an enormous number of fine pores. (Source: LANDY)'. The scope is 'materials'. The hierarchy is: HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE... > PRODUCTS, MATERIALS > product > chemical product > activated carbon. On the right, the AGROVOC concept is displayed: 'Activated carbon, Activated charcoal'. The target concept is 'Activated carbon'. The alignment type is 'Exact match (=)'. The interface also shows navigation options like 'Places', 'Admin', 'Repository', 'Vocabularies', 'Alignments', 'Query', 'Help', and 'Administrator'.

- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta, ja que coincideixen amb lèxic, definició i abast temàtic (àmbit químic). No obstant això, la jerarquia entre els conceptes és diferent, i com es pot observar en els termes superiors associats al concepte avaluat, també hi ha una diferència conceptual entre els conceptes. GEMET categoritza el concepte com a un producte químic, i AGROVOC com una substància química, però ambdós termes són sinònims.

7.15 Alineament 15: risk *versus* risk

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept is displayed: 'risk' with a definition: 'The expected number of lives lost, persons injured, damage to property and disruption of economic activity due to a particular natural phenomenon, and consequently the product of the probability of occurrence and the expected magnitude of damage. (Source: GUNN / RRDA)'. The scope is 'disasters, accidents, risk'. The hierarchy is: SOCIAL ASPECTS, ENVIRONMENTAL POLICY MEASURES > RISKS, SAFETY > risk > cancer risk > chemical risk > environmental risk > major risk > natural risk > nuclear risk > pollution risk > residual risk > technological hazard. On the right, the AGROVOC concept is displayed: 'Risk'. The target concept is 'Risk'. The alignment type is 'Exact match (=)'. The interface also shows navigation options like 'Places', 'Admin', 'Repository', 'Vocabularies', 'Alignments', 'Query', 'Help', and 'Administrator'.

- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta, ja que coincideixen amb lèxic, definició i abast temàtic. Existeixen diferències jeràrquiques, essent GEMET el tesaurus que té un major desenvolupament de descriptors relacionats amb el

¹²⁷ Veure AGROVOC Concept Term 6195. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=6195&mylang_interface=en&myLanguage=EN [Última accés: 01/09/2013]

concepte d'estudi, sobretot a nivell de termes específics (AGROVOV no inclou termes específics). Cal destacar que el terme superior en AGROVOC és "phenomena"¹²⁸ i en GEMET no té terme superior però està agrupat dintre el grup de "riscs i seguretat".

7.16 Alineament 16: sea water versus sea water

- **Avaluació:** alineament amb tipus de relació "exact match" o amb relació d'equivalència exacta, ja que coincideixen amb lèxic, definició i abast temàtic. Ambdós conceptes tenen termes superiors diferents (*salt water versus saline water*), però considerats sinònims. La diferència principal entre els conceptes està en la seva estructura jeràrquica, la d'AGROVOC és més simple i la que ofereix GEMET dona una major granularitat al concepte.

7.17 Alineament 17: benthos versus benthos

- **Avaluació:** alineament amb tipus de relació "exact match" o amb relació d'equivalència exacta. Ambdós termes tenen el mateix abast temàtic, els seus termes superiors són diferents (*aquatic organisms versus aquatic*

¹²⁸Veure AGROVOC Concept Term 6612. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=6612&mylang_interface=en&myLanguage=EN [Últim accés: 10/09/2013]

communities), però són termes que tenen el mateix punt de vista i que es poden considerar sinònims.

7.18 Alineament 18: soil degradation versus soil degradation

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'soil degradation' (ID 7857) is displayed with its definition and a hierarchical tree structure. On the right, the AGROVOC concept 'Soil degradation, Soil deterioration' (ID 7168) is displayed with its definition and a hierarchical tree structure. The interface shows an 'Exact match' relationship between the two concepts.

- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, al igual que els seus termes superiors (*degradation of the environment versus environmental degradation*¹²⁹), però a més a més, GEMET els identifica com a termes amb relació de tipus “exact match”¹³⁰. Estructures jeràrquiques entre els tesaurus es mostren diferents, GEMET inclou molts més nivells superiors que AGROVOC, i aquest darrer inclou més termes específics.

19 Alineament 19: sulphur dioxide versus sulphur dioxide

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'sulphur dioxide' (ID 8202) is displayed with its definition and a hierarchical tree structure. On the right, the AGROVOC concept 'Sulphur dioxide, Sulphur dioxide' (ID 7515) is displayed with its definition and a hierarchical tree structure. The interface shows an 'Exact match' relationship between the two concepts.

- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic (òxids), i fins a dos nivells superiors similars.

7.20 Alineament 20: titanium versus titanium

¹²⁹ Veure AGROVOC Term Code 7168. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=7168&mylang_interface=en&myLanguage=EN [Últim accés: 10/09/2013]

¹³⁰ Veure GEMET 7857 Concept. Accessible a: <http://www.eionet.europa.eu/gemet/concept/7857> [Últim accés: 10/09/2013]

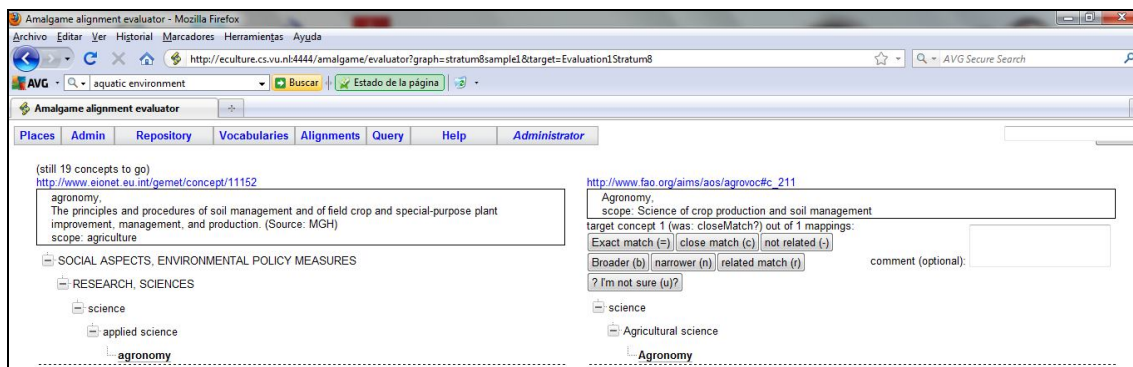
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic i tenen el mateix terme superior (*transition elements*). No obstant això, analitzant tota la jerarquia si s’observen diferències quant a l’abast temàtic en altres nivells superiors. Per exemple, GEMET agrupa el concepte “transition element” com a un element químic i AGROVOC ho considera com un element metàl·lic.

Stratum 8: Mostra aleatòria de 20 alineaments

8.1 Alineament 1: aquatic environment versus aquatic environment

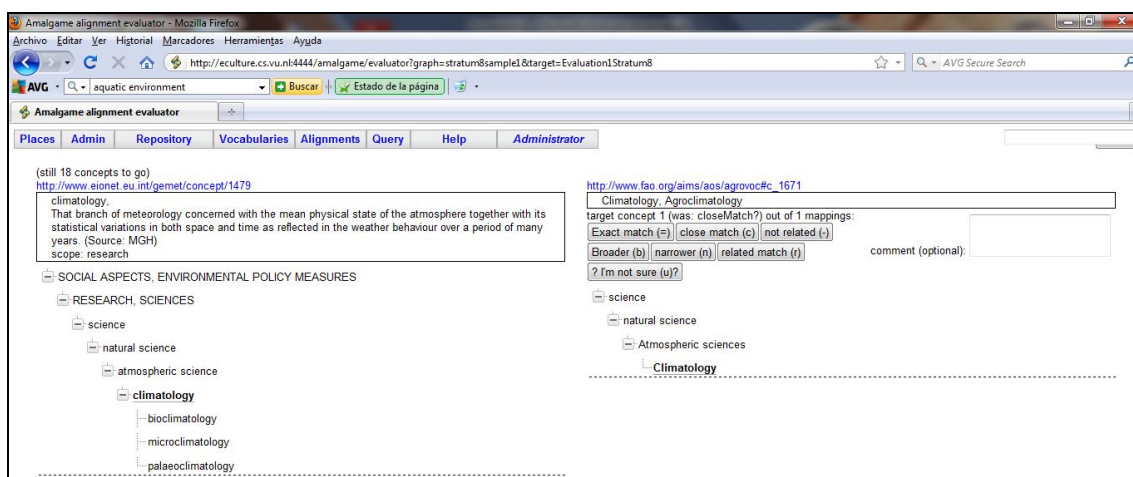
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, i tot i que no tenen el mateix terme superior (*natural environment versus environment*), el terme superior d’AGROVOC “Aquatic environment” si es representat en GEMET com a terme superior de “natural environment”. D’una banda es confirma que l’abast temàtic dels conceptes és equivalent (medi ambient), però també es posa de manifest les diferències jeràrquiques entre els tesaurus. I en aquest cas, a més a més, AGROVOC té un nivell d’especificitat major.

8.2 Alineament 2: agronomy versus agronomy



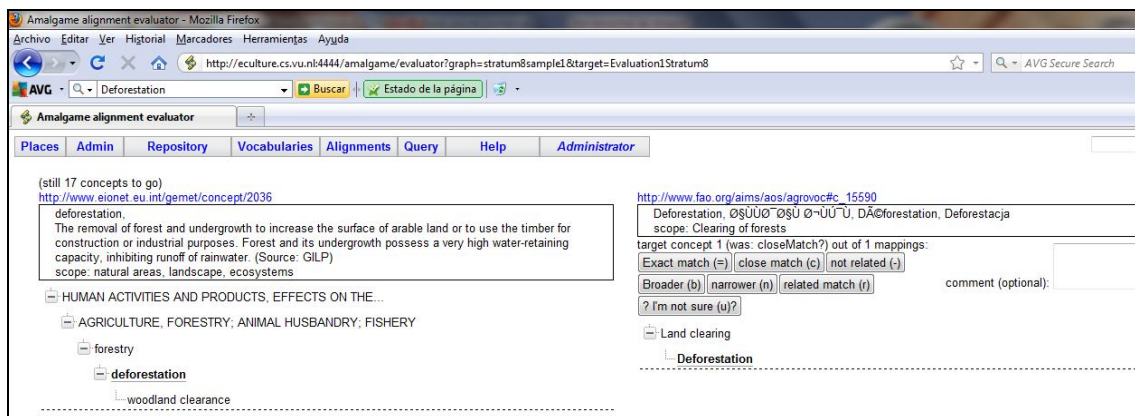
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, i tot i que no tenen el mateix terme superior (*applied science versus agricultural science*), el terme superior d’AGROVOC “Agricultural science” si es representat en GEMET com a terme superior d’“applied science”, és a dir, el punt de vista d’ambdós tesaurus és equivalent (àmbit agrari i científic).

8.3 Alineament 3: climatology versus climatology

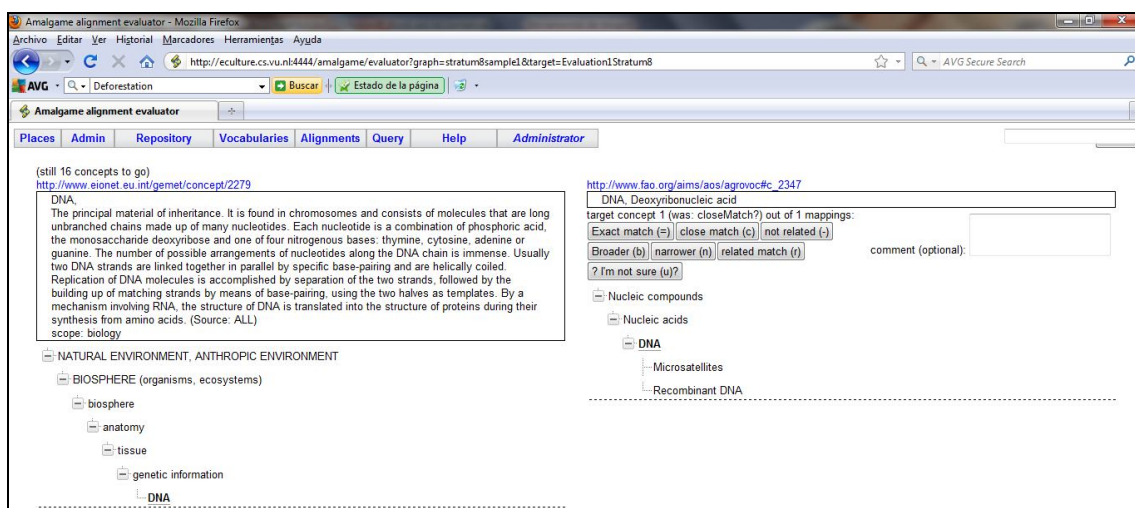


- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, i fins a tres nivells superiors coincidents (*Science → natural science → atmospheric science/s → Climatology*). En aquest cas, el nivell d’especificitat del concepte és major en el tesaurus GEMET, on “climatology” té fins a tres termes específics, mentre que AGROVOC no en té.

8.4 Alineament 4: deforestation versus Deforestation



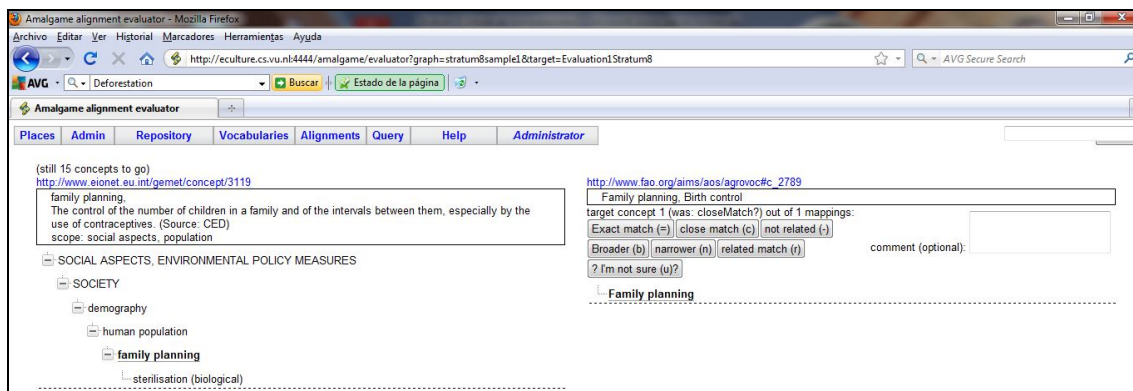
8.5 Alineament 5: DNA versus DNA



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. Ambdós termes tenen el mateix abast temàtic, però no comparteixen ni termes superiors (*genetic information versus nucleic acids*), ni tampoc la seva estructura jeràrquica. Tot i així, tal i com s’indica en el tesaurus GEMET¹³¹, ambdós conceptes són equivalents.

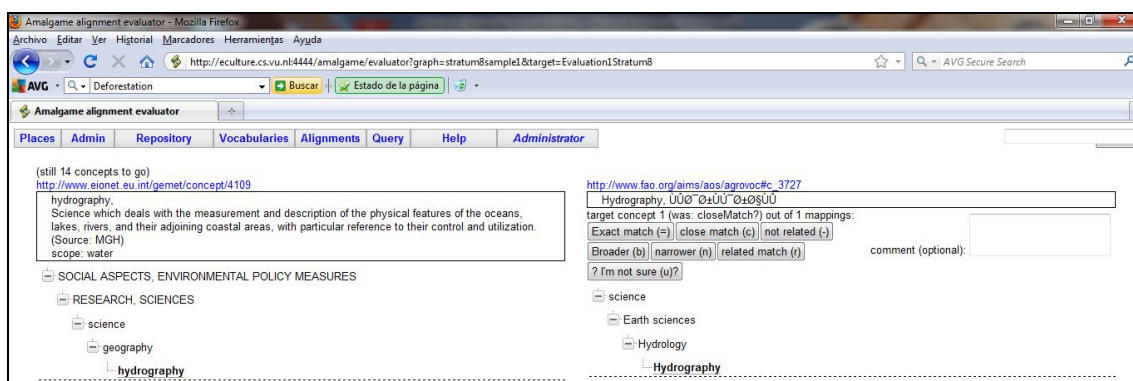
8.6 Alineament 6: family planning versus family planning

¹³¹ Veure GEMET 2279 Concept. Accessible a: <http://www.eionet.europa.eu/gemet/concept/2279> [Últim accés: 10/09/2013]



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. A més a més GEMET¹³², estableix que els conceptes tenen un tipus de relació d’“exact match”. No obstant això, els conceptes equivalents no comparteixen els termes superiors ni l’abast temàtic d’aquests (*human population versus reproduction control*¹³³).

8.7 Alineament 7: hydrography versus hydrography



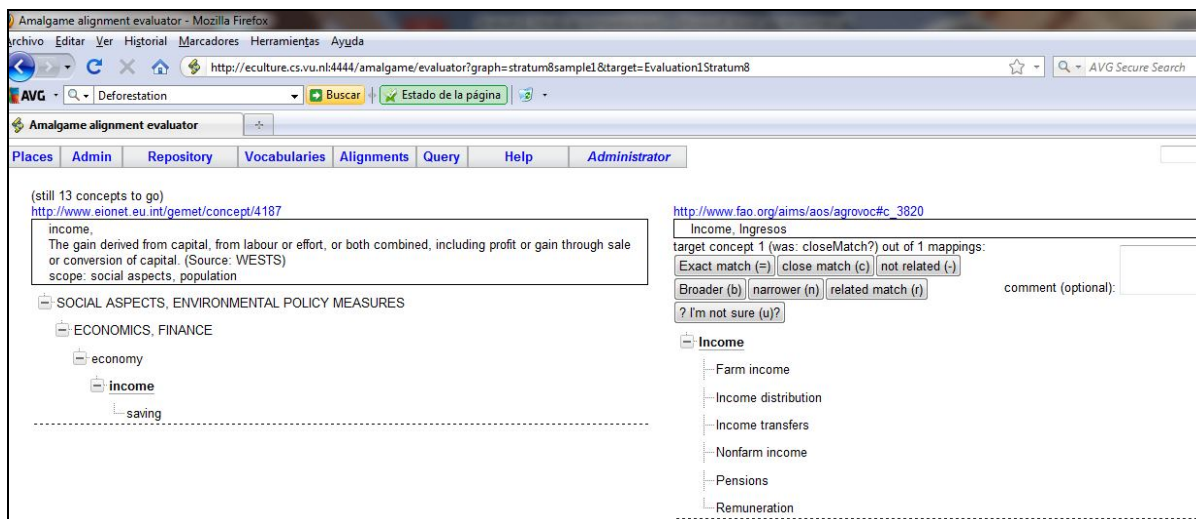
- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. A més a més GEMET¹³⁴, estableix que els conceptes tenen un tipus de relació d’“exact match”. No obstant això, aquests conceptes considerats com a termes sinònims i amb relació d’equivalència exacta, no comparteixen els termes superiors ni l’abast temàtic d’aquests (*geography versus hydrology*).

8.8 Alineament 8: income versus income

¹³² Veure GEMET 3119 Concept. Accessible: <http://www.eionet.europa.eu/gemet/concept/3119> [Últim accés: 10/09/2013]

¹³³ Veure AGROVOC Concept 2789. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=2789&mylang_interface=en&myLanguage=EN [Últim accés: 10/09/2013]

¹³⁴ Veure GEMET 4109 Concept. Accessible: <http://www.eionet.europa.eu/gemet/concept/4109> [Últim accés: 10/09/2013]



- **Avaluació:** alineament amb tipus de relació “exact match” o amb relació d’equivalència exacta. A més a més GEMET¹³⁵, estableix que els conceptes tenen un tipus de relació d’“exact match”. No obstant això, no comparteixen els termes superiors (economy versus entities¹³⁶), i l’abast temàtic en GEMET és molt més ampli (economia, aspectes socials), mentre que en AGROVOC està més centrat en l’àmbit agrícola.

8.9 Alineament 9: infiltration versus infiltration



- **Avaluació:** alineament amb tipus de relació “exact match”, ja que els conceptes són equivalents tant en lèxic com en significat. No obstant això, l’estructura jerarquia d’ambdós tesaurus és diferent tant en conceptes superiors com inferiors, i no es comparteix el mateix “broader term” (hydrologic balance versus water circulation).

8.10 Alineament 10: labour versus labour

¹³⁵ Veure GEMET 4187 Concept. Accessible a: <http://www.eionet.europa.eu/gemet/concept/4187> [Últim accés: 10/09/2013]

¹³⁶ Veure AGROVOC Concept Term 3820. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=3820&mylang_interface=en&myLanguage=EN [Últim accés: 10/09/2013]

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'labour' is displayed with its definition: 'One of the factors of production. It includes all the exertions - manual, physical or mental - by individuals, directed towards the production of wealth. (Source: GOOD) scope: social aspects, population'. Below this is a hierarchical tree structure under 'SOCIAL ASPECTS, ENVIRONMENTAL POLICY MEASURES' and 'SOCIETY', with 'labour' expanded to show sub-terms like 'employment', 'labour force', 'labour market', 'labour relations', 'occupation', 'organisation of work', and 'working condition'. On the right, the AGROVOC concept 'Labour, Labor' is shown with a 'target concept 1 (was: closeMatch?) out of 1 mappings:' and an 'Exact match (=)' relationship. The interface also includes navigation tabs like 'Places', 'Admin', 'Repository', 'Vocabularies', 'Alignments', 'Query', 'Help', and 'Administrator'.

- **Avaluació:** alineament amb tipus de relació “exact match”, ja que els conceptes són equivalents en lèxic, significat i abast temàtic. En aquest cas el thesaurus GEMET destaca per la especificitat expressada pel concepte, ja que inclou fins a 7 “narrower terms”. No obstant això, en la versió actual del thesaurus AGROVOC¹³⁷ si s’amplia el nivell jeràrquic inferior, i es mostra també una gran especificitat temàtica (12 descriptors NT).

8.11 Alineament 11: landscape conservation versus landscape conservation

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'landscape conservation' is displayed with its definition: 'The safeguarding, for public enjoyment, of landscape and of opportunities for outdoor recreation, tourism and similar activities; the concept includes the preservation and enhancement not only of what has been inherited but the provision of new amenities and facilities. (Source: GILP96) scope: environmental policy'. Below this is a hierarchical tree structure under 'SOCIAL ASPECTS, ENVIRONMENTAL POLICY MEASURES' and 'ENVIRONMENTAL POLICY', with 'landscape conservation' expanded to show sub-terms like 'environmental protection', 'environmental conservation', and 'landscape conservation'. On the right, the AGROVOC concept 'Landscape conservation, Landscape management' is shown with a 'target concept 1 (was: closeMatch?) out of 1 mappings:' and an 'Exact match (=)' relationship. The interface also includes navigation tabs like 'Places', 'Admin', 'Repository', 'Vocabularies', 'Alignments', 'Query', 'Help', and 'Administrator'.

- **Avaluació:** alineament amb tipus de relació “exact match”, ja que els conceptes són equivalents tant en lèxic com en significat. No obstant això, els termes immediatament superiors (*broader term*), són diferents i fan que la jerarquia tinguin punts de vista lleugerament diferents (*landscape conservation versus nature conservation*), en GEMET el concepte s’engloba dintre les polítiques mediambientals i la protecció del

¹³⁷ Veure AGROVOC Concept Term 4128. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=4128&mylang_interface=en&myLanguage=EN [Últim accés: 10/09/2013]

medi ambient, i en AGROVOC el concepte fa referència a la gestió de recursos naturals com és la conservació paisatgística.

8.12 Alineament 12: marine geology versus marine geology

The screenshot shows the Amalgame alignment evaluator interface. On the left, the concept 'marine geology' is defined with a URL and a description. Its taxonomy includes: SOCIAL ASPECTS, ENVIRONMENTAL POLICY MEASURES > RESEARCH, SCIENCES > science > natural science > earth science > geology > marine geology. On the right, the concept 'Marine geology' is defined with a URL. Its taxonomy includes: science > Earth sciences > Geology > Marine geology. The alignment is marked as 'Exact match (=)'.

- **Avaluació:** alineament amb tipus de relació “exact match”. Ambdós conceptes són equivalents tant en lèxic, significat com en abast temàtic. Els termes superiors o “broader terms” també són coincidents: “marine geology” < geology < earth science

8.13 Alineament 13: nitrification versus nitrification

The screenshot shows the Amalgame alignment evaluator interface. On the left, the concept 'nitrification' is defined with a URL and a description. Its taxonomy includes: HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE... > CHEMISTRY, SUBSTANCES, PROCESSES > chemical process > chemical reaction > nitrification. On the right, the concept 'Nitrification' is defined with a URL. Its taxonomy includes: Chemical reactions > Nitrification. The alignment is marked as 'Exact match (=)'.

- **Avaluació:** alineament amb tipus de relació “exact match”. Ambdós conceptes són equivalents tant en lèxic, significat com en abast temàtic. Els termes superiors o “broader terms” també són coincidents (*chemical reaction versus chemical reactions*).

8.14 Alineament 14: osmosis versus osmosis

Amalgame alignment evaluator - Mozilla Firefox
 http://eculture.cs.vu.nl:4444/amalgame/evaluator?graph=stratum8sample1&target=Evaluation1Stratum8
 AVG - Osmosis
 Amalgame alignment evaluator
 Places Admin Repository Vocabularies Alignments Query Help Administrator
 (still 7 concepts to go)
 http://www.eionet.eu.int/gemet/concept/5949
 osmosis.
 The passage of a solvent through a semipermeable membrane separating two solutions of different concentrations. A semipermeable membrane is one through which the molecules of a solvent can pass but the molecules of most solutes cannot. There is a thermodynamic tendency for solutions separated by such a membrane to become equal in concentration, the water (or other solvent) flowing from the weaker to the stronger solution. Osmosis will stop when the two solutions reach equal concentration, and can also be stopped by applying a pressure to the liquid on the stronger-solution side of the membrane. The pressure required to stop the flow from a pure solvent into a solution is a characteristic of the solution, and is called the osmotic pressure. Osmotic pressure depends only on the concentration of particles in the solution, not on their nature. (Source: DICCHE)
 scope: biology
 HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE...
 CHEMISTRY, SUBSTANCES, PROCESSES
 physicochemical process
 osmosis
 reverse osmosis
 http://www.fao.org/aims/aos/agrovoc#c_5442
 Osmosis, Osmotic flow
 target concept 1 (was: closeMatch?) out of 1 mappings:
 Exact match (=) close match (c) not related (-)
 Broader (b) narrower (n) related match (r) comment (optional):
 ? I'm not sure (u)?
 Physical phenomena
 Diffusion
 Osmosis

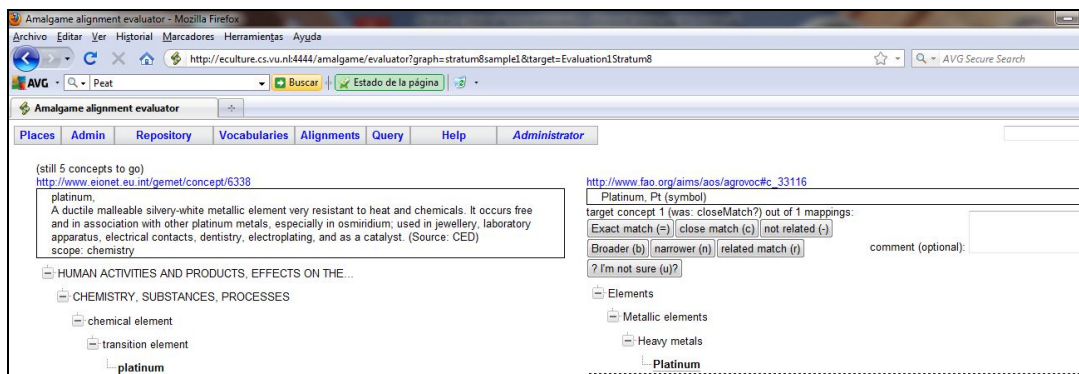
- **Avaluació:** alineament amb tipus de relació “exact match”. Ambdós conceptes són equivalents, tenen el mateix abast temàtic (fenomen o procés fisico-químic). No obstant això, el terme immediatament superior o “broader term” divergeix, “physicochemical process” en GEMET i “difusion” en AGROVOC, però el terme superior de “difusion” (“physical phenomena”), si té un punt de vista equivalent al terme superior d’osmosis representat per GEMET, “physicochemical process”.

8.15 Alineament 15: peat versus peat

Amalgame alignment evaluator - Mozilla Firefox
 http://eculture.cs.vu.nl:4444/amalgame/evaluator?graph=stratum8sample1&target=Evaluation1Stratum8
 AVG - Peat
 Amalgame alignment evaluator
 Places Admin Repository Vocabularies Alignments Query Help Administrator
 (still 6 concepts to go)
 http://www.eionet.eu.int/gemet/concept/6081
 peat.
 Unconsolidated soil material consisting largely of undecomposed or slightly decomposed organic matter accumulated under conditions of excessive moisture. (Source: LANDY)
 scope: materials
 HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE...
 PRODUCTS, MATERIALS
 product
 fuel
 fossil fuel
 peat
 http://www.fao.org/aims/aos/agrovoc#c_5647
 Peat
 target concept 1 (was: closeMatch?) out of 1 mappings:
 Exact match (=) close match (c) not related (-)
 Broader (b) narrower (n) related match (r) comment (optional):
 ? I'm not sure (u)?
 Natural resources
 Nonrenewable resources
 Mineral resources
 Peat

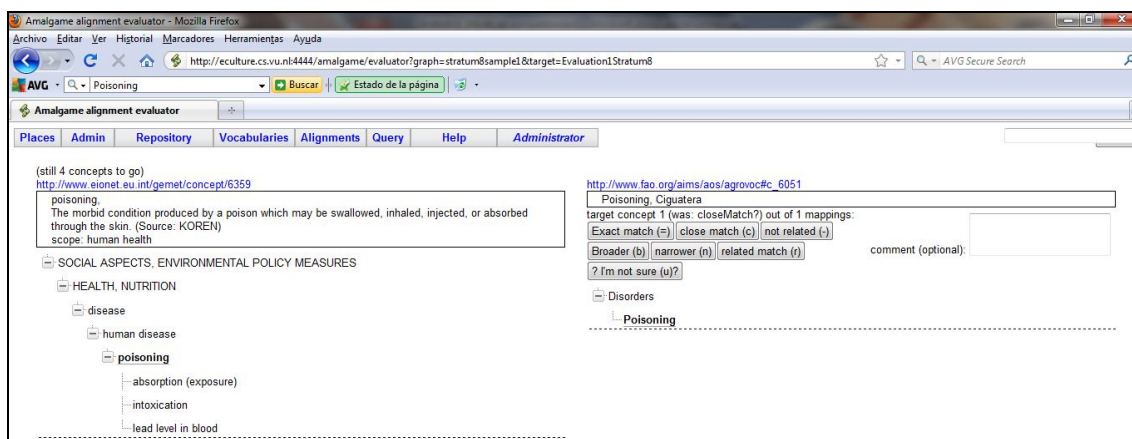
- **Avaluació:** alineament amb tipus de relació “exact match”. Ambdós conceptes són equivalents, tenen el mateix abast temàtic, minerals. No obstant això, el punt de vista d’ambdós tesaurus és diferent, GEMET ho considera el “mineral” com un producte o material i AGROVOC com un recurs natural. Aquesta distinció, no lleva que estem al davant de dos termes equivalents, però evidentment segons els criteris i ús del “mapeig de vocabularis”, són conceptes que es poden identificar amb relació de “quasi-equivalència” (*close match*).

8.16 Alineament 16: platinum versus plantinum



- **Avaluació:** alineament amb tipus de relació “exact match”. Es tracta del mateix concepte però cada tesaurus ho descriu de forma diferent, “platinum” com a element químic (GEMET) i “platinum” com a símbol de l’element químic. Conseqüentment, aquest és un dels casos on el tipus de relació potser també “close match” o tipus de relació de quasi-sinònims, aquest criteri dependrà de l’ús final del mapeig i/o objectiu de l’agent que hi hagi al darrera (elaboració d’un tercer tesaurus, integració de continguts de diferents procedències, etc.).

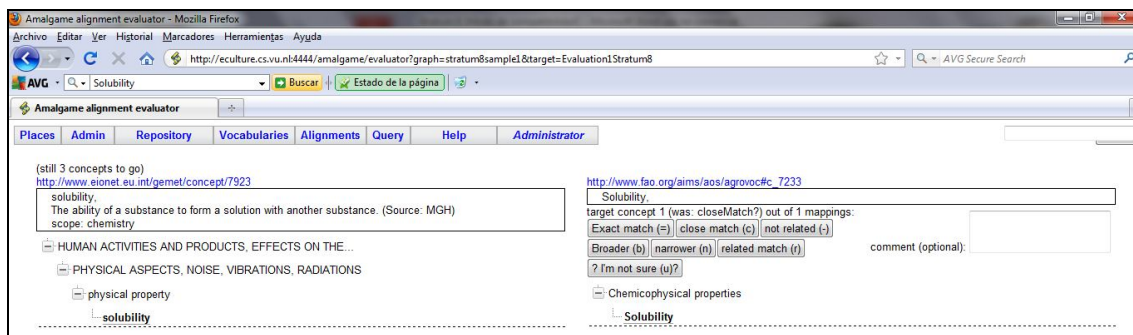
8.17 Alineament 17: poisoning versus poisoning



- **Avaluació:** alineament amb tipus de relació “exact match”. Es tracta del mateix concepte però cada tesaurus ho representa de forma diferent, tot i que tenen un mateix punt de vista, “salut alimentària”. En la jerarquia d’AGROVOC¹³⁸, aquest relaciona el concepte amb diferents termes específics que responen a diferents tipus de desordres o intoxicacions alimentàries (humans o animals). En canvi la jerarquia de GEMET engloba termes més amplis (intoxicació per exposició, intoxicació en general o nivell de plom en sang).

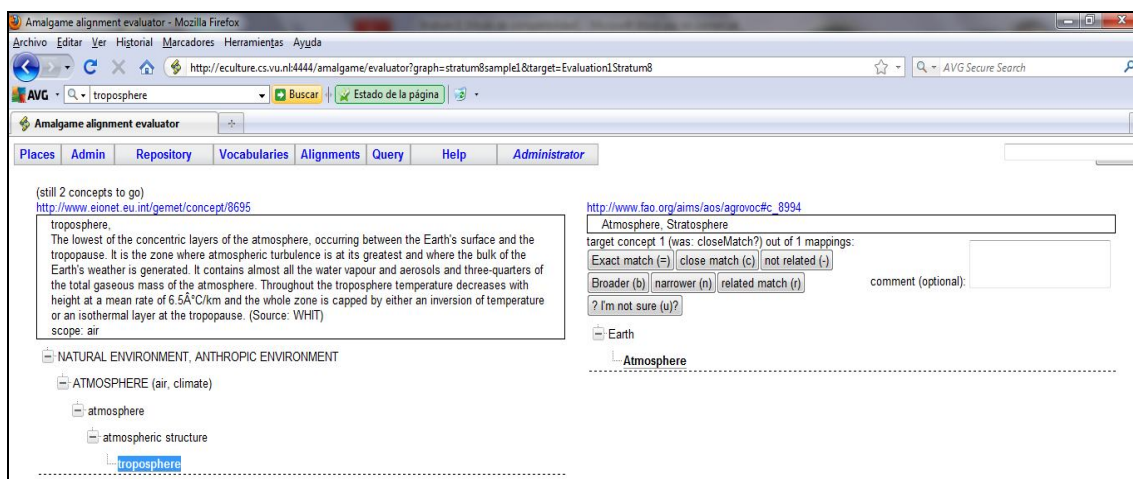
8.18 Alineament 18: solubility versus solubility

¹³⁸ Veure AGROVOC Term Code: 6051. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=6051&mylang_interface=en&myLanguage=EN [Últim accés: 10/09/2013]



- **Avaluació:** alineament amb tipus de relació “exact match”. Ambdós tesaurus tenen el mateix punt de vista i els seus termes genèrics són quasi-sinònims. Si es consulta el tesaurus GEMET¹³⁹, s’observa que el concepte “solubility” és polijeràrquic, ja que contempla dos termes genèrics, “chemical property” i “physical property”, aquests dos termes genèrics representen un sol terme genèric en el cas d’AGROVOC, que utilitza “chemicophysical property” com a terme genèric de “solubility”.

8.19 Alineament 19: Atmosphere is a BT of Troposphere (gemet)



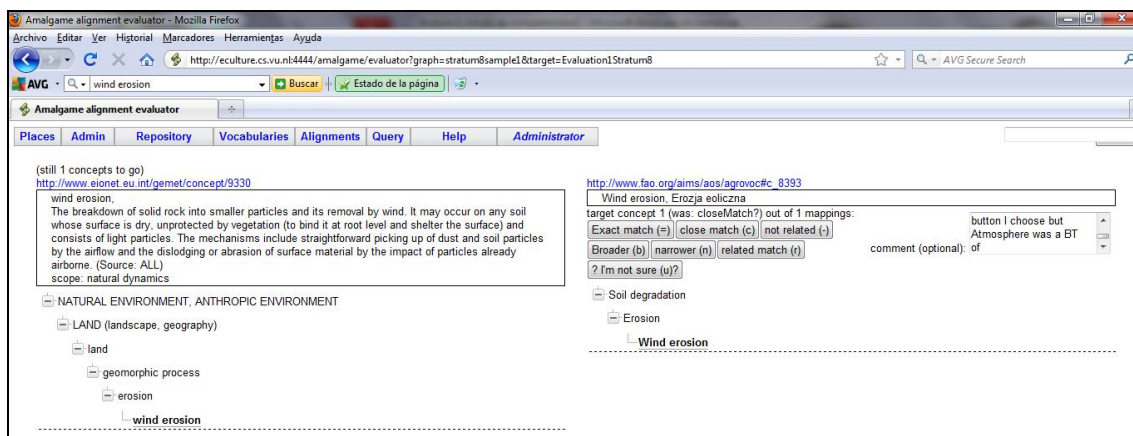
- **Avaluació:** alineament amb tipus de relació “exact match”. Ambdós tesaurus tenen el mateix punt de vista, inclús el tesaurus GEMET en la seva versió 3.1 online¹⁴⁰, contempla el terme “Atmosphere” d’AGROVOC com un tipus de relació d’“exact match”. Cal afegir però, que les jerarquies dels termes en ambdós tesaurus són diferents i que AGROVOC inclou una major especificitat¹⁴¹.

8.20 Alineament 20: wind erosion versus wind erosion

¹³⁹ Veure GEMET 7923 Concept. Accessible a: <http://www.eionet.europa.eu/gemet/concept/7923> [Últim accés: 10/09/2013]

¹⁴⁰ Veure GEMET 617 Concept. Accessible a: <http://www.eionet.europa.eu/gemet/concept/617> [Últim accés: 10/09/2013]

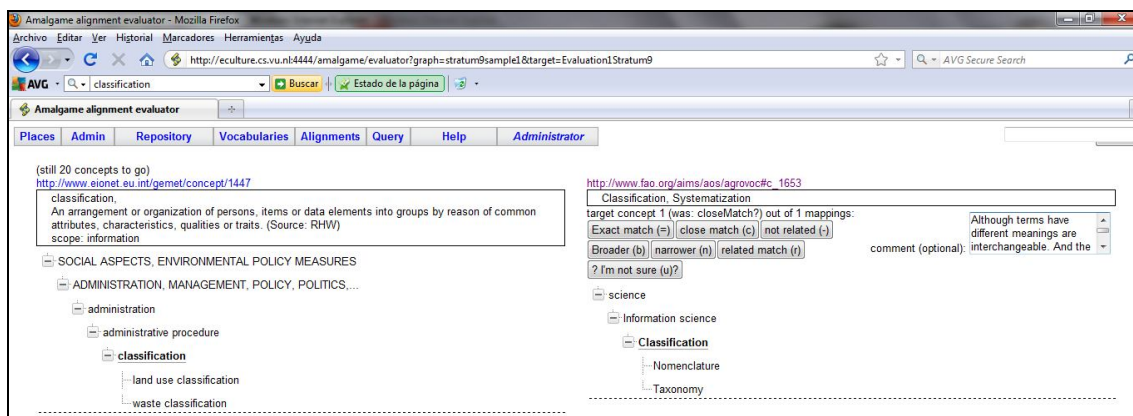
¹⁴¹ Veure AGROVOC Term Code 8994. Accessible: http://aims.fao.org/en/agrovoc-term-info?mytermcode=8994&mylang_interface=en&myLanguage=EN [Últim accés: 10/09/2013]



- **Avaluació:** alineament amb tipus de relació “exact match”. Ambdós tesaurus tenen el mateix punt de vista i comparteixen el mateix terme genèric, “erosion”. GEMET estableix també que ambdós conceptes tenen una relació de tipus “exact match”¹⁴².

Stratum 9: Mostra aleatòria

9.1 Alineament 1: classification versus classification: close match



- **Avaluació:** alineament amb tipus de relació “close match”. Ambdós conceptes tenen el mateix significat i definició (agrupació d’informació), l’abast temàtic són les “ciències de la informació”. Però cada tesaurus usa el terme amb punts de vista diferents, i és per això que s’han identificat com a tipus de relació de “quasi-sinònims”, tot i que aquesta decisió sempre s’ha de prendre amb consens dels agents implicats amb l’avaluació de les alineacions. En aquest cas, GEMET, utilitza el terme amb un abast administratiu i geogràfic, per exemple, per categoritzar l’ús del sòl o els tipus de residus. En canvi, AGROVOC utilitza el terme amb el seu significat tradicional, procés sistemàtic per agrupar la informació

¹⁴² Veure GEMET 9303 Concept. Accessible a: <http://www.eionet.europa.eu/gemet/concept/9330> [Últim accés: 10/09/2013]

en diferents classes, tal i com els seus termes específics indiquen: taxonomies i nomenclatures.

9.2 Alineament 2: caspia sea versus caspia sea

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'Caspian Sea' is shown with its hierarchy: 'NATURAL ENVIRONMENT, ANTHROPIC ENVIRONMENT' > 'LAND (landscape, geography)' > 'land' > 'world' > 'Asia' > 'Western Asia' > 'Caspian Sea'. On the right, the AGROVOC concept 'Caspian Sea' is shown with its hierarchy: 'International inland waters' > 'Asian inland waters' > 'Caspian Sea'. The alignment is marked as an 'exact match'.

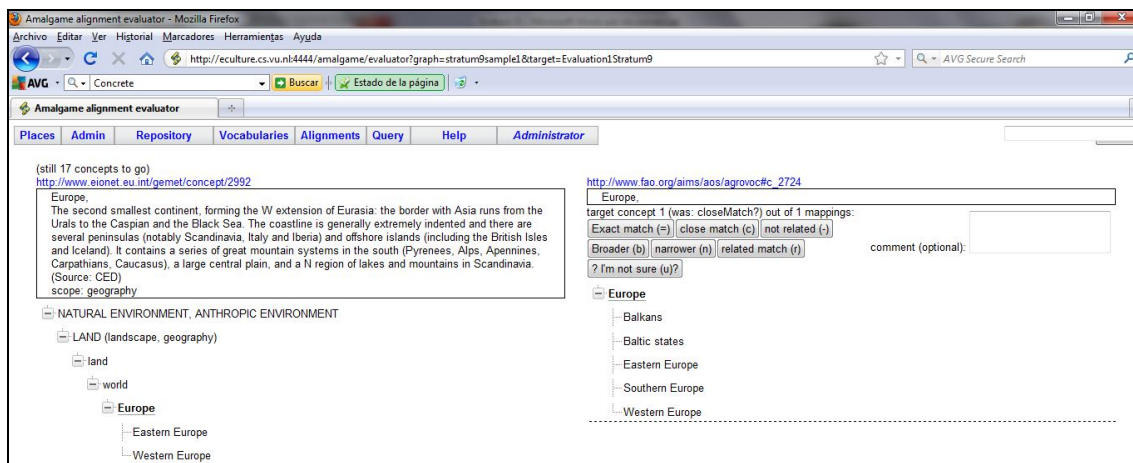
- **Avaluació:** alineament amb tipus de relació “exact match”. Ambdós tesaurus classifiquen els termes amb un terme genèric que inclou un punt de vista geogràfic (Àsia), però la jerarquia en si dels tesaurus té un punt de vista lleugerament diferent, tal i com indiquen els seus termes genèrics. GEMET té un punt de vista completament geogràfic, “Caspian sea > Western Asia > Asia”, i AGROVOC, ofereix un punt de vista geogràfic però també dóna una classificació per tipus o cossos d’aigua, “Caspian sea > Asian inland waters” > International inland waters”.

9.3 Alineament 3: concrete versus concrete

The screenshot shows the Amalgame alignment evaluator interface. On the left, the GEMET concept 'concrete' is shown with its hierarchy: 'HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE...' > 'PRODUCTS, MATERIALS' > 'material' > 'building material' > 'cement' > 'concrete'. On the right, the AGROVOC concept 'Concrete' is shown with its hierarchy: 'materials' > 'Construction materials' > 'Concrete'. The alignment is marked as an 'exact match'.

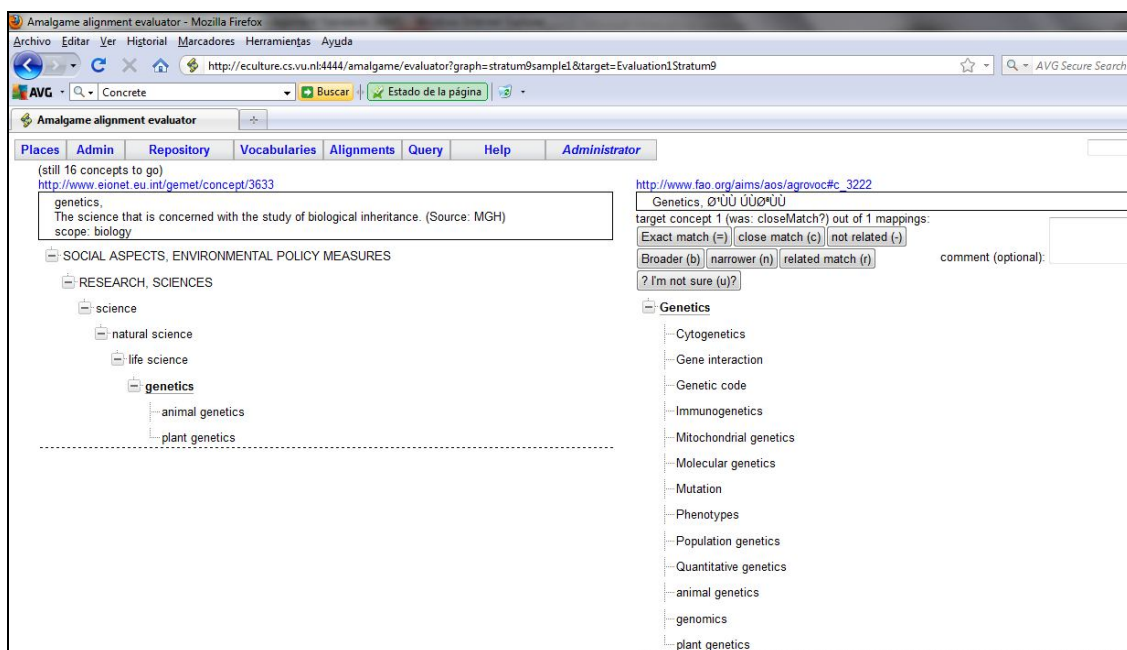
- **Avaluació:** alineament amb tipus de relació “exact match”. Ambdós tesaurus tenen el mateix punt de vista, “materials de construcció”, i comparteixen termes superiors: “building material versus construction materials” i “material versus materials”. No obstant els nivells jeràrquics d’ambdós tesaurus són diferents, com és habitual, GEMET inclou un major nivell jeràrquic superior, i inclou el terme “cement” com a terme genèric de “concrete”.

9.4 Alineament 4: Europe versus Europe



- **Avaluació:** alineament amb tipus de relació “exact match”. Ambdós conceptes tenen el mateix significat i els conceptes tenen el mateix punt de vista geogràfic. Afegir però, que les estructures jeràrquiques dels tesaurus són diferents i concretament el nivell d’especificitat és major en AGROVOC que en GEMET. Cal destacar també la diferència entre els termes genèrics que representen un abast diferent, el “broader term” en GEMET és “World” i en AGROVOC és “Eurasia”¹⁴³, i el concepte “World” en AGROVOC és considerat un terme relacionat.

9.5 Alineament 5: exact match

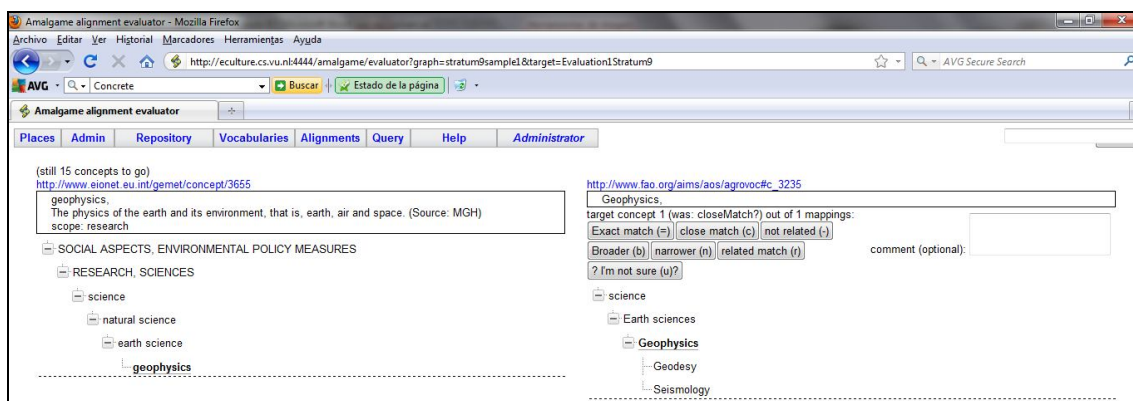


- **Avaluació:** alineament amb tipus de relació “exact match”. Ambdós conceptes tenen el mateix significat i abast temàtic, “biologia” (vegeu definició del concepte en GEMET). A més a més, el propi tesaurus GEMET identifica els termes com a conceptes amb tipus de relació

¹⁴³ Veure AGROVOC Term Code 2724. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=2724&mylang_interface=en&myLanguage=EN [Últim accés: 11/09/2013]

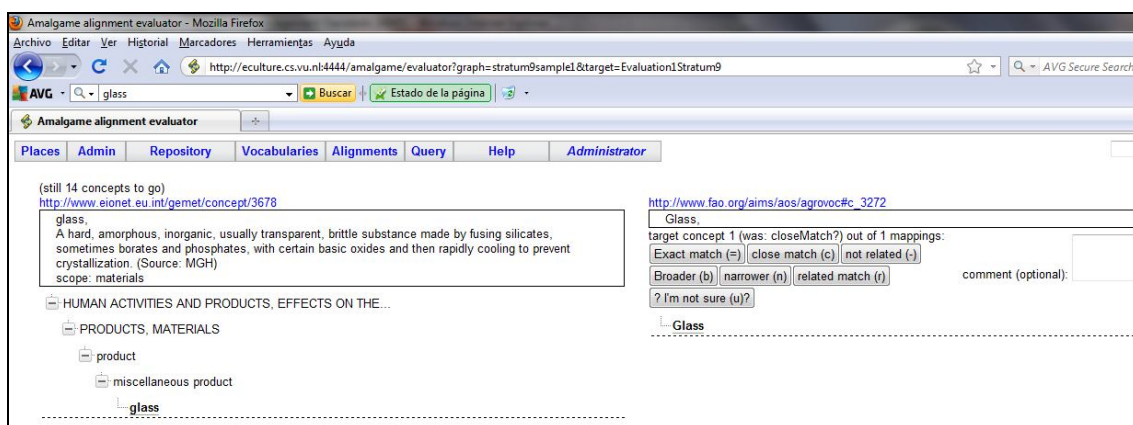
d'equivalència exacta. Els termes genèrics varien, GEMET incorpora el terme “life science” i AGROVOC “biology”¹⁴⁴, però ambdós termes tenen un tipus de relació de “broader terms”. Destacat també que en aquest cas AGROVOC incorpora un major nivell d'especificitat en comparació amb GEMET que només inclou dos termes específics.

9.6 Alineament 6: exact match



- **Avaluació:** alineament amb tipus de relació “exact match”. Ambdós conceptes tenen el mateix significat, abast temàtic i comparteixen també el mateix terme genèric “Earth Science”. La diferència principal dels dos tesaurus rau amb l'estructura jeràrquica, en el cas de GEMET s'inclou un major nombre de nivells superiors jeràrquics i pel que fa a AGROVOC la diferència està en els nivells inferiors (AGROVOC inclou termes específics i GEMET no).

9.7 Alineament 7: glass versus glass



- **Avaluació:** alineament amb tipus de relació “exact match”. Ambdós conceptes tenen el mateix significat i abast temàtic, “materials”. GEMET

¹⁴⁴ Veure AGROVOC Concept Term 3222. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=3222&mylang_interface=en&myLanguage=EN [Últim accés: 11/09/2013]

identifica també el concete “glass” d’AGROVOC com a terme equivalent¹⁴⁵.

9.8 Alineament 8: health versus health

- **Avaluació:** alineament amb tipus de relació “exact match” o relació d’equivalència exacta. Ambdós conceptes tenen el mateix significat i abat temàtic, “human health”. GEMET¹⁴⁶ identifica també el concepte “health” d’AGROVOC com a terme equivalent o tipus de relació “exact match”. En aquest cas, el tesaurus GEMET és el que té una major especificitat quant al concepte analitzat, però no en canvi el la versió actual del tesaurus AGROVOC¹⁴⁷, la qual també ofereix un major nombre de termes associatius.

9.9 Alineament 9: hydrobiology versus hydrobiology

¹⁴⁵ Veure GEMET 3678 Concept. Accessible a: <http://www.eionet.europa.eu/gemet/concept/3678> [Últim accés: 11/09/2013]

¹⁴⁶ Veure GEMET 3678 Concept. Accessible a: <http://www.eionet.europa.eu/gemet/concept/3865> [Últim accés: 11/09/2013]

¹⁴⁷ Veure AGROVOC Term Code 3511. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=3511&mylang_interface=en&myLanguage=EN [Últim accés: 11/09/2013]

- **Avaluació:** alineament amb tipus de relació “exact match” o relació d’equivalència exacta. Ambdós conceptes tenen el mateix significat, abast temàtic (estudi d’organismes vius que viuen en l’aigua) i també comparteixen el terme genèric, “biology”. A més a més, el propi tesaurus GEMET, en la seva versió online (v.3.1, 2012-07-20), exposa que són termes amb relació d’“exact match”.

9.10 Alineament 10: ivory versus ivory

The screenshot shows the Amalgame alignment evaluator interface. On the left, a concept from GEMET is displayed: 'ivory' with the description 'The fine-grained creamy-white dentine forming the tusks of elephants, and the teeth or tusks of certain other large animals such as the walrus; it has long been esteemed for a wide variety of ornamental articles. (Source: EBJGEO)'. Below this is a hierarchical tree structure: 'HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE...' -> 'PRODUCTS, MATERIALS' -> 'material' -> 'natural material' -> 'animal product' -> 'ivory'. On the right, a target concept from FAO is shown: 'ivory' with the URL 'http://www.fao.org/aims/aos/agrovoc#c_16373'. Below this is a hierarchical tree structure: 'Products' -> 'Animal products' -> 'ivory'. The interface includes a search bar with 'glass' and a navigation menu with 'Places', 'Admin', 'Repository', 'Vocabularies', 'Alignments', 'Query', 'Help', and 'Administrator'.

- **Avaluació:** alineament amb tipus de relació “exact match” o relació d’equivalència exacta. Ambdós conceptes tenen el mateix significat, abast temàtic (material/producte) i també comparteixen el terme genèric, “animal product”. A més a més, el propi tesaurus GEMET, en la seva versió online (v.3.1, 2012-07-20), exposa que són termes amb relació d’equivalència “exact match”.

9.11 lineament 11: milk versus milk

The screenshot shows the Amalgame alignment evaluator interface. On the left, a concept from GEMET is displayed: 'milk' with the description 'The whitish fluid secreted by the mammary gland for the nourishment of the young; composed of carbohydrates, proteins, fats, mineral salts, vitamins, and antibodies. (Source: MGH)'. Below this is a hierarchical tree structure: 'HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE...' -> 'PRODUCTS, MATERIALS' -> 'product' -> 'foodstuff' -> 'dairy product' -> 'milk'. On the right, a target concept from FAO is shown: 'Milk, Raw milk' with the URL 'http://www.fao.org/aims/aos/agrovoc#c_4826'. Below this is a hierarchical tree structure: 'Milk' -> 'Buffalo milk' -> 'Camel milk' -> 'Colostrum' -> 'Cow milk' -> 'Ewe milk' -> 'Goat milk' -> 'Human milk' -> 'Mare milk' -> 'Milk fat'. The interface includes a search bar with 'glass' and a navigation menu with 'Places', 'Admin', 'Repository', 'Vocabularies', 'Alignments', 'Query', 'Help', and 'Administrator'.

- **Avaluació:** alineament amb tipus de relació “exact match” o relació d’equivalència exacta. En el propi tesaurus GEMET, en la seva versió

online (v.3.1, 2012-07-20), exposa que són termes amb relació d'equivalència “exact match”. No obstant això, els termes genèrics són diferents, el terme genèric en GEMET és “dairy products”, i en AGROVOC apareixen dos termes genèrics (cas de polijerarquia)¹⁴⁸, “body fluids” i “animal products”.

9.12 Alineament 12: natural gas versus natural gas

The screenshot shows the Amalgame alignment evaluator interface. The search bar contains 'glass'. The main content area displays an alignment between two concepts. On the left, the source concept is 'natural gas' from GEMET (concept 5503), with a description: 'A natural fuel containing methane and hydrocarbons that occurs in certain geologic formations. (Source: LANDY)'. On the right, the target concept is 'Natural gas, Lp-gas (liquid petroleum gas)' from AGROVOC (concept 5087). The alignment is marked as 'Exact match (=)'. Below the alignment, there are options for 'Broader (b)', 'narrower (n)', 'related match (r)', and 'comment (optional)'. Hierarchical trees for both vocabularies are visible below the descriptions.

- **Avaluació:** alineament amb tipus de relació “exact match” o relació d'equivalència exacta. Ambdós conceptes tenen el mateix significat i abast temàtic, recurs de tipus combustible. Recalcar, que els termes genèrics són diferents, “raw material versus fuels”, però com ve s'indica en la web del tesaurus GEMET¹⁴⁹ són termes amb equivalència exacta.

9.13 Alineament 13: phosphorus versus phosphorus

The screenshot shows the Amalgame alignment evaluator interface. The search bar contains 'phosphorus'. The main content area displays an alignment between two concepts. On the left, the source concept is 'phosphorus' from GEMET (concept 6196), with a description: 'A nonmetallic element used to manufacture phosphoric acid, in phosphor bronzes, incendiaries, pyrotechnics, matches, and rat poisons; the white or yellow allotrope is a soft waxy solid, soluble in carbon disulfide, insoluble in water and alcohol, and is poisonous and self-igniting in air; the red allotrope is an amorphous powder, insoluble in all solvents and is nonpoisonous; the black allotrope comprises lustrous crystals similar to graphite, and is insoluble in most solvents. (Source: MGH)'. On the right, the target concept is 'Phosphorus, P (symbol)' from AGROVOC (concept 5804). The alignment is marked as 'Exact match (=)'. Below the alignment, there are options for 'Broader (b)', 'narrower (n)', 'related match (r)', and 'comment (optional)'. Hierarchical trees for both vocabularies are visible below the descriptions.

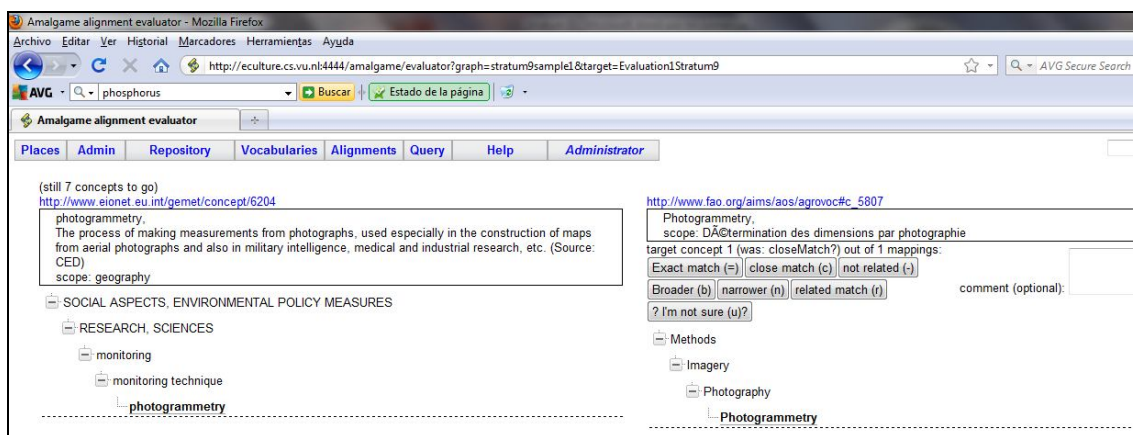
- **Avaluació:** alineament amb tipus de relació “exact match” o relació d'equivalència exacta. Ambdós conceptes tenen el mateix significat i abast temàtic, ‘elements químics’. Els seus termes genèrics són diferents, “element of group versus nonmetals”, però no hi ha discussió

¹⁴⁸ Veure AGROVOC Term Code 4826. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=4826&mylang_interface=en&myLanguage=EN [Últim accés: 11/09/2013]

¹⁴⁹ Veure GEMET 5503 Concept. Accessible a: <http://www.eionet.europa.eu/gemet/concept/5503> [Últim accés: 11/09/2013]

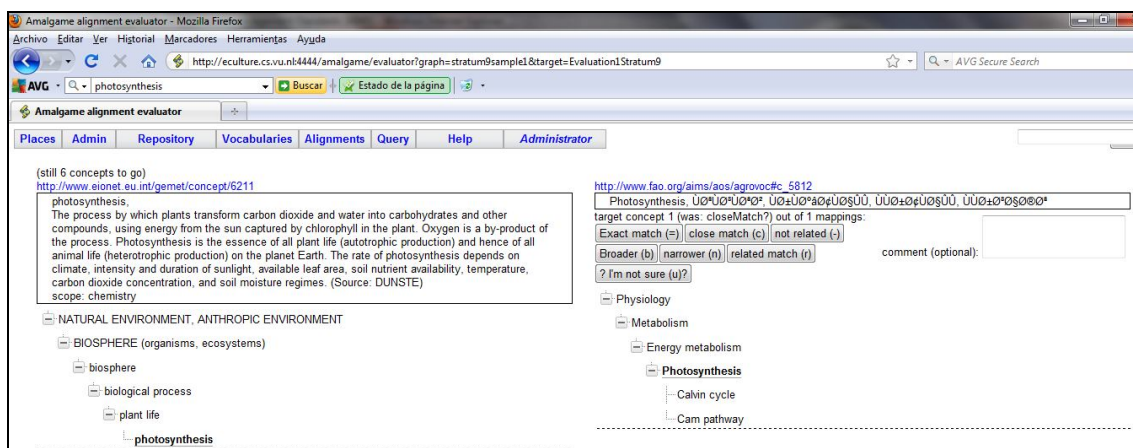
de la seva relació a nivell d' "exact match", tal i com es pot comprovar en la versió 3.1 del tesaurus GEMET (2012-07-20)¹⁵⁰.

9.14 Alineament 14: photogrammetry versus photogrammetry



- **Avaluació:** alineament amb tipus de relació "exact match" o relació d'equivalència exacta. Ambdós conceptes tenen el mateix significat i abast temàtic (procés o mètode de mestura fotogràfica). No obstant això, la jerarquia superior i els seus termes genèrics són diferents entre els tesaurus, "monitoring technique versus photography". Com en d'altres casos, GEMET estableix també que els termes en ambdós tesaurus tenen una relació d'equivalència exacta (v.3.1, 2012-07-20)¹⁵¹.

9.15 lineament 15: photosynthesis versus photosynthesis



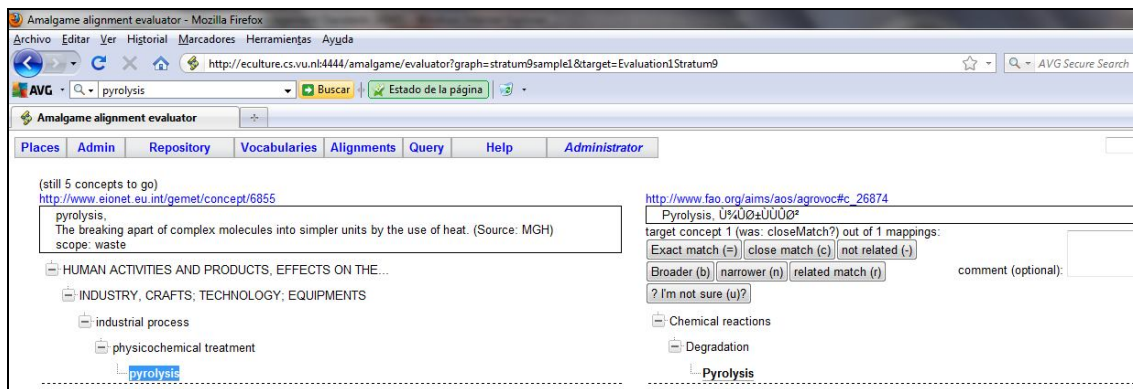
- **Avaluació:** alineament amb tipus de relació "exact match" o relació d'equivalència exacta. Ambdós conceptes tenen el mateix significat, tot i que la jerarquia superior i els seus termes genèrics són diferents entre els tesaurus ("plant life versus energy metabolism"). GEMET, identifica "fotosíntesis" com un procés biològic i químic que forma part de la vida

¹⁵⁰ Veure GEMET 6196 Concept. Accessible a: <http://www.eionet.europa.eu/gemet/concept/6196> [Últim accés: 11/09/2013]

¹⁵¹ Veure GEMET 6240 Concept. Accessible a: <http://www.eionet.europa.eu/gemet/concept/6240> [Últim accés: 11/09/2013]

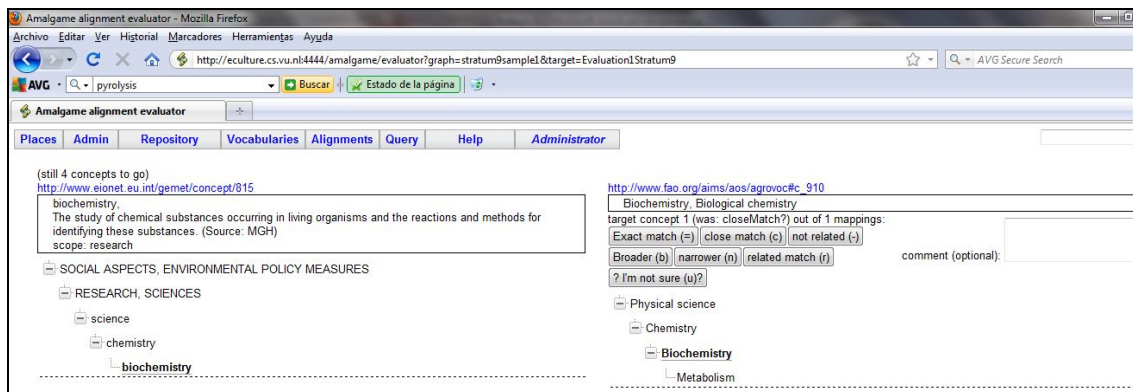
de les plantes, i AGROVOC ho representa com un procés fisiològic o metabòlic de les plantes (AGROVOC). Es conclou però que s'està descrivint el mateix tipus de procés, veure GEMET (v.3.1, 2012-07-20)¹⁵² on descriu que ambdós termes tenen una relació d'equivalència exacta.

9.16 Alineament 16: pyrolysis versus pyrolysis



- **Avaluació:** alineament amb tipus de relació “exact match” o relació d'equivalència exacta. La representació jeràrquica dels termes en cadascun dels tesaurus descriu els termes de formes diferents (*physicochemical treatment versus degradation*), no obstant, el terme té el mateix significat (consulta el concepte nº6855 de GEMET)¹⁵³.

9.17 Alineament 17: biochemistry versus biochemistry



- **Avaluació:** alineament amb tipus de relació “exact match” o relació d'equivalència exacta. Ambdós conceptes tenen el mateix significat, abast temàtic i també comparteixen el terme genèric, “chemistry”. A més a més, el propi tesaurus GEMET, en la seva versió online (v.3.1, 2012-07-20), exposa també que són termes amb relació d'“exact match”.

¹⁵² Veure GEMET 6211 Concept. Accessible a: <http://www.eionet.europa.eu/gemet/concept/6211> [Últim accés: 11/09/2013]

¹⁵³ Veure GEMET 6855 Concept. Url: <http://www.eionet.europa.eu/gemet/concept/6855> [Últim accés: 11/09/2013]

9.18 Alineament 18: thallium *versus* thallium

The screenshot shows the Amalgame alignment evaluator interface. The search term is 'thallium'. The interface displays two concepts being compared:

- Source Concept:** <http://www.eionet.eu.int/gemet/concept/8428>
thallium.
Bluish-white metal with tinlike malleability, but a little softer; used in alloys. (Source: MGH)
scope: chemistry
- Target Concept:** http://www.fao.org/aims/aos/agrovoc#c_7702
Thallium, Tl (symbol)

The alignment is marked as an "Exact match (=)". The interface also shows hierarchical taxonomies for both concepts, with 'thallium' and 'Thallium' at the bottom of their respective trees.

- **Avaluació:** alineament amb tipus de relació “exact match” o relació d’equivalència exacta. Ambdós conceptes tenen el mateix significat i abast temàtic, ‘elements químics’. Els seus termes genèrics són diferents, “element of group III *versus* metallic elements”, però no hi ha discussió, la seva relació és d’equivalència exacta, tal i com mostra GEMET en la seva versió online (v.3.1, 2012-07-20)¹⁵⁴.

9.19 Alineament 19: unemployment *versus* unemployment

The screenshot shows the Amalgame alignment evaluator interface. The search term is 'unemployment'. The interface displays two concepts being compared:

- Source Concept:** <http://www.eionet.eu.int/gemet/concept/8768>
unemployment.
The condition of being without remunerative employment. (Source: CED)
scope: economics
- Target Concept:** http://www.fao.org/aims/aos/agrovoc#c_8061
Unemployment. Structural unemployment

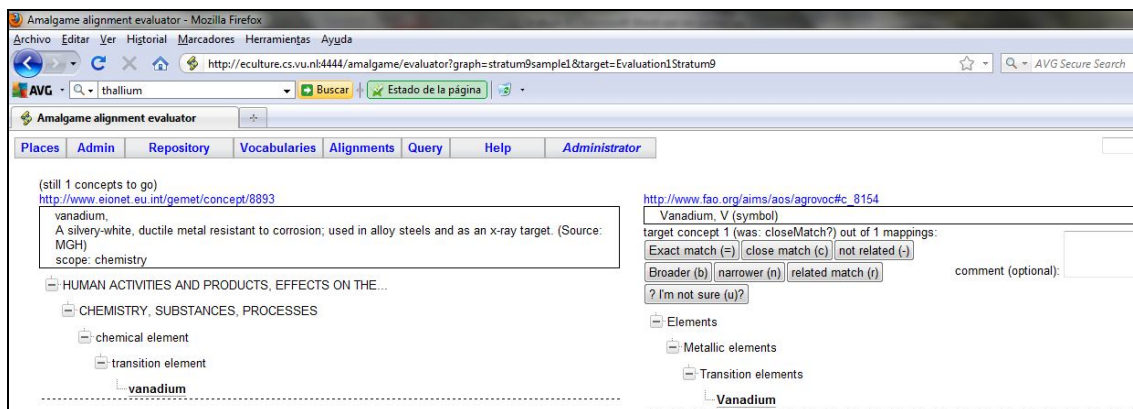
The alignment is marked as an "Exact match (=)". The interface also shows hierarchical taxonomies for both concepts, with 'unemployment' and 'Unemployment' at the bottom of their respective trees.

- **Avaluació:** alineament amb tipus de relació “exact match” o relació d’equivalència exacta. Ambdós conceptes tenen el mateix significat, abast temàtic i també comparteixen el terme genèric, “employment”, i el terme superior d’aquest, “labour”. Al mateix temps també són identificats com termes amb equivalència exacta per GEMET¹⁵⁵.

9.20 Alineament 20: vanadium *versus* vanadium

¹⁵⁴ Veure GEMET 8428 Concept. Url: <http://www.eionet.europa.eu/gemet/concept/8428> [Últim accés: 11/09/2013]

¹⁵⁵ Veure GEMET 8768 Concept. Url: <http://www.eionet.europa.eu/gemet/concept/8768> [Últim accés: 11/09/2013]

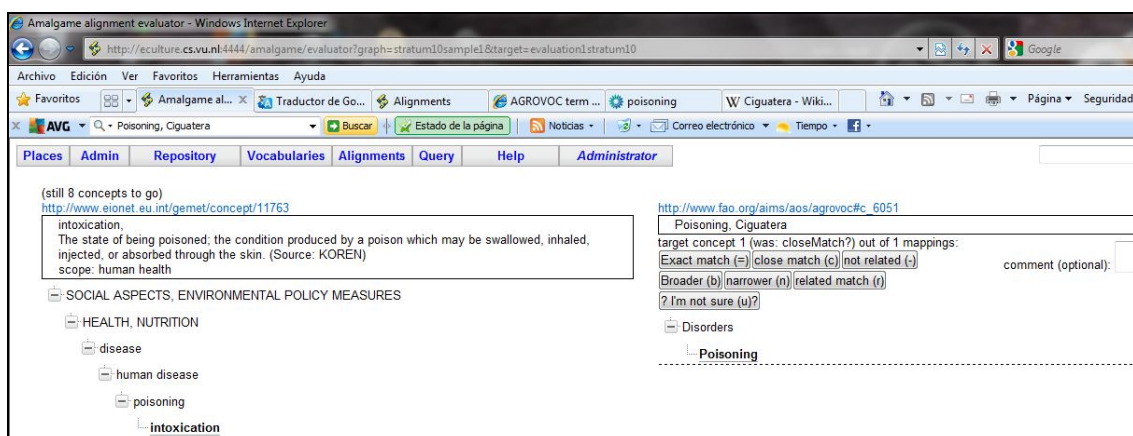


- **Avaluació:** alineament amb tipus de relació “exact match” o relació d’equivalència exacta. Ambdós conceptes tenen el mateix significat, abast temàtic (processos o elements químics) i també comparteixen el terme genèric, “transició elements”. I també trobem que en la descripció del terme en GEMET¹⁵⁶ s’estableix una relació d’equivalència exacta en el terme representat per part d’AGROVOC.

Stratum 10: Mostra aleatòria

Aquest estrat és una mostra aleatòria de només 8 alineacions, tot i que només es poden exposar la meitat dels casos, per qüestions tècniques (no s’han pogut recuperar la totalitat dels casos), no obstant, si sabem els tipus de relacions que es van establir. Resumint, en aquesta avaluació s’estableix les relacions d’equivalència següents: 3 relacions de tipus “exact match”, 3 relacions de tipus “close match”, 3 relacions de tipus “Broater relation” i 1 relació de tipus “related” (termes relacionats o RT).

10.1 Alineament 1: intoxication versus poisoning



¹⁵⁶ Veure GEMET 8893 Concept. Url: <http://www.eionet.europa.eu/gemet/concept/8893> [Últim accés: 11/09/2013]

- **Avaluació:** Relació de tipus “Close Match”. La desició es pren de la base que per al tesaurus AGROVOC “intoxication” és un terme “no-descriptor”¹⁵⁷ i que és substituït pel concepte “Poisoning” o envenenament, per això es decideix establir un tipus de relació de quasi-sinònims (conceptes diferents que s’empren amb el mateix objectiu, etc.). Un dels exemples que ens aporta AGROVOC, és un no-descriptor com “ciguatera”, que és un tipus d’intoxicació alimentària.

10.2 Alineament 2: democracy versus federations

The screenshot shows the Amalgame alignment evaluator interface. On the left, the concept 'democracy' is selected, with its definition: 'A system of governance in which ultimate authority power is vested in the people and exercised directly by them or by their freely elected agents. (Source: APS)'. Below the definition is a hierarchical tree structure showing 'democracy' as a sub-concept of 'political organisation'. On the right, the target concept 'Political systems. Democracy' is selected, with its definition: 'Political systems. Democracy'. Below the definition is a hierarchical tree structure showing 'Political systems' as a parent concept, with 'Democracy' as a sub-concept. The interface also displays a 'target concept 1 (was: closeMatch?) out of 1 mappings:' section with options for 'Exact match (=)', 'close match (c)', 'not related (-)', 'Broader (b)', 'narrower (n)', 'related match (r)', and '? I'm not sure (u)?'. The 'close match (c)' option is selected.

- **Avaluació:** Relació de tipus “close match” o quasi-sinònims. S’identifiquen com a termes quasi-sinònims perquè AGROVOC identifica el concepte “democracy” com un no-descriptor, i el terme acceptat per AGROVOC és “political systems”. El fet de que tinguin connotacions diferents en altres contextos, fa que no es seleccioni una relació de tipus “exact match”. De fet, ‘democràcia’ pot entendre’s com a una doctrina política, però AGROVOC ho entén com a un ‘sistema polític’, i GEMET, ho descriu com una ‘Organització Política’.

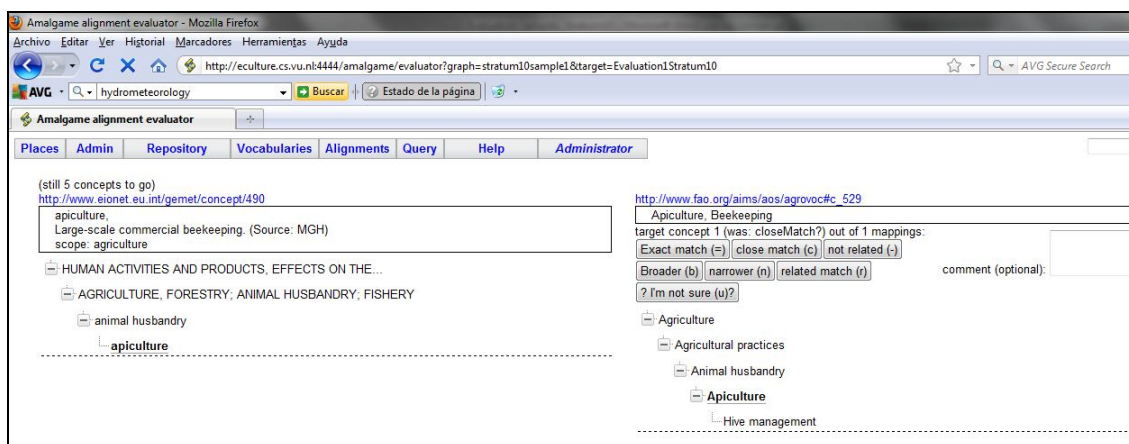
10.3 Alineament 3: hydrometeorology versus precipitation

The screenshot shows the Amalgame alignment evaluator interface. On the left, the concept 'hydrometeorology' is selected, with its definition: 'That part of meteorology of direct concern to hydrologic problems, particularly to flood control, hydroelectric power, irrigation, and similar fields of engineering and water resource. (Source: ZINZAN)'. Below the definition is a hierarchical tree structure showing 'hydrometeorology' as a sub-concept of 'meteorology'. On the right, the target concept 'Precipitation, Amount of precipitation' is selected, with its definition: 'Precipitation, Amount of precipitation'. Below the definition is a hierarchical tree structure showing 'Precipitation' as a sub-concept of 'Meteorological factors'. The interface also displays a 'target concept 1 (was: closeMatch?) out of 1 mappings:' section with options for 'Exact match (=)', 'close match (c)', 'not related (-)', 'Broader (b)', 'narrower (n)', 'related match (r)', and '? I'm not sure (u)?'. The 'close match (c)' option is selected.

¹⁵⁷ Veure AGROVOC Term Code 11819. Url: http://aims.fao.org/en/agrovoc-term-info?mytermcode=11819&mylang_interface=en&myLanguage=EN [Últim accés: 11/09/2013]

- **Avaluació:** Relació d'equivalència associativa o de tipus "related match". El concepte "precipitation"¹⁵⁸ no apareix en GEMET, en canvi sí apareix en AGROVOC, i aquest el descriu com un terme relacionat a "hydrometeorology"¹⁵⁹, conseqüentment, s'estableix un alineament de tipus relacionat.

10.4 Alineament 4: apiculture versus apiculture



- **Avaluació:** alineament amb tipus de relació "exact match" o relació d'equivalència exacta. Ambdós conceptes tenen el mateix significat, abast temàtic (agricultura) i també comparteixen el terme genèric, "animal husbandry". A més a més, el propi tesaurus GEMET¹⁶⁰, en la seva versió online (v.3.1, 2012-07-20), exposa que són termes amb relació d'"exact match".

¹⁵⁸ AGROVOC Term Code 6161. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=6161&mylang_interface=en&myLanguage=EN [Últim accés: 11/09/2013]

¹⁵⁹ AGROVOC Term Code 3733. Accessible a: http://aims.fao.org/en/agrovoc-term-info?mytermcode=3733&mylang_interface=en&myLanguage=EN [Últim accés: 11/09/2013]

¹⁶⁰ GEMET Concept Term 490. Accessible: <http://www.eionet.europa.eu/gemet/concept/490> [Últim accés: 11/09/2013]

Annex 8.4. Anàlisi qualitativa de l'experiment d'alineament de tesaurus multilingües (Avaluador 2, anàlisi de 6 estrats o *stratums*)

Estrats	PrefLabel1 (GEMET)	PrefLabel2 (AGROVOC)	MatchType
s1:14	camping site	Recreation	http://purl.org/vocabularies/amalgame/evaluator#unrelated
s1:14	drainage system	Watersheds	http://www.w3.org/2004/02/skos/core#relatedMatch
s1:14	rural law	Agricultural and rural legislation	http://www.w3.org/2004/02/skos/core#broadMatch
s1:14	justice	Justicia	http://purl.org/vocabularies/amalgame/evaluator#unrelated
s1:14	prescription	Common law	http://www.w3.org/2004/02/skos/core#relatedMatch
s1:14	underground disposal	Runoff	http://purl.org/vocabularies/amalgame/evaluator#unrelated
s1:14	approval	Licences	http://www.w3.org/2004/02/skos/core#relatedMatch
s1:14	surface active compound	Surfactants	http://www.w3.org/2004/02/skos/core#exactMatch
s1:14	longitude	Length	http://purl.org/vocabularies/amalgame/evaluator#unrelated
s1:14	services providing company	Service industry	http://www.w3.org/2004/02/skos/core#closeMatch
s1:14	coal	Carbon	http://www.w3.org/2004/02/skos/core#relatedMatch
s1:14	draught animal	Working animals	http://www.w3.org/2004/02/skos/core#broadMatch
s1:14	fen	Swamps	http://www.w3.org/2004/02/skos/core#relatedMatch
s1:14	financial assistance	Support measures	http://www.w3.org/2004/02/skos/core#broadMatch

s1:14	firing	Burners	http://www.w3.org/2004/02/skos/core#relatedMatch
s1:14	flavouring	Flavour compounds	http://www.w3.org/2004/02/skos/core#exactMatch
s1:14	flue gas	Smokes	http://purl.org/vocabularies/amalgame/evaluator#unrelated
s1:14	forest management	Forestry	http://www.w3.org/2004/02/skos/core#broadMatch
s1:14	gastropod	Snails	http://www.w3.org/2004/02/skos/core#narrowMatch
s1:14	incineration	Burning	http://www.w3.org/2004/02/skos/core#broadMatch
s1:14	intermediate goods	Prepared foods	http://purl.org/vocabularies/amalgame/evaluator#unrelated
s1:14	land planning	Regional planning	http://www.w3.org/2004/02/skos/core#broadMatch
s1:14	landscape management	Landscaping	http://www.w3.org/2004/02/skos/core#narrowMatch
s1:14	manure	Organic fertilizers	http://www.w3.org/2004/02/skos/core#closeMatch
s1:14	arable farming	Tillage	http://www.w3.org/2004/02/skos/core#narrowMatch
s1:14	mixing	Agitating	http://www.w3.org/2004/02/skos/core#closeMatch
s1:14	monitoring system	Monitoring	http://www.w3.org/2004/02/skos/core#broadMatch
s1:14	moor	Bogs	http://www.w3.org/2004/02/skos/core#relatedMatch
s1:14	natural science	Biology	http://www.w3.org/2004/02/skos/core#narrowMatch
s1:14	North Pacific Ocean	North Pacific	http://www.w3.org/2004/02/skos/core#exactMatch

s1:14	nutrient cycle	Cycling	http://www.w3.org/2004/02/skos/core#broadMatch
s1:14	organic substance	Organic matter	http://www.w3.org/2004/02/skos/core#exactMatch
s1:14	overwintering	Winterization	http://purl.org/vocabularies/amalgame/evaluator#unrelated
s1:14	soil improvement	Soil structure	http://www.w3.org/2004/02/skos/core#relatedMatch
s1:14	stock (biological)	Cuttings	http://purl.org/vocabularies/amalgame/evaluator#unrelated
s1:14	stone	Rock	http://www.w3.org/2004/02/skos/core#relatedMatch
s1:14	biological attribute	Biological properties	http://www.w3.org/2004/02/skos/core#exactMatch
s1:14	tree nursery	Plant nurseries	http://www.w3.org/2004/02/skos/core#closeMatch
s1:14	undertaking business	business enterprises	http://purl.org/vocabularies/amalgame/evaluator#unrelated
s1:14	watershed	Watersheds	http://www.w3.org/2004/02/skos/core#exactMatch
s2:14	public function	civil service	http://www.w3.org/2004/02/skos/core#relatedMatch
s2:14	devolution	Devolution	http://www.w3.org/2004/02/skos/core#exactMatch
s2:14	cultivation system	Cropping systems	http://www.w3.org/2004/02/skos/core#exactMatch
s2:14	agricultural exploitation	Farms	http://www.w3.org/2004/02/skos/core#narrowMatch
s2:14	digester	Digesters	http://www.w3.org/2004/02/skos/core#exactMatch
s2:14	ecotype	Ecotypes	http://www.w3.org/2004/02/skos/core#exactMatch

s2:14	forest conservation	Forest protection	http://www.w3.org/2004/02/skos/core#closeMatch
s2:14	incineration	Combustion	http://www.w3.org/2004/02/skos/core#relatedMatch
s2:14	animal husbandry	Animal breeding	http://www.w3.org/2004/02/skos/core#narrowMatch
s2:14	information service	Information services	http://www.w3.org/2004/02/skos/core#exactMatch
s2:14	low flow	Low water	http://purl.org/vocabularies/amalgame/evaluator#unsure
s2:14	pheromone	Pheromones	http://www.w3.org/2004/02/skos/core#exactMatch
s2:14	photograph	Photography	http://www.w3.org/2004/02/skos/core#relatedMatch
s2:14	polder	Reclaimed land	http://www.w3.org/2004/02/skos/core#broadMatch
s2:14	pollution	Contamination	http://www.w3.org/2004/02/skos/core#narrowMatch
s2:14	slaughtering of animals	Slaughtering	http://www.w3.org/2004/02/skos/core#exactMatch
s2:14	poriferan	Sponges	http://www.w3.org/2004/02/skos/core#relatedMatch
s2:14	tanker (ship)	Tanker ships	http://www.w3.org/2004/02/skos/core#exactMatch
s2:14	biosphere reserve	Nature reserves	http://www.w3.org/2004/02/skos/core#closeMatch
s2:14	wood	Forests	http://www.w3.org/2004/02/skos/core#broadMatch
s3:14	biological cycle	Life cycle	http://www.w3.org/2004/02/skos/core#narrowMatch
s3:14	chemical pollution	Chemical contamination	http://www.w3.org/2004/02/skos/core#closeMatch

s3:14	vector	Vectors	http://purl.org/vocabularies/amalgame/evaluator#unrelated
s3:14	controlled burning	Controlled burning	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	agricultural equipment	Farm equipment	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	finances	Finance	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	food colourant	Food colourants	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	alloy	Alloys	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	garden waste	Garden wastes	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	water power	Water power	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	land reclamation	Reclamation	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	mill	Mills	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	mycorrhiza	Mycorrhizae	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	nitrosamine	Nitrosamines	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	ornithology	ornithology	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	percolating water	Infiltration water	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	reflectometry	Reflectometry	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	seed (biology)	Seeds	http://www.w3.org/2004/02/skos/core#exactMatch

s3:14	shipbuilding	Shipbuilding	http://www.w3.org/2004/02/skos/core#exactMatch
s3:14	adhesive	Adhesives	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	demand	Demand	http://www.w3.org/2004/02/skos/core#closeMatch
s4:13	economics	Economics	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	environmental monitoring	Environmental monitoring	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	geographic information system	Geographical information systems	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	hurricane	Cyclones	http://www.w3.org/2004/02/skos/core#broadMatch
s4:13	hydraulics	Fluid mechanics	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	inland fishery	Inland fisheries	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	liquid waste	Liquid wastes	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	marginal land	Marginal land	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	microcomputer	Microcomputers	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	plain	Plains	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	pollutant emission	Pollutant emission	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	pollution load	Pollutant load	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	polymer	Polymers	http://www.w3.org/2004/02/skos/core#exactMatch

s4:13	avalanche	Avalanches	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	research project	Research projects	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	road	Roads	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	sulphide	Sulphides	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	surplus	Surpluses	http://www.w3.org/2004/02/skos/core#exactMatch
s4:13	weed	Weeds	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	arboretum	Arboreta	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	pruning	Pruning	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	vitrification	Vitrification	http://www.w3.org/2004/02/skos/core#relatedMatch
s5:13	cultural heritage	Cultural heritage	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	earthquake	Earthquakes	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	fodder	Forage	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	amine	Amines	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	hybridisation	Hybridization	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	hydrocarbon	Hydrocarbons	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	kerosene	Paraffin	http://www.w3.org/2004/02/skos/core#exactMatch

s5:13	lagoon	Lagoons	http://www.w3.org/2004/02/skos/core#closeMatch
s5:13	landscape management	Landscape conservation	http://www.w3.org/2004/02/skos/core#closeMatch
s5:13	library	Libraries	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	market economy	Market economies	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	natural heritage	Natural heritage	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	parasite	Parasites	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	policy	Policies	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	price	Prices	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	rural population	Rural population	http://www.w3.org/2004/02/skos/core#exactMatch
s5:13	waste reduction	Waste reduction	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	classification	Classification	http://www.w3.org/2004/02/skos/core#closeMatch
s9:13	Caspian Sea	Caspian Sea	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	concrete	Concrete	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	Europe	Europe	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	genetics	Genetics	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	geophysics	Geophysics	http://www.w3.org/2004/02/skos/core#exactMatch

s9:13	glass	Glass	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	health	Health	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	hydrobiology	Hydrobiology	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	ivory	Ivory	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	milk	Milk	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	natural gas	Natural gas	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	phosphorus	Phosphorus	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	photogrammetry	Photogrammetry	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	photosynthesis	Photosynthesis	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	pyrolysis	Pyrolysis	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	biochemistry	Biochemistry	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	thallium	Thallium	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	unemployment	Unemployment	http://www.w3.org/2004/02/skos/core#exactMatch
s9:13	vanadium	Vanadium	http://www.w3.org/2004/02/skos/core#exactMatch