8th NKOS Workshop, Oct 01, Corfu, Greece, 2009

The KOS interoperability in aquatic science field through mapping processes

Carmen Reverté Reverté

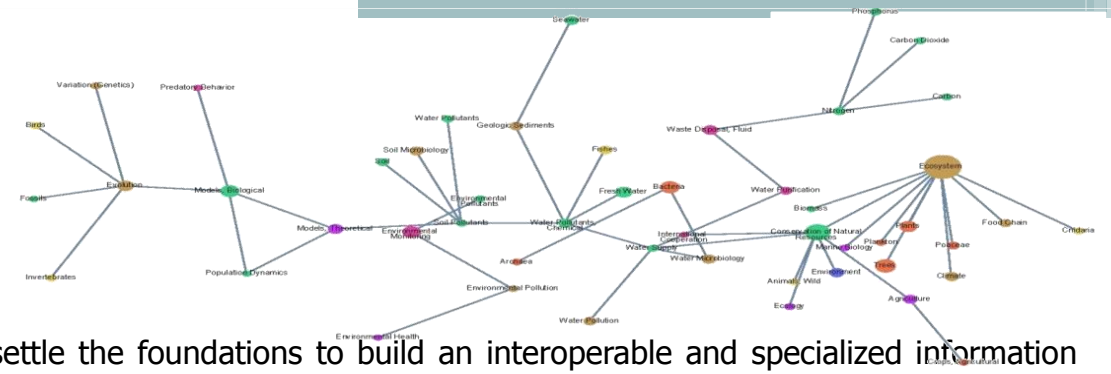
Aquatic Ecosystems Documentation Center. IRTA.

(Sant Carles de la Ràpita, Catalonia, Spain). E-mail: carme.reverte@irta.cat

Dra. Montserrat Sebastià

Information Science and Librarianship Department. University of Barcelona (Catalonia, Spain). E-mail: msebastia@ub.edu

1. Introduction



- **Objective:** The PhD project is focused on settle the foundations to build an interoperable and specialized information system to solve the main problems that they are found in aquatic science area: the Multilingualism, Multidisciplinar and Information dispersion.
- **Context:** In the Aquatic Science exists several organisms and associations which work together to share their knowledge, information exchange and information science diffusion (**IAMSLIC** & **EURASLIC** nets[1]).
  - They only have developed a professional level of cooperation: some repository projects have already developed but they doesn't enough subject coverage and **they aren't advanced information systems**.
  - **Lack of specialized information systems** which covers this heterogenic area (Agriculture and Aquaculture, Biology, Freshwater and Marine Science, Environmental Science, Ecology, Climatic change, Chemistry, etc.).
  - **Heterogeneity & Multidisciplinar** involves information dispersion on the net:

+ Information recourses
+ Information dispersion
+ Information systems without standards (nets, databases, portals, blog's, etc.)
+ use of several Knowledge Organization Systems (controlled vocabularies)
+ use of different metadata domains

- Accuracy
- Precision
- Information Quality

1. International and European Association of Aquatic and Marine Science Libraries and Information Centers (IAMSLIC)

2. Work lines: Study the Semantic Interoperability to provide the simultaneous access to different heterogenic collections

Analysis of the Subject Context Characterization → Information Resources Analysis → Controlled Vocabularies Evaluation → ***Semantic Interoperability*** analysis among aquatic systems → Study of the representative SG's → **ASESG Framework basis**

Table 1. Metadata mapping in Aquatic Science

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 $c 1996*	<DC:publisher>*SPB Academic Publishing*	**1.1.8.8.2 Publisher** *SPB Academic Publishing*
260$c (Date) *1996*	<DC:data>*1996*</>	**1.1.8.2 (Publication date)** *1996*
650$a (Subject); 650 2 / 653 (Subject); *Ex: 650.04 Freshwater plants $x Identification*	<DC:subject>*Freshwater Plants, Identification*</>	**1.6.1.1 (Theme Keyword)** *Freshwater plants – Identification*; **1.6.1.2 (T.K. thesaurus / term uncontrolled)**

Crosswalk

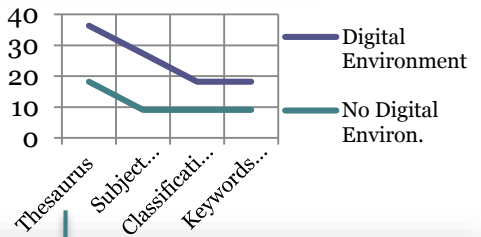
Accuracy

1.Data Structures: Mapping data structures through associative processes between data elements and structures are possible because of the use of standards (DC & MARC).

2.Categorical Data (standards): controlled vocabularies Mapping processes for the information integration is difficult because everybody use different & simultaneous vocabularies & terminologies (not always standards).

Semantic tendencies in aquatic science libraries: use of thesaurus for indexing and Information Retrieval (RI)

Important

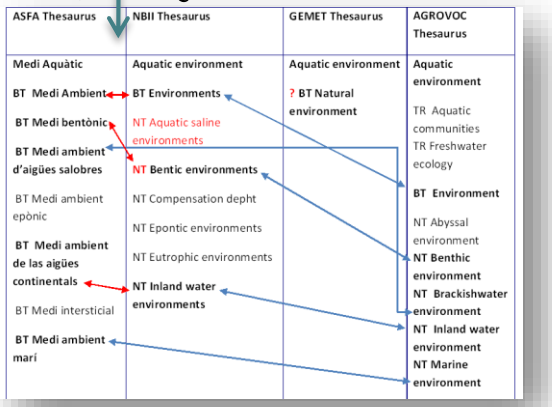


Two research lines in SI	Process	Advantages/Disadvantages	
Standardization – Proactive (everybody can share and access to the data using a common standard)	Metadata meaning and schemes for sharing the same KOS: authority controls, geographic names and common identifiers	+ Stable process in a large period of time.	- Flexible. Used in general information systems.
Interpretation – Reactive (interchange and consistent terminology to obtain a full *SI*)	Translation, mapping or correlated processes of metadata, content standards (*crosswalk*) and controlled vocabularies mapping like interpretative tools within system.	+ Flexible and selective. Used in multilingual and specialized digital information systems.	- Stable in a large period of time because of changes in research areas.

3.Factual Data (authority data): particular date. It is needed **codification rules** to identify repeated items (geographic names, authors, dates, etc.)

2.1.Categorical Data: *Standardization* is difficult because is related to user communities (culture language and ideology). Full **equivalent relations** aren't always possible

Consistency

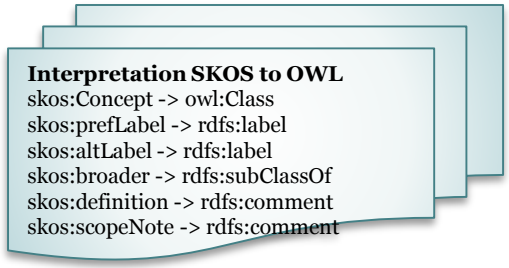


➢ Subject indexing + Subject access = search support & accuracy in Information Retrieval (**IR**).

3. Semantic Interoperability Problems & Solutions

- **Possible solution:** use of Mapping process through ontologies. The controlled vocabularies are converted to data schemes (like metadata standards):
  - Representing several controlled vocabularies in the same system
  - + Interoperable between them
  - + useful for Multilingual voacavularies
  - Use of automatic mapping processes
- **Problem:** although ontologies are represented to facilitate the information interchange in semantic web, aren't enough developed for terminological representation.
  - The most standardized models of ontologies are: OWL and SKOS
  - Aquatic ontologies tendencies: Use of SKOS format, but the automatic tools developed are for OWL.

- **Solution:** Convert Aquatic Science thesaurus from SKOS to OWL.
  - At the moment we are testing only three thesaurus in SKOS and they are located in ThManager ontology software: Agrovoc (FAO), Gemet (EEA), Unesco Thesaurus.
- Conversion tool proposed: MiklosNagy tool (OAEI-2009). It is need an structural change process:



- Mapping process tool: FALCON & RiMOM (OWL) suggested for several OAEI members, and in a recent future the ThManager tool (SKOS) when the mapping process being available.

4. Conclusions & Future lines

Conclusions

- Ensure the Semantinc Interoperability involves the use of standardization methods (metadata, ontologies and controlled vocabularies) and interpretation methods (data mapping) to provide accuracy in the systems.
- The Multidisciplinary of this field involves a deep analysis to redesign a new vocabulary.
- Build a prototype of Subject Gateway specialized in aquatic science:
  - Cover the lack of aquatic information systems and offer more quality services for researchers and aquatic science professionals.
  - Design a collaborative model for unifying methodologies among aquatic science information systems.

Future Lines

- Assure the Semantic Interoperability: doing an integration policy based on Cooperation, Coordination and Sustainability among different professionals. The partners proposed are IAMSLIC and EURASLIC nets (where there are all the aquatic science communities and governments).
- Encourage different professionals (researchers & librarianship) for the scientific information diffusion in the Subject Gateway.
- Contribute to the development from the Semantic Web to new standards and information technologies

5. Bibliography

- (c2008) Crosswalks tools in Marine Metadata Interoperability. USA, National Science Foundation. URL: http://marinemetadata.org/tools
- Caracciolo, C. (2008). *Results of the Ontology Alignment Evaluation Initiative 2008.* Ontology Alignment Evaluation Initiative (OAEI)
- DOERR, M. (2004) Semantic interoperability : Theoretical Considerations. ICS-FORTH. URL: http://www.ics.forth.gr/ftp/tech-reports/2004/2004.TR345_Semantic_Interoperability_Theoretical_Considerations.pdf
- Euzenat, J., [et al.]. (2006). *Results of the Ontology Alignment Evaluation Initiative 2006.* Ontology Alignment Evaluation Initiative (OAEI).
- Euzenat, J., [et al.]. (2007). *Results of the Ontology Alignment Evaluation Initiative 2007.* Ontology Alignment Evaluation Initiative (OAEI).
- Gasevic, D. (2005). Searching Web Resources Using ntology Mappings.
- (2004) Final European Interoperability Framework. European Communities. URL: http://ec.europa.eu/idabc/servlets/Doc?id=19529
- Lauser, B., Johannsen, G. & Caracciolo, C. (2008) Comparing Human and Autmatic Thesaurus Mapping Approaches in the Agricultural Domain. Proceedings International Conference On Dublin Core and Metadata Application Berlin.
- Nagy, Miklos. 2007. Ontology Alignment Evaluation Initiative : A note about translation from SKOS to OWL. URL: http://oaei.ontologymatching.org/2007/skos2owl.html
- ThManager tool: http://thmanager.sourceforge.net/