A Marine Platforms Ontology: Experiences and Lessons

Luis Bermudez[1], John Graybeal[1], Robert Arko[2]

[1] Monterey Bay Aquarium Research Institute
700 Sandholdt Road, Moss Landing, CA, 95039 USA
{bermudez, graybeal}@mbari.org
[2] Lamont-Doherty Earth Observatory of Columbia University
61 Route 9W, Palisades, NY 10964 USA
arko@ldeo.columbia.edu

Abstract. The community-centered creation of ontologies is an essential process to increase semantic interoperation across disciplines. The Marine Metadata Interoperability Project (MMI) is keenly aware of the need for such ontologies to advance the sophisticated interaction of research communities and their data systems. This paper presents the creation of a marine platforms ontology, fostered by MMI and jointly developed by a team of data system developers from the marine science and ontological communities.

Keywords: Ontology, Platform, Community, Experience, Marine, Lessons, Sensor.

1. Introduction

In August 2005, the Marine Metadata Interoperability Project (hereafter MMI) [6] held a workshop [7] to coordinate the mapping of science domain vocabularies. The sensor team at the workshop determined that a comprehensive ontology of sensors used by the marine science community did not exist. However, they decided that the best starting point for the creation of such ontology was the Global Change Master Directory (GCMD [15]) instrument keywords, and the related Semantic Web for Earth and Environmental Terminology (SWEET [16]) sensor ontology, which is based on the GCMD keywords. During their work, the team found that GCMD top-level categories presented a mixture of domain and measurement types; many sensors would fit into more than one class; and there is a mixture and overlap of instruments and platforms concepts. The workshop produced an initial reformulation of the GCMD/SWEET sensors list, organizing their work according to the techniques by which the various sensors operate. This provided at least a reasonably consistent basis for categorizing the major sensor groups.

Based on interest from the workshop, and needs expressed by members of the community, some members of the team undertook to continue their development of the sensor ontology. Announcements were distributed to workshop participants, MMI mailing lists, and at workshops and meetings attended by MMI representatives, and 7 participants agreed to meet to advance the project. The team discussed the best sensor ontology approach, but ran into two problems: some of the terms or classifications for sensors depended on their relationships to the platforms on which they were deployed (and no platform controlled vocabulary was available); and the best approach for pursuing a sensor ontology was uncertain. As a prerequisite to creating a sensor ontology, the team began work on a platform ontology, expecting to leverage the experience gained in that process. The following use cases, which are being considered in the MMI interoperability demonstration OOSTethys, were identified for this ontology:

- Guide a marine data repository to tag data sets with the appropriate platform tag (for example, “satellite”).
- Help a data portal discover data. (For example, when searching for remote sensing data, data tagged with the ”satellite“ term can be retrieved, since the ontology knows that "satellite" data is also remotely sensed data.)
- Help observatories categorize their platforms in a way that is consistent and interoperable.
- Help asset managers assess, model, and control their observational assets based on known characteristics and capabilities. (For example, the ontological classification includes the property ”mobility”, and so supports queries for mobile platforms.)
- Enable other domains to better categorize their platforms, in a way that is interoperable with the marine domain.
- Allow characterization of instruments to include the platforms upon which the instruments can be deployed.

The MMI served as host for the ontology development effort, providing telephone and web conferencing, a web host for notes and publication of results, and the leadership to promote accomplishment of the project. The ontology project relates closely to a number of other marine metadata activities sponsored by MMI, including variable mapping and ontology serving technologies. The project’s technical leader was the “ontological host”, updating the ontology based on the team’s discussions, publishing it in OWL and text formats, and developing many alternate and more detailed classifications for the team to consider. This was a key role in maintaining progress and the interest of the team’s members.

2. Strategy, Tools and Team Composition

The work was accomplished by exchanging ideas in a mailing list; meeting via five Web conference calls, lasting 2-3 hours each; and meeting once in person. The team was composed of 7 participants representing projects from Europe and United States (Table 1). From this group, two were ontology experts while the others were domain experts. The kickoff day was March 17, 2006, and we set a milestone to have version 1.0 beta by May 17. We used the MMI site to publish the most important discussions and resolution of our meetings and to publish the versions of the ontology in OWL as well a text file with the classification of the concepts in simple text format.

Table 1. Participants Platform Ontology Creation

Institution	*Research Project*	*Role of Participant*
Monterey Bay Aquarium Research Institute	Shore Side Data System (SSDS) http://www.mbari.org/ssds/	Domain Expert
British Oceanographic Data Centre	Sea Search http://www.sea-search.net/	Domain Expert
Lamont-Doherty Earth Observatory of Columbia University	Marine Geoscience Data System (MGDS) http://www.marine-geo.org/	Domain Expert
NASA Jet Propulsion Laboratory	Semantic Web for Earth and Environmental Terminology (SWEET) http://sweet.jpl.nasa.gov/	Ontology Expert
Natural Environmental Research Council	DataGrid http://ndg.badc.rl.ac.uk/	Domain Expert
University of Alabama in Huntsville	Information Technology and Systems Center (ITSC) http://www.itsc.uah.edu/	Domain Expert
Monterey Bay Aquarium Research Institute	Marine Metadata Initiative http://marinemetadata.org/	Ontology Expert / Facilitator

The first meeting was spent reviewing the classic “Pizza Ontology” tutorial [8]. The non-ontology experts in the team familiarized themselves with creating classes and relationships, creating properties with domains and ranges, and creating restrictions on classes, learning along the way about the tools that would be used by the team.

The team also reviewed existing platform vocabularies in the marine science community (Table 2). Most of the existing vocabularies were encoded as a plain list of terms, lacking any type of hierarchical organization and lacking any definitions of the terms. Some of these vocabularies were converted to simple ontologies to find commonalities among them, resulting in a “first draft” ontology.

The language selected to encode the ontology was the Web Ontology Language, OWL-DL[2], due to the availability of editor tools and reasoners that support this language. We used two ontology editors: Protégé [19] and Swoop [11]. We also used Pellet OWL Reasoner [10] to debug the ontology, checking that the classification was appropriate and that no problems were presented (e.g. finding inconsistent classes).

Table 2. Platform vocabularies

Vocabulary name	*URL*	*Has Categori-zation*	*Has Defini-tions*	*Available in Ontology format*
Common Data Index Terms – Sea Search Partners	http://www.bodc.ac.uk/projects/european/seasearch/data_dictionaries/	Basic	No	At the MMI site
GCMD Instrument Keywords	http://gcmd.nasa.gov/Resources/valids/sensors.html	No	No	At the MMI site
Alliance for Coastal Technologies Platform Types	http://www.act-us.info	No	No	No
SWEET Sensor Ontology *(no platform types)*	http://sweet.jpl.nasa.gov/ontology/sensor.owl	No	No	Yes
MBARI SSDS Device Types	http://ssdspub.mbari.org:8080/access/device.jsp	No	No	Yes at the MMI site

3. Description of Ontology Created

The series of versions of the ontologies created are available at: http://marinemetadata.org/platformonts. It contains 5 object properties, 0 datatype properties and approximately 150 classes.

Figure 1, shows the hierarchy of properties created, all of which had a domain *Platform*. The left side of the figure shows the first version of the properties and the right side shows the latest version. All of the platform qualifiers were moved towards concepts, as shown in Figure 2, to allow the platforms' qualifiers to be expressed in an extensible way. For instance, this allowed us to expand the concept of *mobility* in *motility*, *immotility*, *constraintMobility*. This also allowed us to create mobility types based on the availability of a crew. So, instead of saying that "*Platform isMobile true*" and "*Platform hasOccupants true*", we said "*Platform hasPlatformQuality withCrewMotility*".

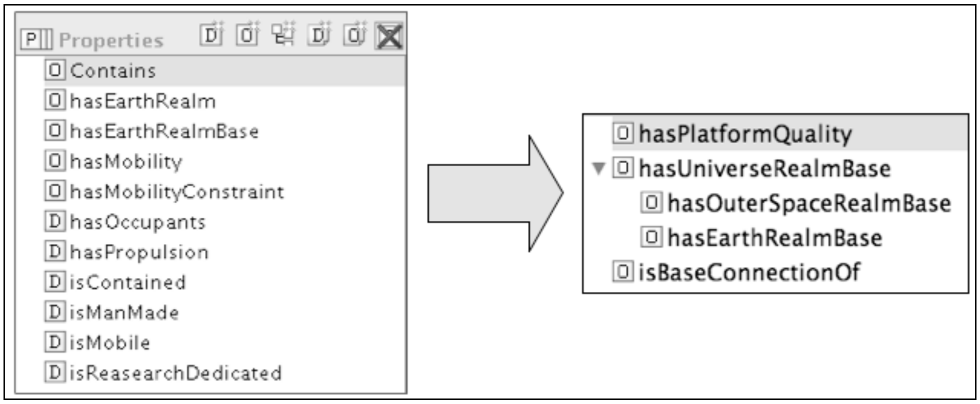

Figure 1. Properties in The Platform Ontology. Left version 2006-04-10. Right version 2006-05-14

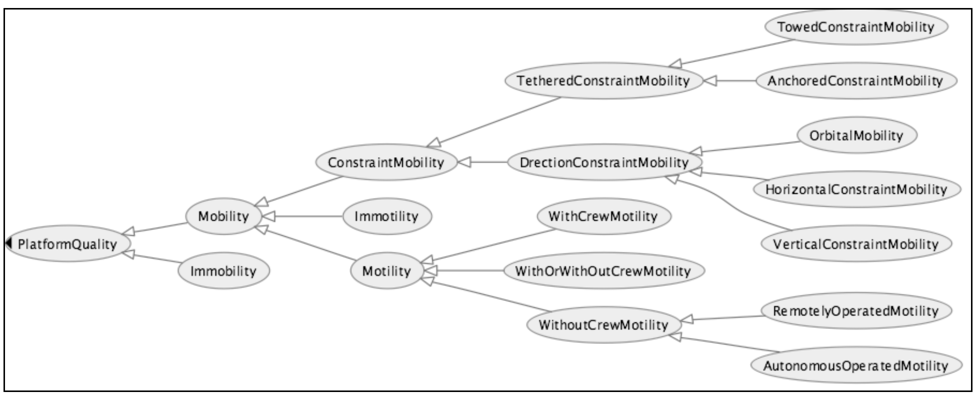

Figure 2. Hierarchy of Platform Qualities

All the platform types were define as classes. Each class was defined by creating a quantifier restriction on two main properties: *hasPlatformQuality* and *hasUniverseRealmBase*. The quantifier restrictions are composed of existential quantifiers (i.e. *someValuesFrom*) and universal quantifiers (i.e. *allValuesFrom*) [8]. Figure 3, shows an example of a definition for an Autonomous Underwater Vehicle (AUV). For a thing to be an AUV it should not be an animal, it should have no crew, it should move spontaneously and independently (*AutonomousOperatedMotility*), and it should be based in the water body subsurface. It is important to note that the set of individuals of a universal restriction also contains the "individuals that do not have any relationship along this property to any other individuals" [8]. For this reason the existential restriction (*someValuesFrom*) was also added when defining an AUV. The two additional restrictions added describe all the individuals that have at least one earth realm base *WaterBodySubsurface* and all the individuals that have at least one platform quality *AutonomousOperatedMotility*.

```
Class  (AutonomousUnderwaterVehicle partial
   Platform
   complementOf (Animal)
   restriction (hasEarthRealmBase allValuesFrom (WaterBodySubsurface))
   restriction (hasEarthRealmBase someValuesFrom (WaterBodySubsurface))
   restriction (hasPlatformQuality someValuesFrom (AutonomousOperatedMotility))
   restriction (hasPlatformQuality allValuesFrom (AutonomousOperatedMotility))
)
```

Figure 3. Definition of an Autonomous Underwater Vehicle

Three main types of platforms were defined based on the universe realm [16] where the platform is located: *AirAndOuterSpaceBasedPlatform, EarthSurfaceBasedPlatform and WaterBasedPlatform*. For example the class *WaterBasedPlatform* is defined as a platform that includes individuals that have at least one property whose values are of type *WaterBodyRealm* (See Figure 5). When running the classifier, *WaterBasedPlatform* included terms like *AUV, Boat*, *Ship*, *ResearchVessel*, *Towfish*, and *Buoy; AirAndOuterSpaceBasedPlatform* group included platforms like *AirShip*, *Airplane*, *Balloon*, *Glider*, *Helicopter*, *Kite,* and *Satellite*; and *EarthSurfaceBasedPlatform* group includes terms like *BenthicNode* and *LandCrawler*.

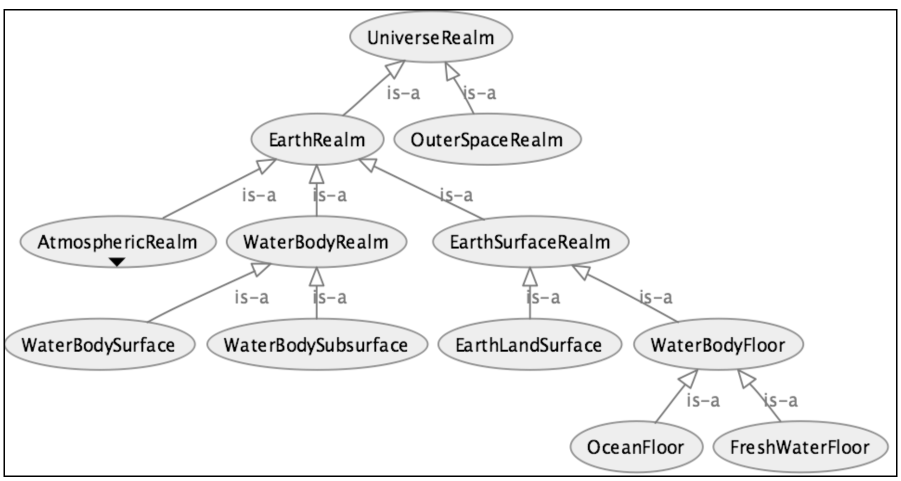

Figure 4. Universe Realms

```
Class (WaterBasedPlatform partial
   Platform
   restriction (hasEarthRealmBase someValuesFrom (WaterBodyRealm) )
)
```

Figure 5. Definition of a *WaterBasedPlatform*

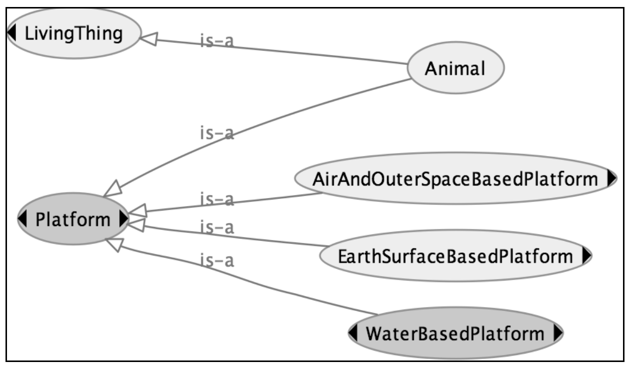

Figure 6. Root Platform Classes

4. Ontology Design Principles

4.1. Classes versus instances

Ontologies are composed of classes that represent conceptualizations. The ontology presented here describes the main terminology for marine platform types, and does not include any instance. The ontology is in OWL-DL, leveraging the ability of reasoners to classify the concepts being defined.

4.2. Name constructs

In RDF [5] based ontologies, such as OWL, every term is a resource represented by a Unique Resource Identifier (URI) [4]. Every URI in RDF is composed of a base URI and a fragment identifier, which is typically the string after the "#" character [12]. The fragment identifier is the default name of a resource displayed by most of the ontology tools. Careful attention was placed in construction of the fragment identifier string, even though we were aware that friendlier human labels can be added with the *rdfs:label* property.

The fragment identifier as part of the URI is subject to rules of character encoding [4]. For example spaces and forward slash are not allowed, and it should start with a letter or underscore. But outside of the syntax, which is checked by the ontology tools, the grammatical composition of the fragment identifiers was not straightforward to realize. Using common sense and previewing random ontologies in the Web, we came up with the following rules for naming a resource:

1) Common marine terms were preferred to logic terms. (For example ResearchVessel instead of ResearchShip, as a subclass of ship).
2) Class names should be nouns.
3) Nouns should be preceded by adjectives, following English grammatical conventions.
4) The singular form of a noun was preferred over its plural form, following Object Oriented Design Principles.
5) Camel Case was preferred, as opposed to using hyphens or underscored characters to separate two or more words inside in a fragment. (For example, ResearchVessel instead of Research_Vessel or Research-Vessel).

5. Integration with CDI Sea Search Dictionary

Ontology integration consists of bringing together two or more ontologies expressing the result of such agreement in a new ontology [9, 14]. Integration is also known as merging [13]. In the integration process, two terms from two different ontologies are related by making an equivalent, subsumption, disjoint or instance relationship among them [13]. We use the term *alignment* as the process of creating a subsumption relation between two classes and *mapping* as the process of creating an equivalent relation. [14].

The participants agreed that the ontology created (hereafter MMI-Ontology) should at least be integrate with the Common Data Index (CDI) platform terms [18] created by the European data management initiative Sea Search Partners [17]. CDI Platform dictionary is a plain list with no explicit hierarchy of terms. The MMI-Ontology created is an attempt to have a more encompassing and structured controlled vocabulary for marine platforms. The implications are that data being tagged by CDI terminology can be further discoverable using the MMI-Ontology.

To integrate the CDI platform terms with the MMI-Ontology we first created the CDI platform terms to an ontology using the tool Voc2OWL [20], which allows to create ontologies from a text file by specifying the details of the conversion in a simple form. To perform the integration, we imported the CDI-ontology into our working space in Protégé and saved the relations into the MMI-Ontology. Figure 7 presents an excerpt of the integration of Ship-related terms. The CDI-ontology classes have the prefix *cdi*, while the MMI ontology classes do not have any prefix. We also agreed that if necessary we could modify the MMI-Platform ontology to guarantee that both ontologies were consistent and coherent.

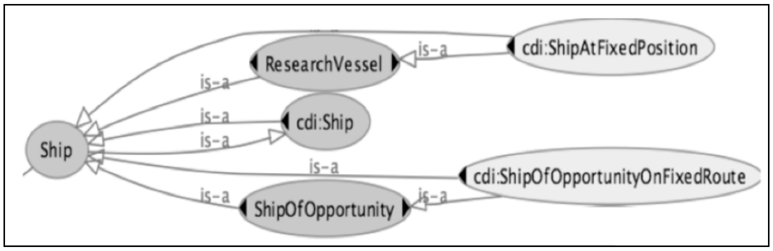

Figure 7. Integration of the Ship terms in MMI-Ontology and CDI-Ontology

6. Conclusions and Future Work

In this paper, we report our experience in creating a community-centered ontology whose domain lacks a formalized controlled vocabulary for sensor and platform concepts. We describe the logistical details and tools utilized so that other communities may leverage our experience. We delineate practical design ontology principles and present integration with an important platform terminology used in European marine community. The team concluded that the following lessons were learned:

- Creating ontologies from controlled vocabularies is more objective than creating a plain dictionary of terms or a thesaurus. For example, the exercise revealed that the semantics unrelated to platform properties (e.g actual routing rather than routing capability) had in the past been unnecessarily included in platform classification.
- Ontology represents a community consensus, and designers should expect to spend a lot of time discussing the concepts and the relations among them. We also discussed some of the terms outside the group; this brought new insights, enriching the process.
- Ontology experts are an important part of the team. Discussions about the qualifiers representation in classes and the inclusion of terms like "immobility" were guided by the team's ontology experts.
- Commonly held terms tend to be ambiguous. For example, the term "CTD" is widely used in the marine science community to refer to both an instrument (which measures any or all of conductivity, temperature, and pressure) and a platform (which is typically towed in the water column and carries numerous sensors and samplers). To solve the semantic heterogeneity problem we labeled the former CTDInstrument, and the latter CTDPlatform.

Work spawned by this development activity encompasses at least three new areas: continued improvement and availability of the platforms ontology; progression to development of the originally envisioned sensor ontology; and community engagement, via the MMI Sensor Metadata Interoperability workshop and other activities, in improving and using the resulting ontologies.

The MMI platforms ontology is a more sophisticated and comprehensive ontology than any existing platform reference of which the team was aware. The ontology is being tested in the MMI interoperability demonstration, OOSTethys [3], and its being used in the Autonomous Ocean Sampling Network [1]. The interoperability demonstration will provide practical usability feedback, making the ontology more robust.

The team that developed the platform ontology has started to consider a related sensor ontology, debating the best approach to this task. A sensor ontology will be much more complicated than the platform ontology, for several reasons:

- there are many more existing instruments, instrument makers, and instrument types,
- there is no clear set of organizing principles, or properties,
- there are no community-standard sensor type or naming conventions, and
- many more potential users are interested in classifying sensors and using the results.

The sensor ontology work continues at http://marinemetadata.org/sensorsont, and we invite other members of the marine community to join.

The platform ontology has been published on the MMI web site and SourceForge, and comments have been solicited in numerous community venues. Both the platform and sensor ontologies will be presented and discussed at the MMI Sensor Metadata Interoperability workshop, in Portland, Maine (October 19-20, 2006).

To be most effective, access to these ontologies must be widespread, simple, and persistent, and mechanisms must be provided for the public to comment, and revisions to be gracefully adopted. MMI is pursuing these tasks, and interested parties are encouraged to visit the MMI site to take advantage of the most recent improvements to its ontologies and related services, as well as contribute their own expertise and insights.

Acknowledgements. This activity was supported by the National Science Foundation through grant ATM-0447031. We would like to thank the other participants in this work: Roy Lowry, Rob Raskin, Kevin O'Neil and Marilyn Drewry.

References

1. Autonomous Ocean Sampling Network (AOSN), http://www.mbari.org/aosn
2. Bechhofer, S., Harmelen, F.v., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P., Stein, L.A.: Owl Web Ontology Language Reference, W3C Recommendation, (2004)
3. Bermudez, L.E., Bogden, P., Bridger, E., Creager, G., Forrest, D., Graybeal, J.: Toward an Ocean Observing System of Systems, in Proceedings of the Oceans'06 MTS/IEEE-Boston
4. Berners-Lee, T., R., F., Masinter, L.: Uniform Resource Identifiers (Uri): Generic Syntax, W3C/IETF Network Working Group, (1998)

5. Brickley, D., Guha, R.V.: Rdf Vocabulary Description Language 1.0: Rdf Schema, W3C, (2004)
6. Graybeal, J., Bermudez, L.E., Bogden, P., Miller, S., Watson, S.: Marine Metadata Interoperability Project: Leading to Collaboration, in Proceedings of the Local to Global Data Interoperability - Challenges and Technologies Symposium, Sardinia, Italy (June 19-24, 2005)
7. Graybeal, J., Watson, S., Bermudez, L.E., Galbraith, N., Stocks, K., Subramanian, V.: Marine Metadata Interoperability Workshop: Advancing Domain Vocabularies Workshop Report, MBARI, Moss Landing, CA, (2006)
8. Horridge, M., Knublauch, H., Rector, A., Stevens, R., Wroe, C.: A Practical Guide to Building Owl Ontologies Using the Protege-Owl Plugin and Co-Ode Tools, The University of Manchester (2004)
9. Klein, M.: Combining and Relating Ontologies: An Analysis of Problems and Solutions, in Proceedings of the Workshop on Ontologies and Information Sharing, IJCAI'01, Seattle, USA (August 4-5 2001)
10. Maryland Information and Network Dynamics Lab Semantic Web Agents Project (MINDSWAP): Pellet, http://www.mindswap.org/2003/pellet/
11. Maryland Information and Network Dynamics Lab Semantic Web Agents Project (MINDSWAP): Swoop, http://www.mindswap.org/2004/SWOOP/
12. McBride, B.: Resource Description Framework (Rdf): Concepts and Abstract Syntax, W3C, (2004)
13. McGuinness, D.L., Fikes, R., Rice, J., Wilder, S.: An Environment for Merging and Testing Large Ontologies, in Proceedings of the Principles of Knowledge Representation and Reasoning (KR2000), San Francisco, USA (2000)
14. Mena, E., Kashyap, V., Illarramendi, A., Sheth, A.P.: Managing Multiple Information Sources through Ontologies: Relationship between Vocabulary Heterogeneity and Loss of Information, in Proceedings of the 3rd Workshop Knowledge Representation Meets Databases (KRDB'96), Budapest, Hungary. (August 13, 1996)
15. Olsen, L.M., Major, G., Leicester, S., Shein, K., Scialdone, J., Weir, H., Ritz, S., Solomon, C., Holland, M., Bilodeau, R., Northcutt, T., Vogel, T.: Global Change Master Directory (Gcmd) Earth Science Keywords, http://gcmd.gsfc.nasa.gov/Resources/valids/keyword_list.html
16. Raskin, R., Pan, M.: Semantic Web for Earth and Environmental Terminology (SWEET), in Proceedings of the Semantic Web Technologies for Searching and Retrieving Scientific Data (SCISW), Sanibel Island, Florida (October 20th, 2003)
17. Sea Search, http://www.sea-search.net/
18. Sea Search Data Dictionaries, http://www.bodc.ac.uk/projects/european/seasearch/data_dictionaries/
19. Stanford Medical Informatics: Protege, http://protege.stanford.edu/
20. Voc2OWL, http://marinemetadata.org/voc2owl