Ontologies and Ontology Extension for Marine Environmental Information Systems

Adam Leadbetter[1], Torill Hamre[2], Roy Lowry[1], Yassine Lassoued[3], and Declan Dunne[3]

[1] British Oceanographic Data Centre, Liverpool, United Kingdom
[2] Nansen Environmental and Remote Sensing Center, Bergen, Norway
[3] Coastal and Marine Resources Centre, Cork, Ireland

Abstract. An increasing number of data delivery and processing services are available online from different environmental information systems. However, these services are often described using different terminology and may therefore be difficult to find. Integrating services from multiple providers and combining services into a workflow is likewise complicated. Development of an interoperable environmental data portal requires a detailed definition of the semantics of the data being requested and the services being called. Existing descriptions of semantics, captured in controlled vocabularies or ontologies, represent valuable knowledge that must be capitalised on. By ontology extension, we propose to “bridge” existing semantic resources, with the potential to represent more generic relationships and multilingual search and discovery. This ontology extension will be part of a toolkit for building environmental portals by means of standards for web-based data delivery and processing services.

Key words: ontology extension, semantic infrastructure, marine environmental information system, GMES, INSPIRE

1 Introduction

Today, a steadily growing wealth of marine data from a wide range of disciplines acquired in real-time, near-real time and delayed mode, are available from online information systems around the world. While substantial achievements have been made in reaching consensus on standards for data and metadata formats, system protocols and exchange mechanisms by standardisation organisations such as the Open Geospatial Consortium, Inc. (OGC), the International Standardisation Organization (ISO) and the World Wide Web Consortium (W3C), there are still a number of gaps prohibiting full interoperability between environmental information systems.

The NETMAR project [1] is part of the effort of the FP7 Information and Communication Technologies (ICT) Programme to develop interoperable components for the establishment of a Single Information Space in Europe for the Environment (SISE) and Shared Environmental Information System (SEIS). NETMAR aims to build on the results of the FP6 projects ORCHESTRA and Semantic Web services Interoperability for Geospatial Decision Making (SWING), plus the IST-FP7 integrated project Sensors Anywhere (SANY) to provide a toolkit for building environmental portals in a coherent manner by means of chained OGC Web Services (WxS), Open-source Project for a Network Data Access Protocol (OPeNDAP) services and W3C standards controlled by Business Process Execution Language (BPEL) workflow. NETMAR will also aim to utilise results emerging from ongoing and new projects addressing interoperability in environmental information systems.

The expected results of NETMAR will be a suite of components that can be used by user communities interested in building a portal for marine environmental data, and will offer search, download and integration tools for a range of satellite, in situ and model data from open ocean and coastal areas. Further processing of these data will also be available in order to provide statistics and derived products suitable for decision making. Within NETMAR, this toolkit will be used to develop a pilot European Marine Information System (EUMIS).

2 Ontologies for the Marine Domain

NETMAR has conducted a survey of existing ontologies that can be used as building blocks for EUMIS and other Global Monitoring for Environment and Security (GMES) downstream services and systems. Some of these are described briefly below.

GEneral Multilingual Environmental Thesaurus (GEMET) [2] is a repository maintained by the European Environment Information and Observation Network (EIONET) using the procedural models specified in ISO 19135 'Geographic information - Procedures for item registration'. GEMET has the potential to host and serve multiple semantic resources (termed registers) maintained by separate governance authorities under common technical governance. At the time of writing, there are two registers: the GEMET concept thesaurus and the Infrastructure for Spatial Information in Europe (INSPIRE) spatial data themes. The GEMET concept thesaurus is a rich multilingual resource that should be included in part if not in its entirety in the NETMAR semantic infrastructure. The INSPIRE themes can be used to provide a semantic linkage to INSPIRE for the NETMAR products and services.

The Natural Environment Research Council (NERC) Vocabulary Server (NVS) [3] is a collection of controlled vocabularies that cover many subjects and facets of the oceanographic domain. Currently, NVS holds some 135 publicly accessible lists, of which 30 lists containing about 45% of all defined terms, were found to be of interest to NETMAR. For instance, the SeaDataNet Parameter Disciplines list that provides a hierarchy of topics/themes for parameter classification, several other SeaDataNet vocabularies, as well as lists originating from the British Oceanographic Data Centre, SeaVoX and several ISO standards. Many of the list entries are linked together by a set of relationships that approximate to the Simple Knowledge Organisation System (SKOS) model. Therefore, the NVS could become the base semantic resource on which the NETMAR extended ontology is developed.

The NASA Semantic Web for Earth and Environmental Terminology (SWEET) ontologies [4] Version 2.0 (beta release as of August 2010) are a modular network of 150 ontologies containing 4600 concepts. It was initially populated from NASA's Global Climate Change Master Directory (GCMD) Science Keywords, but the content has undergone significant subsequent development. The network is designed as a faceted upper-level ontology for Earth system science. Significant emphasis has been placed on the development of orthogonal concepts that may be combined to produce additional concepts. For example, the ontologies include substances such as 'water', 'air' and 'blood' and properties such as 'temperature' rather than pre-combined concepts like 'air temperature'. This facilitates discovery searches along either the substance axis or the property axis.

Other identified semantic resources include the GCMD; Marine Metadata interoperability's (MMI) Ontology Repository & Registery (ORR); the International Council for the Exploration of the Seas (ICES) Reference Code (RECO) system; the United States Geological Survey (USGS) Thesaurus; the USGS Information Bank; data and keyword ontologies from the InterRisk project; Geoscience Markup Language (GeoSciML) vocabularies; World Register of Marine Species (WoRMS) Aphia Records; and Quantities, Units, Dimensions and Types (QUDT) ontologies.

3 Ontology Extension

To develop an interoperable environmental data portal, the NETMAR project requires a detailed definition of the semantics of the data being requested and the services being called. A major goal of the project is therefore to develop a multi-domain and multilingual ontology of marine data and services to provide semantic interoperability [5] within the system. This will allow searching across human languages as well as across scientific domains. In order to achieve this, NETMAR will focus on data sets described by metadata records which adhere to the ISO19115 standard in which the fields are populated from controlled vocabularies. Initially, NETMAR will also only support data marked up in English, but these data will be discoverable in a variety of human languages.

The approach taken to building an ontology for NETMAR will be to analyse existing semantic resources and provide mappings between them, gluing together the definitions and worksflows of the OGC WxS services. The concept of ontology extension is well documented in such domains as computer science [6, 7], bioinformatics [8] and eLearning [9], but has not previously been demonstrated in the oceanographic community. The mappings between the terms aim to be more general than the standard "narrower than", "broader than" type seen in the thesauri or simple ontologies implemented by previous projects. Tools for the development and population of ontologies will also be needed in the semantic infrastructure. NETMAR will provide such tools for selected parts of the marine domain as there will be instances in which existing resources cannot sufficiently describe newly encountered data or services.

Fig. 1 shows a flow chart for the process of ontology extension in NETMAR. An existing ontology typically represents knowledge within a single domain, with “simple” or “richer” relationships between the concepts. Besides describing the domains, an ontology gives the ability to reason about the concepts and their relationships. If the ontology on the other hand is extended, i.e. linked to other ontologies, it is possible to increase the breadth and depth of the represented knowledge and also infer knowledge across domains. Linking to multilingual ontologies further extends the reasoning capabilities by allowing the user to specify search parameter in their native language, while the semantic infrastructure will also find resources described using these same or related terms in other human languages. Ontology extensions and the use of multilingual ontologies further enable the chaining of web and data processing services, by allowing verification that the input data to each processing step fulfil given criteria.

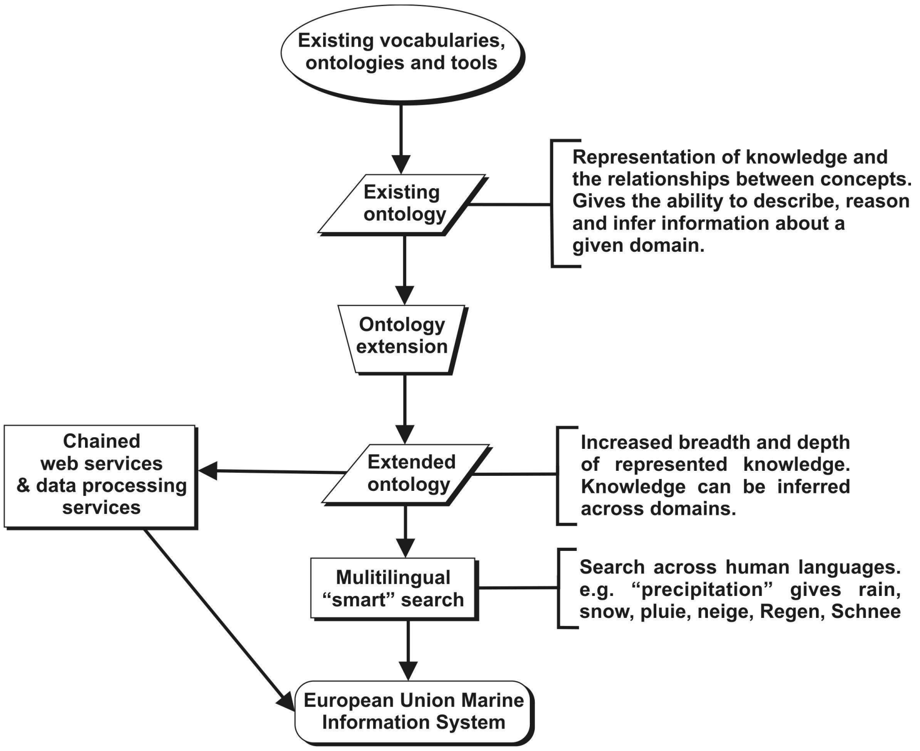
Existing vocabularies, ontologies and tools
Existing ontology
Representation of knowledge and the relationships between concepts. Gives the ability to describe, reason and infer information about a given domain.
Ontology extension
Chained web services & data processing services
Extended ontology
Increased breadth and depth of represented knowledge. Knowledge can be inferred across domains.
Mulitilingual “smart” search
Search across human languages. e.g. “precipitation” gives rain, snow, pluie, neige, Regen, Schnee
European Union Marine Information System

Fig. 1. A flow chart for the process of ontology extension within NETMAR (after [10]).

There are a number of challenges presented by ontology extension, and the literature contains a range of automated and semi-automated schemes to overcome these issues [11]. However, due to the domain specificity of the NETMAR project, a manual approach is to be used. The major imitations of this approach are that the ontologies being extended have been developed without being made interoperable and that human identification of related semantic concepts is required in order to produce the mappings between the various semantic resources. This is a time consuming process, and one which requires science domain expertise but is possibly open to human error, and as such the construction of the NETMAR extended ontology will take time and effort. The formation of a NETMAR ontology governance body [12] goes some way to mitigating the concerns over domain knowledge and human error.

It would be possible to use programmatic methods to produce the extended ontology, but as noted in [13] the re-use of existing algorithms for this purpose is difficult, in part due to the heterogeneous nature of the distributed resources [8], and the need for human validation is not removed. The identified base semantic resource for NETMAR [1] contains some 100,000+ mappings which have been generated by hand and these mappings already bridge between several of the semantic resources identified above. As such the NETMAR ontology will be able to operationally support "smart" discovery (a search for "precipitation" returns concepts also labelled "rain", "pluie", "Regen" etc. [14]) of concepts, products and services, which has previously only been demonstrated using small, custom built ontologies.

4 Semantic Infrastructure

Standard technologies and tools are needed to develop a semantic infrastructure that supports the planned ontology extensions and tools for creation and management of semantic resources, web service and processing workflows. NETMAR has conducted a survey of ontology and semantic framework technologies. Due to the broad compatibility between its members, its recommendation by the W3C and its large software base, we recommend the use of the Resource Description Framework (RDF) family of languages to represent ontologies. The SPARQL Protocol and RDF Query Language (SPARQL) is the recommended query language due to the ubiquitous nature of support for it and its high level of extensibility. For ontology publication, the Mulgara server has the easiest mechanisms for entering data and then querying it, with a simple HyperText Transfer Protocol (HTTP) interface to a SPARQL and interactive Tucana Query Language (iTQL) endpoint.

Jena is a powerful complete ontology framework that provides two mechanisms for storing data, one as a wrapper to a relational database, the other using a bespoke system similar to the Mulgara server. The querying mechanism is an extension of SPARQL called ARQ which provides access to Extensible Stylesheet Language Transformations (XSLT) functions that make complex queries possible. Jena also provides classes for creating and editing ontologies programmatically.

Additional tools needed include an ontology editor and a concept mapping tool, for which Semantic Turkey and CMAPTools Ontology Editor are recommended within the NETMAR project. Other tools, such as Protégé SWOOP and PoolParty were reviewed [15] but found to be lacking for the needs of the NETMAR project, on the whole due to problems loading large (20,000+ terms) vocabularies into these tools. The Marine Metadata Interoperability (MMI) project's Vocabulary Integration Environment may be utilised to build bridges between ontologies stored in the MMI Ontology Registry and Repository and other semantic resources.

Several semantic frameworks have been developed over the past few years for Earth Information Systems (EIS). A review of the most popular ones has been carried out as part of NETMAR [16]. The review included, but was not limited to, the InterRisk semantic framework [17], the International Coastal Atlas Network prototype mediator [18], the MMI and the Ocean Observatories Initiative (OOI) semantic frameworks [19], the ORCHESTRA Semantic Catalogue Architecture [20], and OOSTethys [21]. These existing semantic frameworks have been designed for various purposes (data/metadata interoperability, data discovery, etc.) and therefore offer different capabilities and functionalities. Nevertheless, they all deal with simple resources such as datasets, documents or non-complex data and metadata services with limited types of relationships (in general: broader term; narrower term; equivalent term; and related term). The semantic infrastructure provided by NETMAR will aim to move beyond the approach of existing frameworks. Using the tools described above, the extended ontology and a richer set of relationships, the semantic infrastructure will support functional interoperability [5] through the discovery of complex services (such as OGC Web Processing Services), allow service chaining and describe the propagation of uncertainty through the processing scheme.

5 Conclusions

Development of an interoperable environmental data portal requires a detailed definition of the semantics of the data being requested and the services being called. NETMAR has therefore identified a number of ontologies, both specific to the marine domain as well as more generic resources that can be used as building blocks in a semantic infrastructure. These semantic resources will be analysed and mappings provided between them, to extend the different resources, thus allowing cross-resource search and discovery. In addition, the ontology extensions will form the basis for the semantic infrastructure that will support chaining of OGC WxS services, where for each step it can be verified that the proposed input data satisfy the criteria for the service. By integrating the semantic infrastructure into the data and processing elements it will be possible to enforce correct connections between components (e.g. do not send chlorophyll data to a component that only knows how to process sea surface temperature data). It also becomes possible to pass additional metadata along the chain, such as the error introduced by each component.

The NETMAR project approach makes possible the reuse and building upon the large number of concepts and relationships already defined by existing semantic resources, while facilitating more generic relationships in extensions. It is expected that the methodology for the ontology extensions will be applicable to other domains.

Acknowledgement

The NETMAR (Open service network for marine environmental data) project is supported by the European Union Framework Programme 7, under grant agreement no. 249024.

References

1. NETMAR (Open service network for marine environmental data), http://netmar.nersc.no/
2. GEMET, http://www.eionet.europa.eu/gemet
3. NERC Vocabulary Server, http://vocab.ndg.nerc.ac.uk/
4. NASA SWEET Ontologies, http://sweet.jpl.nasa.gov/ontology/
5. Heiler, S.: Semantic interoperability. ACM Computing Surveys 27(2): 271-273 (1995)
6. Ouksel, A., Sheth, A.: Semantic interoperability in global information systems. ACM SIGMOD Record 28(1): 5-12 (1999)
7. Pazienza, M. T., Stellato, A., Henrikson, L., Paggio, P., Zanzotto, F. M.: Ontology mapping to support ontology-based question answering. In Proceedings of the Meaning 05 Workshop (2005)
8. Mungall, C., Bada, M., Berardini, T., Deegan, J., Ireland, A., Harris, M., Hill, D., Lomax, J.: Cross-product extensions of the Gene Ontology. Journal of Biomedical Informatics (in press)
9. Read, T., Verdejo, F., Barros, B.: Incorporating interoperability into a distributed eLearning system. In D. Lassner & C. McNaught (Eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2003 (pp. 273-282). Chesapeake, VA: AACE (2003)
10. Leadbetter, A., Lowry, R., Clements, O.: Semantics in NETMAR (open service NETwork for MARine environmental data). Geophysical Research Abstracts, 12:2638 (2010)
11. Conroy, C.: Towards ontology mapping for ordinary people. In Proceedings of the ESWC 2008 Ph.D. Symposium 21-25 (2008)
12. Lowry, R.: Establishment of NETMAR ontology governance body. NETMAR (Open service network for marine environmental data) Deliverable D3.1. European Commission Information Society and Media Directorate-General Grant Agreement Number 249024 (2010)
13. Zhadnova, A., Shaviko, P.: Community driven ontology matching. In Proceedings of the 3rd European Semantic Web Conference 34-49 (2006)
14. Latham, S., Cramer, R., Grant, M., Kershaw, P., Lawrence, B., Lowry, R., Lowe, D., O'Neill, K., Miller, P., Pascoe, P., Pritchard, M., Snaith, H., Woolf, A.: The NERC DataGrid Services. Philosophical Transactions of the Royal Society A. 367: 1015-1019 (2009)
15. Leadbetter, A., Clements, O.: Review of available ontology tooling. NETMAR (Open service network for marine environmental data) Deliverable D3.2. European Commission Information Society and Media Directorate-General Grant Agreement Number 249024 (2010)

16. Lassoued, Y.: Review of Semantic Frameworks. NETMAR (Open service network for marine environmental data) Deliverable D4.1. European Commission Information Society and Media Directorate-General Grant Agreement Number 249024 (2010)
17. Coene, Y., Truong-Minh, H., Lassoued, Y.: Discovery standards in HMA-T and Discovery in FP6 InterRisk. Presentation at Ontology and Discovery Workshop, ESRIN, Frascati, Italy, 4 March 2009 (2009)
18. Lassoued, Y., Wright, D., Bermudez, L., Boucelma, O.: Ontology-based Mediation of OGC Catalogue Service for the Web: A Virtual Solution for Integrating Coastal Web Atlases. In Proceedings of the 3rd International Conference on Data and Software Engineering ICSOFT 2008 , Berlin, Germany (2008)
19. Bermudez, L.: OOI Semantic Prototype. Presentation at ICAN 4 Workshop, Trieste, Italy, 16-20 November 2009 (2009)
20. Usländer, T.: Reference Model for the ORCHESTRA Architecture (RM-OA) V2 (Rev 2.1). Open Geospatial consortium Inc. (2007)
21. OOSTethys Architecture web page, http://www.oostethys.org/System