Putting meaning into NETMAR – the open service network for marine environmental data

Adam M. Leadbetter, Roy K. Lowry & D. Oliver Clements

Taylor & Francis
Taylor & Francis Group

Putting meaning into NETMAR − the open service network for marine environmental data

Adam M. Leadbetter[a]*, Roy K. Lowry[a] and D. Oliver Clements[b]

[a]*British Oceanographic Data Centre, Liverpool, UK;* [b]*Plymouth Marine Laboratory, Plymouth, UK*

The open service network for marine environmental data (NETMAR) project uses semantic web technologies in its pilot system which aims to allow users to search, download and integrate satellite, *in situ* and model data from open ocean and coastal areas. The semantic web is an extension of the fundamental ideas of the World Wide Web, building a web of data through annotation of metadata and data with hyperlinked resources. Within the framework of the NETMAR project, an interconnected semantic web resource was developed to aid in data and web service discovery and to validate Open Geospatial Consortium Web Processing Service orchestration. A second semantic resource was developed to support interoperability of coastal web atlases across jurisdictional boundaries. This paper outlines the approach taken to producing the resource registry used within the NETMAR project and demonstrates the use of these semantic resources to support user interactions with systems. Such interconnected semantic resources allow the increased ability to share and disseminate data through the facilitation of interoperability between data providers. The formal representation of geospatial knowledge to advance geospatial interoperability is a growing research area. Tools and methods such as those outlined in this paper have the potential to support these efforts.

Keywords: geoinformatics; oceanography; geospatial data integration

Introduction

The World Wide Web (WWW) may be viewed as a vast library, storing documents, images and other media in formats accessible to a human audience through a browser. However, the content of the World Wide Web is largely unstructured. The semantic web is an extension of these fundamental ideas of the World Wide Web into a web of data. This is achieved through the annotation of documents with hyperlinks to machine-readable documents. These machine-readable documents describe the meaning of given fields or keywords (Berners-Lee, Hendler, and Lassila 2001, Shadbolt and Berners-Lee 2008).

The deployment of semantic web technologies is well documented in the fields of biosciences (Smith et al. 2007), business (Jin et al. 1998) and linguistics (Biébow and Szulman 1999). However, in the marine environmental domain, the semantic component of projects such as SeaDataNet (Lowry and Williams, forthcoming), InterRisk (Coene, Truong-Minh, and Lassoued 2009), the International coastal atlas

---

*Corresponding author. Email: alead@bodc.ac.uk

network (ICAN; Lassoued et al. 2008) and OOSTethys (http://www.oostethys.org) have been generally aimed at the mark-up and discovery of metadata and data and with a range of non-complex relationships. These systems also sit in isolation, with no semantic relationships defined between the concepts each of them defines.

The open service network for marine environmental data (NETMAR) project uses these semantic web technologies in its pilot system. This pilot system aims to allow users to search, download and integrate satellite, *in situ* and model data from open ocean and coastal areas. The NETMAR pilot system offers flexible data and service discovery, access and chaining facilities using Open Geospatial Consortium (OGC) and World Wide Web Consortium (W3C) standards. It uses a semantic framework coupled with web published semantic resources in order to identify and access distributed near-real time, forecast and historical data. The NETMAR system also enables further processing of such data to generate composite products and statistics suitable for decision-making in diverse marine application domains.

The primary purpose of developing semantic resources for NETMAR is the support of operational 'smart discovery' (Latham et al. 2009). This is the process by which users of the system are able to locate and therefore utilise datasets using search terms that are different but semantically linked to data-set labels such as keywords. A frequently quoted example is the location of datasets labelled 'rainfall' using the search term 'precipitation'. In a global context, the issue of human language arises which means that operational 'smart discovery' also needs to be able to link datasets labelled in one European language to search terms supplied in another. A secondary goal of this activity is to semantically validate the inputs to OGC Web Processing Services (WPSs) and to mark-up the outputs of these services. This facilitates the simplification of orchestrating chains of these services in a workflow engine. A third goal is to implement data interoperability approaches to coastal web atlases (CWAs) in an ICAN pilot.

This paper describes the NETMAR semantic resource through first introducing the spectrum in complexity which exists in semantic web resources. It then demonstrates the results of constructing an interconnected semantic resource (a machine-readable document, published on the Internet using standard semantic web technologies and which can be accessed through a dereferencable Uniform Resource Identifier [URI]) for the NETMAR project. The use of semantic web technologies and the previously described semantic resource to promote the discovery and use of complex services (e.g. OGC WPSs) is then shown. A system which facilitates the chaining of these complex services with semantic validation of the inputs is also introduced.

Knowledge organisation

Spectrum of detail in semantic resources

The semantic web relies on machine-readable resources for the annotation of data and metadata elements. Within these resources, there exists a broad spectrum in the complexity and the level of detail encoded. The points of this spectrum, often called the 'ontology spectrum' or 'semantic spectrum', are defined by McGuinness (2003).

In overview, the simplest point on the spectrum is the catalogue [or glossary, as in Navigli and Velardi (2008)] that simply provides a list of terms used to annotate data and their definitions. A catalogue is not necessarily published to either the World Wide Web or semantic web. Nor does a catalogue require a group of experts

to provide governance over its content. Complexity is increased in the controlled vocabularies (Warner 2002) and data dictionaries (Kumpati 1985, IBM 1993) which add, respectively, naming conventions and publication on top of the catalogue. A data model (van Renssen 2005) then adds some definition of the relationship between the entities of a data dictionary. A taxonomy further adds a tree hierarchy with the concepts at the bottom of the tree inheriting the properties of those concepts at the top of the tree (Navigli and Velardi 2008). A concept can be viewed as the definition of an idea or notion, a unit of thought (Miles and Bechhofer 2009).

A more complex point of the semantic spectrum is the ontology. An ontology describes all the concepts in a given domain and the relationships between them, and following best practice (Cox 2012), are conceptualised using URI for each of the data elements, relationships and property types. An ontology can be parsed by a reasoning engine and then used to infer knowledge about the domain.

Faceted discovery

An ontology may be described as either a formal or informal ontology. An informal ontology (Altheim 2008) is much like a taxonomy in that it describes a hierarchy of concepts with defined entry points at the broadest level of the hierarchy. The concepts below an entry point of the hierarchy may be informally inferred to be of the type defined by that entry point. The grouping of concepts below the entry point is termed a facet, and may be used to describe a domain. For example, facets of interest to a marine environmental scientist may include the parameter being measured, the instrument used to make the measurement and the platform from which the instrument was deployed. Relationships between the concepts of different facets may be defined. For example, if one considers a Conductivity–Temperature–Depth probe to be an instrument, it can be related to measurement of conductivity, and in turn calculations of salinity. These relationships can be used to create data discovery interfaces in which a user is able to explore their area of knowledge. The result is that users may find data that could have remained undiscovered using traditional search techniques.

This differs from the creation of a formal ontology, in that a formal ontology explicitly constrains the concepts it contains through the definition and use of classes and attributes (Gruber 2009). An ontology class defines a group of concepts that belong together because they share some properties. Attributes can be used in an ontology to restrict the values which a concept may take. Therefore, there is no need to infer inheritance through the facet. Ontological definitions of, in particular, observed environmental parameters often do not carry the required level of detail to identify exactly what has been measured. Therefore, ontologies may be extended through detailed parametric descriptions from data dictionaries, data models or taxonomies.

Ontology extension

Developing any level of semantic resource, from a data dictionary to a formal ontology, is a valuable activity as it provides a new level of interoperability between the metadata and data annotated with terms defined by that resource. However, a

much richer semantic resource may be developed by looking for commonality between concepts defined in different resources. The relationships between these common concepts can then be described and encoded. This is the process known as ontology extension. It is of particular use when an authoritative or legally required resource exists and it is wished to create links to the concepts published by that resource. Examples of these resources include: the World Register of Marine Species (WoRMS, http://www.marinespecies.org/) which aims to publish the most authoritative list of names of marine species; the European Research Vessel Infobase (http://www.rvinfobase.eurocean.org/) which maintains a searchable database of the European research vessels including vessel specifications and contact information and European Commission INSPIRE directive code lists.

The ontology extension approach also allows the reuse of classes defined by other ontologies or to define the class to which the concepts defined in a data model or taxonomy belong. For example, classes defining marine environmental science facets were developed by Rensselaer Polytechnic Institute's Tetherless World Constellation for the Woods Hole Oceanographic Institute's Biological and Chemical Oceanography Data Management Office (Chandler et al. 2011). These classes are instantiated by concepts from SeaDataNet data dictionaries served by the Natural Environment Research Council (NERC) Vocabulary Server (NVS) in the UK.

The approach to ontology extension in NETMAR is described by Leadbetter et al. (2010).

The NVS version 2.0

Resource registry

The data and service discovery portal developed by the NETMAR project requires the use of data and metadata resources from distributed sources. Therefore, the semantic resource used to describe those data and services must be published online. The semantic resource must also provide dereferenceable URI for each term it describes. This allows proper annotation of metadata describing services and data alongside the construction of new services which are aware of the meaning provided by the semantic resource. These principles are adhered to in the NVS, which was originally developed in 2006. The content it serves has been used for: metadata mark-up with verifiable content; populating dynamic drop down lists; semantic cross-walk between metadata schemata and so-called smart search in projects and programmes including: the NERC Data Grid (Lawrence et al. 2009), SeaDataNet (Schaap and Lowry 2010), Geo-Seas and the European Marine Observation and Data Network.

In order to support the requirements of the user community identified by the NETMAR project, several enhancements were required to the existing NVS, and therefore a version 2.0 (NVS2.0) was developed. The major upgrades consisted of: a move to the latest version of the W3C's Simple Knowledge Organization System (SKOS) specification (Miles and Bechhofer 2009) for encoding the data dictionaries and taxonomies served through the NVS; the ability to serve multilingual titles and definitions for resources and the provision for mappings to external resources enabling the results of ontology extension to be delivered. As reported by Ma et al. (2011), SKOS is an excellent tool for describing and publishing geoscientific data dictionaries

and taxonomies. At its core is the W3C's (2004) recommended standard resource description framework used for data interchange across the World Wide Web.

Implementing SKOS

SKOS, however, has a limited set of data types and properties, which cannot always be used to encapsulate all of the meaning necessary in a production system. Therefore, decisions must be taken as to how to implement the SKOS standard with an enriched and enhanced set of properties. Ma et al. (2011) took the approach of generating a new namespace for the geological time scale domain, whereas the NVS2.0 registry uses a range of available, external namespaces in order to maintain interoperability but also to enrich the SKOS model. These namespaces and their use within NVS2.0 are detailed in Table 1.

The SKOS specification also introduces two approaches to the aggregation of concepts: as concept collections or as concept schemes. The SKOS specification only loosely distinguishes between the two, as follows. SKOS concept collections are labelled and/or ordered groups of SKOS concepts. Collections are useful where a group of concepts shares something in common, and it is convenient to group them under a common label, or where some concepts can be placed in a meaningful order. A SKOS concept scheme can be viewed as an aggregation of one or more SKOS concepts. Named and typed semantic relationships between those concepts may also be viewed as part of a concept scheme.

A concept within NVS2.0 is registered to one, and only one, concept collection (analogous to a controlled vocabulary – that is a data dictionary which can be used for lookup). A hierarchical grouping of concepts with one or many top level entry points is encoded as a concept scheme (analogous to a taxonomy). An NVS2.0 concept scheme incorporates concepts registered to one, or many, concept collections. This data model is illustrated in Figure 1. As an aid to the ontology extension process, the concepts registered within an NVS2.0 concept scheme do not have to be served from NVS2.0 itself but can be registered with any openly accessible resource on the World Wide Web. For example, the NETMAR Oceanography Thesaurus (http://vocab.nerc.

Table 1. Summary of namespaces beyond RDF and SKOS used within NVS2.0.

| Namespace	Usage notes
Dublin Core http://purl.org/dc/terms/	Adding title, description, publisher and creator to data dictionaries and taxonomies; dates to resources and replacement terms for deprecated concepts.
ISO/TC211 http://www.isotc211.org/schemas/grg/	Adding provenance information (register owner and register manager) to data dictionaries.
RDF schema http://www.w3.org/2000/01/rdf-schema#	Adding additional comments describing the content governance body to data dictionaries.
Web ontology language http://www.w3.org/2002/07/owl#	To allow exact matching outside of taxonomies (specifically, SKOS concept schemes) and for adding version control information to data dictionaries and individual entities, including deprecation status

RDF, Resource Description Framework.
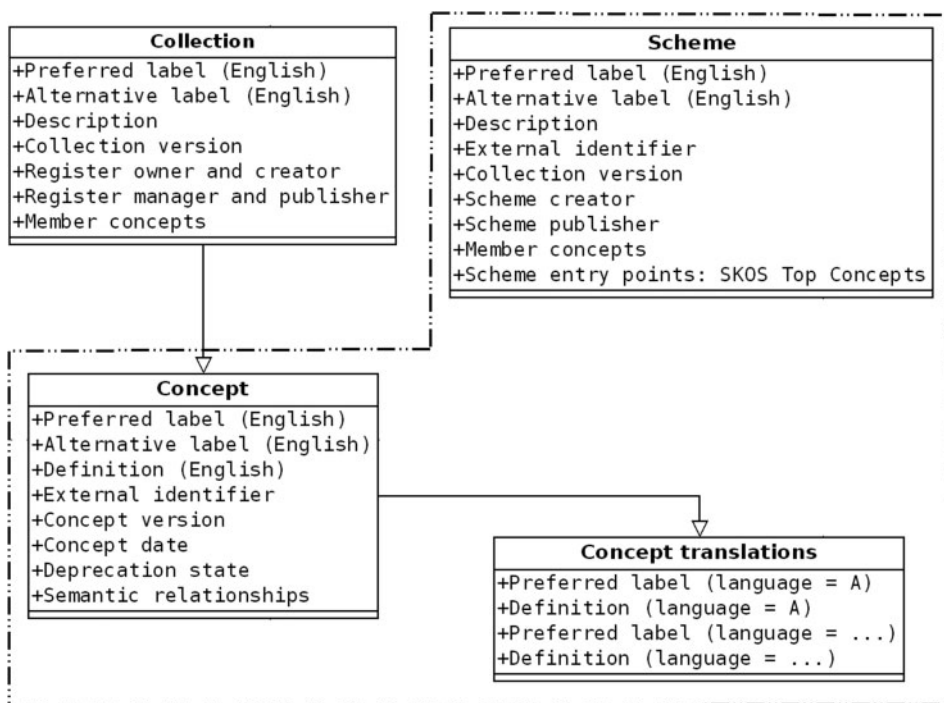
Figure 1. The data model employed within NVS2.0. A SKOS concept registered on NVS2.0 is associated with one, and only one, SKOS concept collection from which its URI is derived. However, the concept may then be accessed by many concept schemes.

ac.uk/scheme/NETMAR_OCEAN) contains biological entity references served directly from the Biodiversity Information Standards (Taxonomic Database Working Group [TDWG]) Life Science Identifiers database. The alternative approach, and one which is commonly taken, would be to cache a copy of these resources within NVS2.0. The benefit of not caching is that the data within NVS2.0 cannot become unsynchronised with the copy held by its originator. The approach of referencing World Wide Web resources also follows more closely both the original semantic web vision (Berners-Lee, Hendler, and Lassila 2001) the Linked Data principles (Bizer, Heath, and Berners-Lee 2009). However, one issue that this approach raises is that of service availability: what happens if a resource becomes deregistered from the Web, or if the mapped Uniform Resource Locator (URL) is no longer dereferenceable? To this end, an automated script is run once a month to check that there is still a valid payload returned from each individual external reference stored within NVS2.0.

As collaborative and distributed data networks of this type become more prolific, the question of the provenance of data for tasks such as quality assessment and the management of remote data sources becomes more important. This is also true of the information served from semantic resources. The W7 model for data provenance (Ram and Liu 2006) has been shown to be adaptable for capturing domain-specific provenance information (Ram and Liu 2009). Orlandi and Passant (2011) also show that this model is ideal for describing formally published semantic web resources. The W7 model asks seven questions about a data resource in order to model its

provenance: *what*, *when*, *where*, *how*, *who*, *which* and *why*. Within NVS2.0, the *what*, *when*, *who* and *why* are delivered as part of the SKOS concept collection or concept scheme documents. Actions which may affect a resource (*what*) within NVS2.0 are its creation, modification or deprecation. These events are versioned and time stamped (*when*) and have agents with either the role of register owner or register manager associated with them (*who*). The W7 model explains *why* in terms of the aim of an event acting upon a resource. The goal of a SKOS concept collection or concept scheme served by NVS2.0 is delivered in the header information of that document. *Which* application was responsible for the event is implicit in that all concepts registered on NVS2.0 are created, modified or deprecated on NVS2.0. The *how* and *where* are not currently captured within NVS2.0 payload documents.

Another important aspect of provenance for users marking up their data using data dictionary services is version control of the served dictionaries and assurance that a concept will not be deleted from a dictionary once published, ensuring validity of previously created metadata. A concept registered on NVS2.0 cannot be deleted, but may be deprecated. Once deprecated a flag on the concept is set to indicate that it should no longer be used to mark-up data or metadata. The back office system maintains mappings from deprecated concepts to their replacements and in future the NVS payloads will also return this information, allowing clients to navigate to replacement concepts.

Instead of deprecation, the context of a concept may be altered. The actions taken will vary based on the scope of that contextual change. If a change means the definition of the concept broadens, its version number would be incremented as would the version number of the concept collection it is registered with. This is acceptable because the broader definition still correctly describes any existing use of the concept in data or metadata. However, any narrowing of meaning is handled by deprecating the concept, then creating a new replacement concept. An unmanaged narrowing of a concept definition may mean that any data or metadata referencing that concept would be rendered invalid. In this managed system the version number of both list and deprecated concept increases; the new concept has a version number of 1.

Versioning within NVS2.0 is controlled by the collection – every 24 hours a new version of a concept collection is published if that concept collection contains new, modified or deprecated concepts. The URI of the collection contains a field used to specify which version of the collection is returned. The concepts each carry their own individual version number.

Supported interface paradigms

Once a well-defined semantic resource with good provenance information has been designed and populated, it needs to be made available to consumers. NVS2.0 meets the W3C's definition of a web service as a software system designed to support interoperable machine-to-machine interaction over a network (Booth et al. 2004). As such, NVS2.0 may be accessed through the Simple Object Access Protocol (SOAP; Box et al. 2000) protocol and a Web Services Definition Language document is published (http://vocab.nerc.ac.uk/vocab2.wsdl) describing the service. However, SOAP web services do not provide dereferenceable URLs to concepts, so a RESTful (Fielding and Taylor 2002) service is also published (http://vocab.nerc.ac.uk) based around URL access to the knowledge registered with NVS2.0.

Application programming interface methods

Prior to designing the application programming interface (API) methods for NVS2.0, a survey of 13 available semantic resources in the earth sciences domain was undertaken. Of those surveyed, at the time seven offered an API:

- GEneral Multilingual Environmental Thesaurus (GEMET, http://www.eionet. europa.eu/gemet);
- NVS1.X (http://vocab.ndg.nerc.ac.uk/);
- Marine Metadata Interoperability Ontology Registry and Repository (MMI ORR, http://mmisw.org/orr/#b) as described by Graybeal, Isenor, and Rueda (2012);
- United States Geological Survey thesaurus (USGS, http://www.usgs.gov/ science/about/);
- Geosciences Markup Language vocabularies (GeoSciML, http://srvgeosciml. brgm.fr/eXist2010/brgm/client.html);
- WoRMS;
- Spatial Information Services Stack Vocabulary service (SISSVoc, https://www. seegrid.csiro.au/wiki/Siss/VocabularyService) developed by AuScope and the Commonwealth Scientific and Industrial Research Organisation.

The available methods were summarised by Leadbetter, Clements, and Lowry (2011) (reproduced below in Table 2). Following the result of this survey, the NVS2.0 API was specified with nine methods including the most common emerging standards in vocabulary access. All nine API methods are available through the SOAP service, with a subset (six) being available through the ReSTful interface. The methods and their availability in the two API paradigms are shown in Table 3.

Results and discussion

The NETMAR semantic resource

Using the API described above, NVS2.0 was made available to the NETMAR project. NETMAR identified four pilot studies for the semantically aware data and service discovery portal it is developing. Following a survey, three pilots were identified as totally within the oceanographic domain and therefore able to share a common semantic resource. The fourth, the ICAN (O'Dea et al. 2011), spanned a number of knowledge domains linked by the concept of 'coast' which required development of a separate semantic resource. This provided a much less complex approach to ICAN semantics than that developed by Wright et al. (2008).

The NETMAR semantic resource is designed around the faceted approach outlined above and describes: observed parameter, measuring instrument, observing platform, observing platform class, project and vertical coverage. A concept map of this faceted resource is shown in Figure 2.

The facets are interconnected at a narrow level of description, placing the NETMAR semantic resource into the taxonomy category of the semantic spectrum. These mappings have all been hand-coded as the accuracy of automated mappings is not guaranteed. This is, however, a time-consuming process and future work in this area may assess how automated ontology matching techniques can be better utilised.

Table 2. Survey of vocabulary server API methods.

Resource	What data dictionaries?	Get data dictionary	Get term	Get related terms	Verify term exists in registry	Regular expression search	SPARQL search	Keyword search
GEMET	Yes		Yes	Yes		Yes		Yes
NVS2.0	Yes	Yes	Yes	Yes	Yes		Yes	Yes
MMI ORR	Yes	Yes	Yes	Yes	Yes		Yes	
USGS			Yes	Yes				Yes
GeoSciML Vocabularies	Yes	Yes	Yes			Yes		Yes
WoRMS			Yes					Yes
SISSVoc	Yes	Yes	Yes	Yes	Yes		Yes	Yes

Note: In the interests of clarity, only common methods are shown.
SPARQL, SPARQL Protocol and RDF Query Language.

Table 3. API methods available in NVS2.0.

Method	Notes	Available in SOAP	Available in REST
GetCollections	Return a catalogue of available data dictionaries (SKOS concept collections)	Yes	Yes
GetConceptCollections	Return a given SKOS concept collection	Yes	Yes
GetConcept	Return a given concept	Yes	Yes
GetSchemes	Return a catalogue of available taxonomies (SKOS concept schemes)	Yes	Yes
GetConceptSchemes	Return a given SKOS concept scheme	Yes	Yes
GetRelatedConcepts	Return a given concept, and all those concepts to which it is related	Yes	
GetTopConcepts	Returns top level entry points of a given SKOS concept scheme	Yes	
SearchVocab	Search NVS2.0	Yes	
VerifyConcept	Verify the existence of given concept within NVS2.0	Yes	Yes (through HTTP headers)

REST, Representational State Transfer.

Due to the large number of concepts stored in NVS2.0 (100,000 +), there are performance issues with the commonly used ontology editing tools, such as Protégé. This means that the mappings are simply drafted in a spreadsheet, which can then be loaded into the relational database management system which stores the NVS2.0 knowledge base.

Each of the facets in the NETMAR semantic resource has an explicitly declared entry point at the broadest level of description, and the concepts below these entry points in the hierarchy can be inferred to inherit the facet to which they belong, therefore moving the NETMAR resource into the realm of informal ontology. Furthermore, the ocean biogeochemistry data ontology described by Chandler et al. (2011) defines the class 'parameter' and 'instrument' which are instantiated by concepts from the narrowest description level of the NETMAR semantic resource. This moves portions of the NETMAR resource into formal ontologies.

A slightly different approach was implemented by the ICAN community. ICAN aims to make CWAs interoperable across jurisdictional boundaries and to create a smart search mediator for OGC Catalogue Services. The authors of local CWAs often use local vocabularies to describe their content and for interoperability, a global resource mapped to these local resources must be defined. Earlier work in this area (Lassoued et al. 2008, Wright et al. 2008, Graybeal, Isenor, and Rueda 2012) took the approach of mapping local Web Ontology Language (OWL) ontologies stored both as files on File Transfer Protocol (FTP) servers and within the MMI ORR to a global OWL ontology. However, the perception within the ICAN community, while not necessarily quantifiable by usage statistics, is that there was little uptake following an initial concerted effort. This perception may have stemmed from a combination of
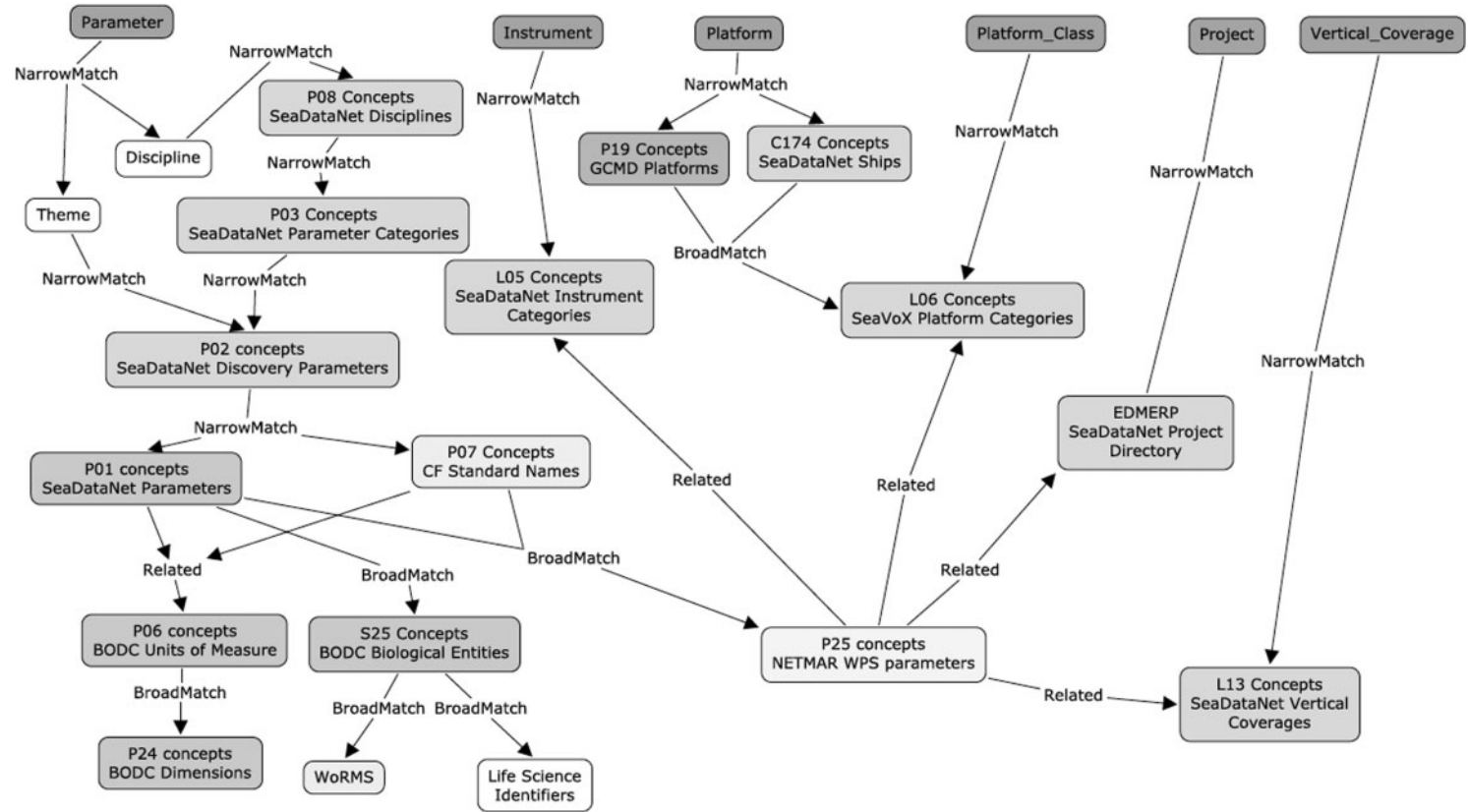
Figure 2. A concept map of the NETMAR semantic resource (http://vocab.nerc.ac.uk/scheme/NETMAR_OCEAN). The shading indicates the governance body responsible for maintaining the content of each collection used to source concepts for the scheme. SeaVoX is the SeaDataNet and MarineXML Vocabulary Content Governance Group; GCMD is NASA's Global Change Master Directory; BODC is the British Oceanographic Data Centre. Concepts within P01 are used within SeaDataNet, but the content governance resides with BODC.

unfortunate events (e.g. failed recordings of the tutorial session) and a lack of experience in creating semantic resources in some parts of the community. The NETMAR project has taken a SKOS-based approach to semantics within ICAN. As a proof of concept, a mapping between the Oregon State CWA mark-up vocabularies and a global set of mark-up terms defined by ICAN has been published at http://vocab. nerc.ac.uk/scheme/ICANCOERO/. A schematic of this approach is shown in Figure 3. This approach also has the benefit of being able to map into existing ICAN resources served by MMI ORR through ontology extension. A tutorial cookbook document has been created for the ICAN community which attempts to address some of the issues the community previously faced in this activity (Dunne, Leadbetter, and Lassoued 2012).

Client layer

In addition to these machine-to-machine interactions, the NETMAR project has also built a number of client layer applications on top of NVS2.0. These applications allow human users to interact with the underlying semantic resource. The simplest of these is a XSLT style sheet over the SKOS concept schemes, which allows users to browse the SKOS concept hierarchy (Figure 4). This renders the underlying SKOS eXtensible Markup Language (XML) as HyperText Markup Language (HTML) within a web browser, the downside being that if a human user wishes to read the XML they must select the 'view source' option in their browser. The advantage is that one URL provides a human- and a machine-readable form of the hierarchy. A more complex client is built on top of the NETMAR semantic resource and extends the OGC Catalogue Service mediator presented by Lassoued et al. (2008). This client (Figure 5) allows the user to both browse and search the semantic resource. Once a term of interest has been located, the user may then discover and download datasets tagged with semantically related terms.

Service validation

The NETMAR project also aims to bring semantic mark-up to WPSs to assist in chaining and orchestration of these services (de Jesus et al. 2012). For a service chain to be semantically valid, streams connected to service inputs must contain the correct measured phenomenon in the correct units of measure. The NETMAR semantic
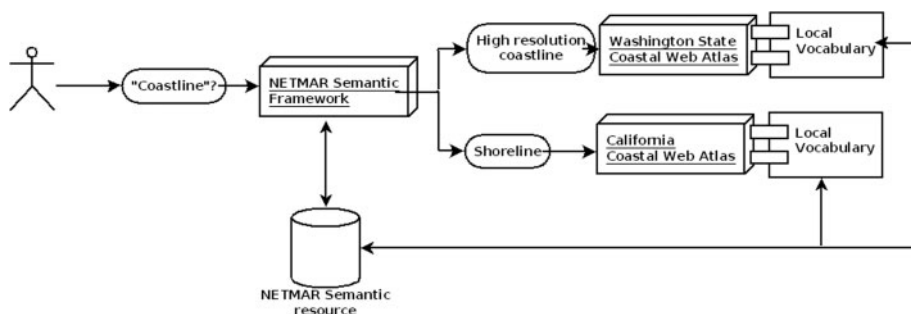

Figure 3. An illustration of a simple semantically enabled CWAs search mediator. A fully operational system would incorporate many more CWA nodes than the two illustrated here.

NETMAR Oceanography Thesaurus

Thesaurus centred on parameter descriptions from the BODC PUV and CF Standard Names mapping to parameter discovery, units, instrument, platform, project and coverage concepts. Underpins discovery and semantic service chain validation for the NETMAR sea ice and oceanography pilots.

Created by: NETMAR Ontology Governance Group

Published by: Natural Environment Research Council

instrument(+)
project(+)
parameter(+)
    Definition: Keyword identifies a phenomenon or group of phenomena in the dataset

    Alternative concept labels: parameter

    discipline(+)
        Definition: Keyword identifies a branch of instruction or specialized learning

        Alternative concept labels: discipline

        Terrestrial(+)
        Chemical oceanography(+)
        Biological oceanography(+)
        Fisheries and aquaculture(+)
        Environment(+)
        Cryosphere(+)
        Physical oceanography(+)
        Atmosphere(+)
        Marine geology(+)
    theme(+)
platform(+)
platform_class(+)
vertical_coverage(+)

Figure 4. Browsing the NETMAR semantic resource through an XSLT style sheet.

resource includes measured phenomenon concepts of fine (http://vocab.nerc.ac.uk/collection/P01/current/) and coarse (http://vocab.nerc.ac.uk/collection/P25/current/) granularity plus both units of measure concepts (http://vocab.nerc.ac.uk/collection/P06/current/) and their related physical dimensions (http://vocab.nerc.ac.uk/collection/P24/current/).

If the streams are tagged with fine-grained phenomena plus their units of measure and the inputs tagged with coarse-grained phenomena plus their units of measure, a 'traffic light' semantic checker (Figure 6) may be built. This does a simple check to ensure that the stream and input units of measure are an exact match and uses the semantic resource to 'smart check' that the stream fine-grained phenomenon matches the input coarse-grained phenomenon. The result is either red for mismatch or green for a match. Amber results from a unit of measure check mismatch that is identified by a 'smart check' using the semantic resource as a match in physical dimensions (e.g. metres and millimetres that share the dimension length). Amber check fails may be rectified by the inclusion of a unit conversion service into the workflow, whereas a red check fail indicates a more serious workflow design error. This tool may be embedded inside a workflow editor or orchestration engine so that compatibility between WPSs may be established prior to executing a workflow.

Conclusions

This paper has presented the development of an interconnected semantic resource which has been used to facilitate data, metadata and web service discovery. It has also
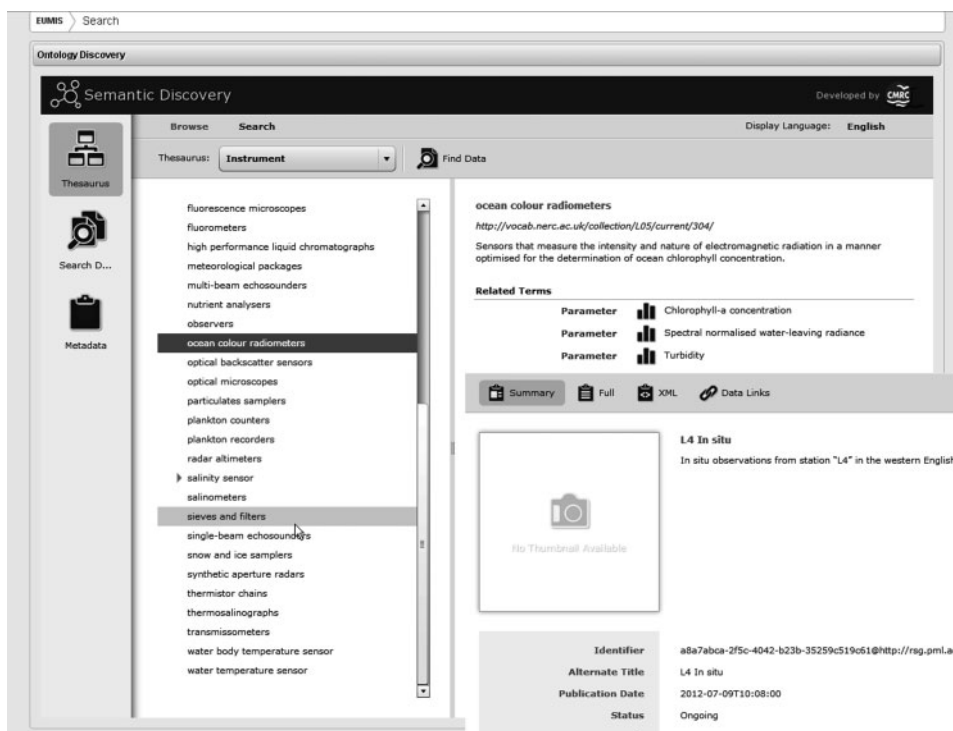
Figure 5. The NETMAR semantically enabled data browse and search interface.

been used in the chaining and orchestration of WPSs and the interconnection of CWAs across jurisdictional boundaries. In particular, the latter of these use cases has demonstrated the need to provide users with comprehensive instructions and simple tools in order to use semantic web resources. The cookbook produced for the ICAN community has provided the instructions, but unfortunately the tool development remains to be completed. The user community's perception, rightly or wrongly, as to the level of technical expertise required to enter into previous efforts also highlights the emerging notion that human, social networks are of equal importance to digital networks in environmental science informatics applications (Allison et al. 2012). Developing cookbook, or similar, approaches with a low barrier of technical expertise to use a system promotes the use of digital earth technologies in non-specialist fields.

Benefits to the wider digital earth community from the development of inter-connected semantic resources include the increased ability easily to share and disseminate data as the World Wide Web becomes more semantic through projects such as schema.org (http://schema.org/). Indeed, digital earth projects such as SeaDataNet (Schaap and Lowry 2010) already use the content of NVS2.0 in order to facilitate interoperability between data providers. The field of geosemantics, that is the formal representation of geospatial knowledge to advance geospatial interoperability, is growing and as such now has a working group at the Open Geeospatial Consortium. Tools and methods such as those outlined in this paper have the potential to support these efforts.

There remains, however, the potential for future developments to the semantic resource described within this paper. These would begin with the incorporation of mappings from deprecated terms to their replacements within payload documents.
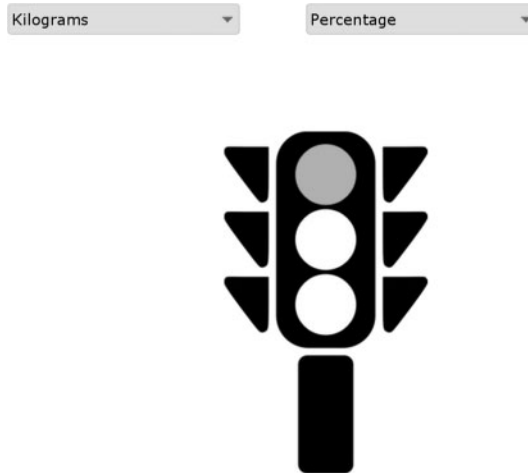
Figure 6. A semantically aware tool for checking the compatibility of input units to an OGC WPS.

The further addition of the type attribute from the OWL XML presentation syntax will allow member concepts of the ocean biogeochemistry ontology (Chandler et al. 2011) to be declared. This additional attribute would also allow for top level concepts to be related to upper level ontologies such as the Basic Formal Ontology (Grenon and Smith 2004) or the General Formal Ontology (Herre et al. 2006) to further the semantic interoperability of the content served by NVS2.0.

The ICAN thesaurus will be enhanced by the addition of mappings to more nodes of the network. The next nodes to be targeted include the Marine Irish Digital Atlas (http://mida.ucc.ie/). This additional content within the thesaurus will empower the ICAN CSW mediator to begin to be demonstrated with full interoperability across international borders.

To further enhance the semantic checking of WPS compatibility, the possibility of semantically encoding the conversations between units will be investigated. This may be beyond the scope of a simple SKOS approach, although there is a precedent in NVS2.0 resources for encoding structured information as JavaScript Object Notation strings within the concept definition. This work would enable the removal of the amber light from the traffic light unit checker, as a semantically aware WPS would be able to use the encoded conversion factor to change the input units to its desired units.

As well as providing the semantic support to the NETMAR project described above, NVS2.0 will be used to support the EU FP7 project SeaDataNet-II and to host vocabularies for some of the EU INSPIRE spatial data themes. Due to the large number of concepts and collections served by NVS2.0, many of the projects which use its content cache the collections in which they are interested on a daily basis. In support of this use case, a web syndication feed, based on Atom, is being developed so that content harvesters may subscribe to concept collection updates.

Coastal and Marine Research Centre, University College Cork, Ireland. The Traffic Light Unit Checker was developed by Jorge de Jesus at the Plymouth Marine Laboratory, United Kingdom.

References

Allison, M. Lee, Rob Atkinson, David Arctur, Simon Cox, Ian Jackson, Stefano Nativi, and Lesley Wyborn. 2012. "International Convergence on Geoscience Cyberinfrastructure." *Geophysical Research Abstracts* 14: EGU2012-13563-1. http://meetingorganizer.copernicus.org/EGU2012/EGU2012-13563-1.pdf

Altheim, Murray. 2008. "Informal Ontology Design: A Wiki-Based Assertion Framework." In *Proceedings of Balisage: The Markup Conference 2008*. Balisage Series on Markup Technologies, Vol. 1.

Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. "The Semantic Web." *Scientific American* 284 (5): 34–43. doi:10.1038/scientificamerican0501-34.

Biébow, Brigitte, and Sylvie Szulman. 1999. "TERMINAE: A Linguistics-Based Tool for the Building of a Domain Ontology." *Lecture Notes in Computer Science* 1621: 49–66.

Bizer, Christian, Tom Heath, and Tim Berners-Lee. 2009. "Linked Data – The Story so Far." *International Journal on Semantic Web and Information Systems* 5 (3): 1–22. doi:10.4018/jswis.2009081901.

Booth, David, Hugo Haas, Francis McCabe, Eric Newcomer, Michael Champion, Chris Ferris, and David Orchard. 2004. "Web Service Architecture." World Wide Web Consortium. February 11. http://www.w3.org/TR/2004/NOTE-ws-arch-20040211/

Box, Don, David Ehnebuske, Gopal Kakivaya, Andrew Layman, Noah Mendelsohn, Henrik Nielsen, Satish Thatte, et al. 2000. "Simple Object Access Protocol (SOAP) 1.1." World Wide Web Consortium. May 8. http://www.w3.org/TR/2000/NOTE-SOAP-20000508/

Chandler, Cynthia, Molly Allison, Robert Groman, Patrick West, Stephan Zednick, and Andrew Maffei. 2011. "Developing an ontology for ocean biogeochemistry data." *Geophysical Research Abstracts* 13: EGU2011-8578. http://www.w3.org/TR/2000/NOTE-SOAP-20000508/

Coene, Yves, Hung Truong-Minh and Yassine Lassoued. 2009. "Discovery Standards in HMA-T and Discovery in FP6 InterRisk". Paper presented at Ontology and Discovery Workshop, ESRIN, Frascati, Italy, March 4.

Cox, Simon. 2012. "Best Practice in Formalizing a SKOS Vocabulary." Solid Earth and Environment GRID, Commonwealth Scientific and Industrial Research Organisation (CSIRO). October 29. https://www.seegrid.csiro.au/wiki/Siss/VocabularyFormalizationInSKOS

de Jesus, Jorge, Peter Walker, Michael Grant, and Steve Groom. 2012. "WPS Orchestration Using the Taverna Workbench: The eScience Approach." *Computers & Geosciences* 47: 75–86. doi:10.1016/j.cageo.2011.11.011.

Dunne, Declan, Adam Leadbetter, and Yassine Lassoued. 2012. "ICAN Semantic Interoperability Pilot Cookbooks." NETMAR: Open Service Network for Marine Environmental Data. European Commission within the Seventh Framework Programme (2007–2013) Theme ICT 2009.6.4 Project 249024. Deliverable 7.9.2. http://netmar.nersc.no/sites/netmar.nersc.no/files/D7.9.2_ICAN_semantic_cookbooks_r2_20120731_0.pdf

Fielding, Roy, and Richard Taylor. 2002. "Principled Design of the Modern Web Architecture." *ACM Transactions on Internet Technology* 2 (2): 115–150. doi:10.1145/514183.514185.

Graybeal, John, Anthony Isenor, and Carlos Rueda. 2012. "Semantic Mediation of Vocabularies for Ocean Observing Systems." *Computers & Geosciences* 40: 120–131. doi:10.1016/j.cageo.2011.08.002

Grenon, Pierre, and Barry Smith. 2004. "SNAP and SPAN: Towards Dynamic Spatial Ontology." *Spatial Cognition and Computation* 4 (1): 69–104. doi:10.1207/s15427633scc0401_5.

Gruber, Tom. 2009. "Ontology." In *Encyclopedia of Database Systems*, edited by Ling Liu and M. Tamer Özsu, 1963–1965. Berlin: Springer-Verlag.

Herre, Heinrich, Barbara Heller, Patryk Burek, Robert Hoehndorf, Frank Loebe, and Hannes Michalek. 2006. *General Formal Ontology (GFO): A Foundational Ontology Integrating Objects and Processes. Part I: Basic Principles.* Leipzig: Research Group Ontologies in Medicine (Onto-Med).

IBM. 1993. *IBM Dictionary of Computing.* New York: McGraw-Hill.

Jin, Zhi, David Bell, F. George Wilkie, and D. Leahy. 1998. "Automatically Acquiring Requirements of Business Information Systems by Reusing Business Ontology." In *Proceedings of Workshop on Applications of Ontologies and Problem-Solving Methods, 13th European Conference on Artificial Intelligence*, 78–87.

Kumpati, Murari. 1985. Patent description. US Patent 4774661, filed November 19.

Lassoued, Yassine, Dawn Wright, Luis Bermudez, and Omar Boucelma. 2008. "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.*

Latham, Sue, Raymond Cramer, Michael Grant, Philip Kershaw, Bryan Lawrence, Roy Lowry, Dominic Lowe, et al. 2009. "The NERC DataGrid Services." *Philosophical Transactions of the Royal Society A* 367: 1015–1019. doi:10.1098/rsta.2008.0238.

Lawrence, Bryan, Roy Lowry, Peter Miller, Helen Snaith, and Andrew Woolf. 2009. "Information in Environmental Data Grids." *Philosophical Transactions of the Royal Society A* 367: 1003–1014. doi:10.1098/rsta.2008.0237.

Leadbetter, Adam, Oliver Clements, and Roy Lowry. 2011. "Emerging Standards in Vocabulary Server Access Methods." *Geophysical Research Abstracts* 13: EGU2011–2143.

Leadbetter, Adam, Torill Hamre, Roy Lowry, Yassine Lassoued, and Declan Dunne. 2010. "Ontologies and Ontology Extension for Marine Environmental Information Systems." In *Proceedings of ENVIP 2010, CEUR Workshop Proceedings 679.* Aachen: Sun SITE Central Europe.

Lowry, Roy, and Geoff Williams, forthcoming. "Putting Meaning into SeaDataNet." Mediterranean Marine Sciences.

Ma, Xiaogang, Emmanuel John Carranza, Chonglong Wu, Freek van der Meer, and Gang Liu. 2011. "A SKOS-Based Multilingual Thesaurus of Geological Time Scale for Interoperability of Online Geological Maps." *Computers & Geosciences* 37 (10): 1602–1615. doi:10.1016/j.cageo.2011.02.011.

McGuinness, Deborah. 2003. "Ontologies Come of Age." In *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, edited by Dieter Fensel, James Hendler, Henry Lieberman, and Wolfgang Wahlster, 171–196. Cambridge, MA: MIT Press.

Miles, Alistair, and Sean Bechhofer. 2009. "SKOS Simple Knowledge Organization System Reference". World Wide Web Consortium. August 18. http://www.w3.org/TR/2009/REC-skos-reference-20090818/

Navigli, Roberto, and Paola Velardi. 2008. "From Glossaries to Ontologies: Extracting Semantic Structure from Textual Definitions." In *Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, edited by Paul Buitelaar and Philipp Cimiano, 71–87. Amsterdam: IOS Press.

O'Dea, Elizabeth, Edward Dwyer, Valerie Cummins, and Dawn Wright. 2011. "Potentials and limitations of Coastal Web Atlases." *Journal of Coastal Conservation* 15 (4): 607–627. doi:10.1007/s11852-011-0150-7.

Orlandi, Fabrizio, and Alexandre Passant. 2011. "Modelling Provenance of Dbpedia Resources Using Wikipedia Contributions." *Web Semantics: Science, Services and Agents on the World Wide Web* 9 (2): 149–164. doi:10.1016/j.websem.2011.03.002.

Ram, Sudha, and Jun Liu. 2006. "Understanding the Semantics of Data Provenance to Support Active Conceptual Mapping." *Lecture Notes in Computer Science* 4512: 17–29.

Ram, Sudha, and Jun Liu. 2009. "A New Perspective on Semantics of Data Provenance" In *Proceedings of First International Workshop on the role of Semantic Web in Provenance Management, CEUR Workshop Proceedings*, Vol. 526. Aachen: Sun SITE Central Europe.

Schaap, Dick, and Roy Lowry. 2010. "SeaDataNet – Pan-European Infrastructure for Marine and Ocean Data Management: Unified Access to Distributed Data Sets." *International Journal of Digital Earth* 3 (Supplement 1): 50–69. doi:10.1080/17538941003660974

Shadbolt, Nigel, and Tim Berners-Lee. 2008. "Web Science Emerges." *Scientific American* 299 (4): 76–81. doi:10.1038/scientificamerican1008-76.

Smith, Barry, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis Goldberg, et al. 2007. "The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration." *Nature Biotechnology* 25 (11): 1251–1255.

van Renssen, Andries. 2005. "Gellish, A Generic Extensible Ontological Language." PhD diss., Delft University of Technology.

W3C. 2004. "Resource Description Framework (RDF)". World Wide Web Consortium. February 10. http://www.w3.org/RDF/

Warner, Amy. 2002. "A Taxonomy Primer". School of Information, University of Texas at Austin. http://www.ischool.utexas.edu/ ~i385e/readings/Warner-aTaxonomyPrimer.html

Wright, Dawn, Yassine Lassoued, Luis Bermudez, Tim Nyerges, Tanya Haddad, and Ned Dwyer. 2008. "Semantic Mediation as a Gateway to Interoperability, with a Case Study of the International Coastal Atlas Network (ICAN)." In *Geographic Information Science: 5th International Conference, GIScience 2008, Park City, Utah, September 23–26, Extended Abstracts*, edited by Thomas Cova, Harvey Miller, Kate Beard, Andrew Frank, and Michael Goodchild, 201–218. Berlin: Springer.