A SKOS approach to semantically enabling the International Coastal Atlas Network

Abstract. The International Coastal Atlas Network (ICAN) is an informal group of organisations that have been meeting since 2006 to specify and implement data interoperability approaches to Coastal Web Atlases (CWAs). Two of the goals to improve CWA interoperability are the development of an ontology to provide mappings between keywords used to markup local Coastal Web Atlases and to develop a CWA search mediator capable of performing a smart search using these mappings. ICAN semantic development began with the assembly of a Web Ontology Language (OWL) based coastal erosion ontology in 2007 with some work to convert local level controlled vocabularies to OWL following. However, little mapping work followed and the possibility was raised that OWL and associated tools provided too great a technological barrier for the ICAN community to enter into this activity. This paper presents an overview of potential resources which can be used by ICAN and an alternative approach to creating semantic content and relationships for the ICAN community, based on the less verbose Simple Knowledge Organization System (SKOS).

Keywords: Vocabularies, thesauri, semantics, coastal, atlas

1. Introduction

The International Coastal Atlas Network (ICAN)\(^1\) is an informal group of organisations that have been meeting since 2006 to specify and implement data interoperability approaches to coastal web atlases (CWAs). In recent years significant momentum has occurred in the development of internet resources for decision makers, scientists and the general public who are interested in the coast. A key aspect of this trend has been the development of CWAs, based on web enabled geographic information systems. A CWA has been defined by the ICAN community as: a collection of digital maps and datasets with supplementary tables, illustrations and information that systematically illustrate the coast, oftentimes with cartographic and decision support tools, all of which are accessible via the Internet.

The strategic aim of ICAN is to share experiences and to find common solutions to CWA development (e.g. user and developer guides, handbooks and articles on best practices, information on standards and web services, expertise and technical support directo-
ries, education, outreach, and funding opportunities) while ensuring maximum relevance and added value for the end-users.

The application scope of CWAs is broad. Driving factors for CWA development include the need for:

- Better planning to cater for increased population pressures in the coastal zone (e.g. the UN estimate that by 2020 75% of the world’s population will be living within 60 km of the coastal zone [6])
- Decision support systems in relation to climate change scenarios in vulnerable coastal regions
- Information to facilitate assessments of risk to natural hazards (including tsunamis and floods)
- Access to data and maps to support both integrated coastal zone management and marine spatial planning (MSP) as a tool for better coastal and marine area management
- More efficient and effective coastal and marine area governance including access to relevant data and information
- Information on resource availability and exploitation including habitat and species information, as well as ecological and community resilience

The ICAN community aims to make ongoing atlas developments more interoperable in order to better support data discovery, visualisation, and access across borders. This includes smart search where, for example, a user search keyword of "water" will discover appropriate datasets flagged with "seawater", "freshwater". Current data discovery, access and visualisation of atlas datasets is limited and fragmented. Users must individually search standalone atlases. Atlas developers require guidelines to support improved interoperability and connectivity between atlases, both legacy and new. Improved atlas interoperability will be achieved by:

- Publication of metadata, which includes vocabulary keywords, from each atlas node using standard Open Geospatial Consortium (OGC) Catalogue Service for the Web (CSW)
- Development of an informal ontology that provides mappings between the vocabulary keywords of atlases
- Development of an atlas search mediator which is capable of performing smart discovery on top of standard CSW using ontology mappings (Figure 1)
- Subsequent visualisation and/or download of discovered datasets using established standards such as OGC Web Map Services, Web Feature Services and Web Coverage Services

The problems demonstrated in this scenario are:

- Improved interoperability between atlases across borders
- Creating a CSW smart search mediator which uses ontology for searching based on keyword semantics

This paper presents an overview of the content of a group of existing semantic resources which contain knowledge of interest to the ICAN community and guidelines developed for the those in the ICAN community wishing to deploy semantic resources to describe their Coastal Web Atlases or to connect their local atlases to existing semantic resources. This is achieved within the semantic framework developed for the EU FP7 NETMAR project.

1.1. Earlier work

The development of ICAN semantics began with the assembly of a Coastal Erosion ontology to support semantically enabled atlas layer discovery following the workshop at Corvallis in 2007. Semantically enabled, or "smart", discovery is where terms supplied by the search client are used to locate metadata that has been marked up using different but semantically related terms. A much quoted example is a search that returns datasets tagged "rainfall" for a search term of "precipitation".

This work was based on strategies and technologies provided by the Marine Metadata Interoperability (MMI) initiative. The strategy adopted was to take local vocabularies and ontologies, convert them into Web Ontology Language (OWL) classes using tools like Protégé or TopBraid Composer and then map them

---

1http://www.opengeospatial.org/
2http://www.opengeospatial.org/standards/wms
3http://www.opengeospatial.org/standards/wfs
4http://www.opengeospatial.org/standards/wcs
5https://marinemetadata.org/
Fig. 1. A use case diagram illustrating a simple version of semantically enabled Coastal Web Atlas search mediation. A fully operational system would incorporate more than the two local Coastal Web Atlas nodes illustrated here.

to a global ontology, again comprising OWL classes, using the MMI VINE tool.

Some initial work, the conversion of local vocabularies from the Irish MIDA and Oregon coastal atlases into OWL, was completed but very little mapping work was done subsequently. The possibility was mooted that the OWL-based approach, particularly some of the tools, created a technological barrier that ICAN community members were unwilling to surmount. The work described below surveys other potential resources and provides a possible alternative solution, at a lower level of semantic complexity.

2. Existing resources

2.1. GEneral Multilingual Environmental Thesaurus (GEMET)

GEMET is a repository maintained by the European Environment Information and Observation Network (EIONET) using the procedural model specified in International Organization for Standardization (ISO) standard 19135 "Geographic information – Procedures for item registration". Consequently, it has the potential to host and serve multiple semantic resources (termed registers) maintained by separate content governance authorities under common technical governance.

At the time of writing there were two registers: the GEMET concept thesaurus and the Infrastructure for Spatial Information in Europe (INSPIRE) spatial data themes.

A large number of concepts (just over 6500 in the 2004 version) are mapped to these themes and interlinked by hierarchical relationships (narrower and broader). The emboldened themes in the above include some concepts that are considered to be relevant ICAN. All themes, concepts and definitions have been translated into 28 languages other than English.

The concepts have also been classified by mapping them to a set of 30 group terms, identified through presentation in upper case, that in turn map to 3 supergroups. For example, the "oceanography" concept is mapped to the "RESEARCH" and "SCIENCES" groups. A full list of the groups and super-groups is given at http://www.eionet.europa.eu/gemet/about.

The semantic scope of GEMET is heavily skewed towards the European Environment Agency’s area of interest of which the marine environment forms a small part, although there is a larger overlap with the coastal atlas domain. The mapping (like every mapping) is incomplete. For example, the concept of "marine biology" is totally divorced from that of "oceanography".

The INSPIRE themes are held in a separate register to the concept thesaurus (accessed through the "INSPIRE Spatial Data Themes" tab on the web interface) as a flat controlled vocabulary (keys, terms and definitions) with no mapped concepts, internal linkages or mappings to external vocabularies or thesauri. The entries are provided in 20 languages.

2.2. Marine Metadata Interoperability (MMI) Ontology Repository & Registry (ORR)

Marine Metadata Interoperability Ontology Registry and Repository (MMI-ORR) [3] is a semantic framework based on the ISO19135 model that allows registered users to create, upload and edit controlled vocabularies, thesauri and ontologies. Each resource is documented with provenance and usage metadata.
Once uploaded the resources are open to anybody for browsing and download, unless flagged as a test resource in which case they are only visible to the originator. Each resource is given a URL that may be extended RESTfully into a concept URL through concatenation of the subject label. Both list and term URLs can sometimes resolve to the same XML document.

Note that MMI-ORR is a registry as well as a repository. For example, there is an entry for the INSPIRE glossary that is served from the INSPIRE registry at the EC Joint Research Centre in Italy.

At the time of review latest version of the framework contained on the order of 200 resources, of which the following were considered of relevance to ICAN

- Oregon Coastal Atlas-ICAN Mapping. A demonstration, and therefore sparsely populated, mapping between the Oregon Coastal Atlas and the International Coastal Atlas Network (ICAN) global term list
- Oregon Coastal Atlas Thesaurus. The terms from the Oregon Coastal Atlas that were used in the demonstration mapping are presented in a thesaurus structure
- ICAN demonstration thesaurus. The global term list used to provide the overarching ontology for the ICAN pilot demonstrator mappings
- Oregon Coastal Atlas Draft Erosion Ontology. A well populated collection of terms relevant to coastal spatial datasets
- Coastal Erosion Topics. Classes representing coastal erosion topics used by the ICAN community

2.3. Global Change Master Directory (GCMD) keywords

The National Aeronautics and Space Administration (NASA) GCMD (at the time of writing) publishes the following controlled vocabularies with the currency dates as shown.

- Science Keywords 2008-02-05
- Services Keywords 2009-10-01
- Data Centers 2009-09-22
- Projects 2009-06-10
- Instruments 2009-06-10
- Platforms 2009-06-10
- Locations 2009-12-22
- Data Resolution 2008-02-05
- Chronostratigraphic Units 2008-02-05
- URL Content Types 2009-06-10

Three of these (Science Keywords, Instruments and Platforms) have significant interest for ICAN semantics. However, the Data Resolution list may have relevance to the description of uncertainty and the URL Content Types list may be useful for marking up service bindings.

The Platforms (things that carry instruments) and Instruments (things that make measurements) lists are served as flat term lists containing a name plus an optional abbreviation. The terms vary in granularity and are obviously forming a hierarchical classification (e.g. "AIRCRAFT" and "Airbus A340-600") but no linkages are included in the published version.

The Science Keywords are an extensive list of parameter terms aimed at data discovery (i.e. metadata keyword population) over a wide range of science domains. This can be clearly seen from the topics that form the top level of the keyword hierarchy.

- Agriculture
- Atmosphere
- Biosphere
- Biological Classification
- Climate Indicators
- Cryosphere
- Human Dimensions
- Land Surface
- Oceans
- Paleoclimate
- Solid Earth
- Spectral/Engineering
- Sun-earth Interactions
- Terrestrial Hydrosphere

There are four hierarchical levels below this: term, variable_level_1, variable_level_2 and variable_level_3. The entries are published as a list of hierarchical chains such as 'Biological Classification > Animals/Invertebrates > Arthropods > Chelicerates > Arachnids' for spiders and not as a network of interrelated terms.
2.4. United States Geological Survey (USGS) thesaurus

The USGS thesaurus is a hierarchical concept infrastructure specifically intended to help people outside USGS find information on USGS web sites without specific knowledge of the organisational structure and operations of the USGS. It is based on a faceted design with the following top concepts:

- Sciences: major educational fields, fields of study, and professional groupings within USGS. This is a well-populated taxonomy covering disciplines within earth (including ocean), engineering, information, life, planetary and social sciences.
- Methods: techniques, methods, procedures, or strategies for research, management, collection, or analysis of scientific information in USGS. These are heavily biased towards the geological sciences.
- Topics: themes, subjects, and concerns for which USGS information resources are relevant. This taxonomy covers a wide range of subjects at variable levels of depth. Some parts are populated to the stage of completeness, such as the "elements" taxonomy, whilst others, such as "ocean processes" and "coastal processes" have sparse coverage.
- Product types: general representation of the information in a resource, such as a map or data set
- Time periods: geologic time periods and seasons of the year. There is a comprehensive hierarchy of geological time periods, although the US names are used such as Mississippian for Lower Carboniferous. In contrast, in the seasons section "autumn" appears rather than "fall".
- Institutional structures and activities: USGS activities, processes, and organizational concepts relevant to USGS itself.

The concepts have definitions and in some cases synonyms. They are inter-linked through a network of "narrower than", "broader than" and "related to" relationships.

2.5. Geoscience Markup Language (GeoSciML) vocabularies

The GeoSciML vocabulary server presents three collections of vocabularies and thesauri designated as versions "CGI200811", "CGI201001" and "1GE201001". CGI200811 contains resources that are a mixture of flat lists and hierarchical thesauri. The content as might be expected is tightly focussed on the description of geological units, including rock type classifications, descriptions of the geological object geometry and geological processes. In addition there are two general lists describing value qualifiers and semantic relationships between vocabulary qualifiers. Version CGI2010001 is a development of these resources with a noticeable development of thesauri.

The 1GE2010 branch contains 16 objects that again cover geological subjects such as lithology, contact types, geoscience methodology and geologic unit morphology and spatial geometry. Many of these are multilingual (18 languages) thesauri. The "1GE_Ages.xml" object is a multilingual thesaurus of geological ages that appears to be developed from the USGS thesaurus.

2.6. Summary

In order to meet ICAN’s requirements of interoperable semantics and a semantically aware CSW, named and typed relationships between the to the pre-existing content described above and the resources used to label CWAs are required. In many cases this requires that the providers of CWAs publish their controlled vocabularies in some way. Only MMI-ORR of the resources described above provides this functionality. The NETMAR project has taken on a sixth resource, the NERC Vocabulary Server, as the basis of its semantic framework due to its large community user base and the breadth of methods available in its Application Programming Interface.

3. International Coastal Atlas Network cookbook

Following the ICAN workshop in Copenhagen in 2008, it was decided to provide technical training for coastal web atlas developers in the form of cookbooks, especially for people who are starting a new atlas. While [7] covers the more theoretical aspects of coastal web atlas development, the cookbooks are targeted to contain more practical step-by-step instructions. Consequently, the NETMAR project has produced cookbooks to contribute to this technical training material.

The following text forms the ICAN "Introduction to semantics" cookbook for CWA authors wishing to use semantic technologies and resources to describe and markup their CWAs.
3.1. Introduction to semantics

"If HTML and the [World Wide] Web made all the online documents look like one huge book, [semantics] will make all the data in the world look like one huge database" [1]

If data in a distributed system are to be understood elsewhere in that system, or externally to the system, they must be labelled (or "marked up") accordingly. Either the mark up used throughout the system must use a common set of phrases [5], or there must be a means of translating between the phrases used at different points of the system, using common "semantics". The aim of the "semantic web" is to provide these consistent phrases and to define the relationships in a formal manner, resulting in what is often called a "knowledge organization system".

This section provides a tutorial for those who wish to investigate and make use of these technologies, aimed specifically at members of the International Coastal Atlas Network community and more generally at scientists and data managers.

3.2. Why use a knowledge organization system?

One scenario for using knowledge organization systems in the International Coastal Atlas Network (ICAN) is to search through the local atlases for a given data keyword from a central portal. For example, as illustrated below, a user arrives at the ICAN portal and request "coastline" data. The portal software is connected to a global knowledge organization system which is aware that "coastline" is related to both "shoreline" and "high resolution coastline". The user request and this information from the global knowledge organization system are then passed on to the local atlases which search on "coastline", "shoreline" and "high resolution coastline". The local atlases then return the relevant data to the portal and then to the user. This is an implementation of so-called "smart-search" [4].

Other uses of knowledge organization systems include populating metadata elements with standardized content which can be verified and validated by software services; dynamically populating drop down lists in websites and software applications; dynamically moving a metadata record from one metadata scheme to another; and the validation of input parameters and their associated units in Open Geospatial Consortium Web Processing Services.

3.3. What are vocabularies, thesauri and ontologies?

Knowledge organization systems fall broadly into three groups: vocabularies, thesauri and ontologies. These three groups show increasing complexity in their structure.

A vocabulary can be either a list of terms or a list of terms and some text providing a definition of the term. A vocabulary ensures that terms are used, and spelt, consistently. A vocabulary can be extended in its power by providing definitions of concepts. Thesauri expand the knowledge contained within a vocabulary by adding information about the relationships between the terms of the vocabulary. These relationships fall broadly into three categories:

- Synonyms - the current term is synonymous with a given, different term. e.g. "dogs" is synonymous with "canines"
- Broader relations - the current term has a more specific definition than a given different term. e.g. "dogs" has a broader relationship to "pets"
- Narrower relations - the current term has a less specific definition than a given different term. e.g. "dogs" has a narrower relationship to "terriers"

In a more complex thesaurus, the concepts at the top of the hierarchy of broader and narrower relations may be stated explicitly, rather than being inferred by software agents. This provides the simplest form of a formal ontology. A well known example of this form is the Yahoo! web directory or the categorisation of auctions on the eBay homepage. eBay has terms such as "Antiques", "Coins" and "Sporting Goods" as the top level in its hierarchy. Narrower terms sit below these, for example "Sporting Goods" contains "Football", "Golf" and "Sailing". These terms sit above those which are narrower still, "Sailing" having such narrower terms as "Clothing & Shoes", "Life Jackets" and "Rope". In the context of environmental sciences, the Global Change Master Directory can be seen to work in this way. For example, "Ocean" is at the top level, with "Coastal Processes" beneath it and terms such as "Beaches" and "Coastal Elevation" beneath that.

These more complex thesaurus also introduce a fourth category of relationship between concepts, that of a "loose relationship". That is where two terms have a relationship that is not of the broader or narrower type or a synonymous relationship, e.g. "domesticated dogs".

\[http://gcmd.nasa.gov/\]
are "loosely related" to "wild dogs". These loose relationships may allow different pathways to the discovery of a term, making the resource what is known as "orthogonal". For example, eBay has "Walking, Hiking, Trail" in its "Fashion" auction categories and "Boots & Shoes" in its "Sporting Goods" auction categories. If these two were loosely mapped a search for "walking boots" could yield auction results from both categories.

More complex ontologies can be created through declaring a term to belong to a particular class, the addition of property information to the term and the restriction of values data associated with the term may take. For example, if eBay defined the class of "auction" particular individual terms belonging to the "auction" class could be "English auction", "blind auction" or "Dutch auction".

3.4. Defining the content of a knowledge organization system

3.4.1. Can I reuse existing resources?

Where possible it is best to make use of existing knowledge organization systems. This increases the ability to reuse data across systems, known as interoperability. If the reuse of existing systems is not an option, the section below explains how to generate a new knowledge organization system. Any new system should have some specified relationships to an existing system to promote interoperability and flexibility. Details of how to access an existing knowledge organization relevant to the International Costal Atlas Network are provided below.

3.4.2. What is the scope of the knowledge organization system?

While it might be tempting to want to describe and define every imaginable concept in a new knowledge organization system, this would be a very time consuming and frustrating process, and would not make best use of other, pre-existing resources. Instead, it is much better to take the time to identify the specific domain that needs to be described by the terms you wish to define, for example coastal erosion, or names and extents of beaches. In this way work in building the knowledge organization system is tightly defined and the content is coherent, well understood and should not replicate existing resources.

3.4.3. Identifying the content

The challenge of integrating data and information of different kinds at different levels of detail is well defined in computer science literature [2]. In the area of semantics on the World Wide Web, the level of detail a term can describe is known as its granularity. For a given level of a knowledge organization system the definitions of a term may be as broad or as narrow as is necessary, as long as they are not ambiguous.

However, when building a hierarchical thesaurus, it is important that concepts defined at the same level of the hierarchy maintain a similar degree of granularity. If the thesaurus is imagined as a pyramid, making a concept at a given level too narrow or broad in its definition is like placing a too small or too large brick in the wall of the pyramid, and makes the structure unstable. For example, "body of water" should not sit at the same level as "lake" or "reservoir", as these are terms with a narrower relationship or a finer granularity.

As described above, the definition of terms by themselves is useful but the impact of the work can be greatly extended by providing relationships which link the terms together to form networks of knowledge. This enhances the ability of a user to find data labelled with a given term or to translate the metadata from one mark up scheme to another. Relationships can be thought of simply as broader and narrower (for example, in the diagram below the BODC Parameter Discovery Vocabulary is narrower than the SeaDataNet Agreed Parameter Groups and vice versa); loosely related (the BODC Parameter Usage\(^{15}\) and MEDATLAS Parameter Usage\(^{16}\) vocabularies are of similar granularity and are linked this way); and synonyms where two terms may be used interchangeably.

3.5. Ensuring the quality of the content of the Knowledge Organization System

There are two aspects to providing quality assurance, or governance, for a knowledge organization system. The first is to ensure the quality of the content of the knowledge organization system. This includes the names and definitions of terms and the relationships between the terms. A well tested mechanism for managing content governance is setting up an e-mail list of interested parties on which requests for new terms and mappings can be discussed. This is the model which has been implemented by: the Climate and Forecast\(^{17}\) netCDF metadata conventions group; the SeaDataNet and MarineXML Vocabulary Content Gov-

\(^{15}\)http://vocab.nerc.ac.uk/collection/P01/current/
\(^{16}\)http://vocab.nerc.ac.uk/collection/P09/current/
\(^{17}\)http://cf-pcmdi.llnl.gov/
ernance Group (SeaVoX); and the NETMAR ontology governance body. The role of the content governance group is analogous to the International Organization for Standardization (ISO) definition of a "control body".

The second aspect is assuring the technical quality of the system. This includes ensuring that the knowledge organization system is available with the greatest possible up-time; the representation of the system is valid in the chosen scheme (e.g. extensible markup language, XML); and the various versions of the concepts, collections and scheme are maintained and accessible. For example, within the NETMAR project this technical governance is provided by the British Oceanographic Data Centre as the developer and maintainer of the NERC Vocabulary Server (NVS). The role of the technical governance group is analogous to the ISO definition of a "register manager".

3.6. Making the content available

The NETMAR project’s knowledge organization systems are built upon the World Wide Web Consortium’s Simple Knowledge Organization System (SKOS) standard. SKOS is designed to provide a method for the online publication of controlled vocabularies and thesauri. NETMAR publishes two International Coastal Atlas Network thesauri and an Oregon Coastal Atlas thesaurus as XML documents using the SKOS standard. A brief overview of SKOS is therefore provided below.

SKOS is based upon concepts that it defines as a "unit of thought", i.e. an idea or notion such as "shoreline emergency access" or "oil spill". Concepts may also carry other information, such as their relationships to other concepts and information about their provenance and version history. SKOS provides the means for grouping those concepts together as either collections or schemes. A SKOS collection is a grouping of concepts which share something in common and can be conveniently grouped under a common label, for example "SeaDataNet agreed parameter groups" or "ISO19115 topic categories". Similarly, SKOS concept schemes are also groupings of concepts but the relationships between the concepts are a part of the concept scheme. For example, if the eBay auction categories were published as a concept scheme, "Antiques" and "Sporting Goods" can be identified as SKOS top-concepts, the broadest definitions in the pyramids of concepts. The narrower concept definitions such as "Antique Clocks" and "Sailing" can also be delivered in the concept scheme, including their position in the hierarchy of concepts. Therefore, concept schemes are a useful model for the publication of thesauri, for example the "ICAN coastal erosion thesaurus."

SKOS also defines three forms of relationship between concepts. A concept may be broader or narrower than another concept, or related to another concept. The related attribute allows the loose mapping of one concept to another, allowing the resource to become orthogonal. The broader and narrower attributes allow the construction of a hierarchy. If a concept belongs to a hierarchical scheme and is an entry point to that hierarchy (that is, at the top of the tree) it can be declared as a SKOS topConcept. For concepts in the same scheme, the broader and narrower relations may be said to be transitive; that is a concept two levels below a given concept can be inferred to be narrower than the concept in question without explicitly stating a relationship. For example (and illustrated below), eBay has "Sporting Goods" as a top level auction category, or a topConcept. Narrower than this is "Sailing", and still narrower is "Rope". If these relationships were declared as transitive "Rope" could be inferred to be narrower than "Sporting Goods", which is not explicit in the non-transitive SKOS narrower relationship.

The differences between SKOS concept collections and concept schemes are very limited in the W3C’s specification. The NETMAR project has chosen to use schemes as a discovery tool for concepts, and collections to store and publish concepts and for referencing their identifiers.

The NETMAR semantic framework has additionally extended the SKOS model to allow synonyms to be identified using the Web Ontology Language’s sameAs attribute. This clearly allows the labelling of the relationship between two concepts which are identical, which is not a feature of the basic SKOS model.

3.7. Deploying ICAN semantics in the NETMAR semantic framework

3.7.1. Incorporating a Knowledge Organization System

The simplest way for an ICAN community member to develop a new controlled vocabulary or thesaurus

\[\text{https://www.bodc.ac.uk/data/codes_and_formats/seavox/}\]

\[\text{http://vocab.nerc.ac.uk}\]

\[\text{http://www.dgiwg.org/Terminology/faq-other.php}\]

\[\text{http://www.w3.org/TR/owl2-overview/}\]
(or propose new content for an existing vocabulary or thesaurus) for incorporation within the framework is to create two worksheets in a spreadsheet: one for concept names and definitions; the other for relationships between concepts.

The first worksheet, illustrated in Table 1, should contain columns for

- Concept key. An identifier for the concept, unique within the vocabulary. It does not need to carry any meaning.
- Concept name and title
- Concept alternative name (e.g. abbreviation)
- Concept definition

Each concept must only occupy one row of the worksheet. If the definition needs to carry some structured information (such as information regarding the identity of a ship’s hull or the bounding box of a geographic area), this should be encoded according using an alternative to XML, such as the JavaScript Object Notation (JSON) standard, i.e. enclosed in curly brackets and formed of "key": "value" pairs separated by commas. An example is given in Table 1.

The second worksheet should contain three columns describing the relationship between concepts, and is illustrated in Table 2

- The subject of the sentence describing the relationship
- Narrower, broader, related or sameAs mapping
- The object of the sentence describing the relationship

Once complete, the spreadsheet should be submitted with supporting information about the domain scope of the concepts, the content governance for the knowledge organization system and the name and contact details for those authorised to make changes to the resource. The supporting information for the ICAN Coastal Erosion thesaurus, for example, is:

- Domain scope: "Thesaurus containing coastal erosion dataset (including GIS layer) terms compiled by ICAN and mapped to a global thesaurus. Includes both markup and discovery terms from the mapped components."
- Content governance: "International Coastal Atlas Network"

The knowledge organization system will be deployed in the NETMAR semantic framework and further updates can be made by authorised persons through a web interface accessed from the British Oceanographic Data Centre website.

3.7.2. Accessing the Knowledge Oraganisation System of the NETMAR semantic resource

Once deployed within the NETMAR semantic framework, a knowledge organization system can be accessed in much the same way as a web site, using Uniform Resource Locators (URLs) to navigate the NVS. The base URL for the NVS is http://vocab.nerc.ac.uk. Full documentation of the application programming interface to the NVS can be found at this location.

3.8. Extending to existing Knowledge Organization Systems

Labelling data and metadata using a knowledge organization system is a first step to making those data interoperable with other datasets. However, if the knowledge organization system has defined relationships to other systems the likelihood of the metadata and data being discovered and reused alongside other data increases. Linked data is an initiative of the World Wide Web Consortium to create a web of data described knowledge organization systems.

A range of environmental science and geospatial knowledge organization systems exist that may be of interest for bridging a new knowledge organization system too, several of which have been described earlier in this paper. Relationships between a concept in the NVS and any external concept can be specified in the same way as the internal mappings but with the NVS URL replaced by the URL of the external concept as the object of the relationship.

4. Concluding remarks

The NETMAR project has adopted a simple approach to semantics ICAN, in which familiar tools such as Microsoft Excel were used to assemble the local vocabularies and the mappings as a series of spreadsheets. These were then imported into an Oracle database and exported by a Java software layer (NVS) as a thesaurus conforming to the W3C Simple Knowledge Organisation System (SKOS) standard. As a part of this work, the mapping between the Oregon Coastal Atlas and the ICAN Coastal Erosion Global Atlas was completed as shown in Figure 2. The resulting thesaurus is available as a SKOS document from

---

9
Table 1
Example concept preparation for the NETMAR semantic framework

<table>
<thead>
<tr>
<th>Concept Key</th>
<th>Concept name &amp; title</th>
<th>Concept alternative name</th>
<th>Concept definition</th>
</tr>
</thead>
</table>
| 74PQ        | Plymouth Quest PQ    | PQ                       | {
|             |                      |                          | "title": "RV", |
|             |                      |                          | "callsign": "MEEU8", |
|             |                      |                          | "platformClass": "research vessel", |
|             |                      |                          | "commissioned": "2004-03-24", |
|             |                      |                          | "previous_name": "Sigurbjorg"} |

Table 2
Example concept mappings for the NETMAR semantic framework

<table>
<thead>
<tr>
<th>Subject</th>
<th>Relationship</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>74PQ (&quot;Plymouth Quest&quot;)</td>
<td>Is narrower than</td>
<td><a href="http://vocab.nerc.ac.uk/collection/L06/current/31/">http://vocab.nerc.ac.uk/collection/L06/current/31/</a> (&quot;research vessel&quot;)</td>
</tr>
<tr>
<td>74PQ (&quot;Plymouth Quest&quot;)</td>
<td>Is narrower than</td>
<td><a href="http://vocab.nerc.ac.uk/collection/L19/current/SDNKG04">http://vocab.nerc.ac.uk/collection/L19/current/SDNKG04</a> (&quot;platform&quot;)</td>
</tr>
</tbody>
</table>

Fig. 2. The ICAN Coastal Erosion Thesaurus

This thesaurus has been used to test a semantically aware CSW mediator which enable smart discovery of the associated, distributed CWA datasets.

Following requests from the ICAN community at their 2011 workshop in Oostende, Belgium, a draft ICAN Water Quality Global Thesaurus has been developed as shown in Figure 3.

the NVS\textsuperscript{22} with the response returned by the vocabulary server being a SKOS document encoded in RDF-XML. However, a stylesheet has been included to facilitate human browsing of the resource.

\textsuperscript{22}http://vocab.nerc.ac.uk/scheme/ICANCOERO

---
The cookbook has also been used by members of the INSPIRE data format specification drafting teams to develop well governed and openly published code lists that will become enshrined in European legislation.

5. Acknowledgments

This work was undertaken for the International Coastal Atlas community under the auspices of the NETMAR (Open Service Network for Marine Environmental Data) project. NETMAR is partially funded by the European Commission under Theme ICT-2009.6.4 ICT for environmental services and climate change adaptation of the Information & Communication Technologies FP7 Programme.

References


