Project No. 249024

NETMAR

Open service network for marine environmental data

Instrument: *Please tick*	CA	✓ STREP	IP	NOE

ICT - Information and Communication Technologies Theme

D7.9.2 ICAN semantic interoperability pilot cookbooks

Reference: D7.9.1_ICAN_semantic_cookbooks_r2_20120731

Due date of deliverable (as in Annex 1): M0 + 30
Actual submission date: 31 July 2012

Start date of project: 1 February 2010   Duration: 3 years

Coastal and Marine Resources Centre (CMRC)

Revision 1

	NETMAR
	Open service network for marine environmental data Project Reference: 249024 Contract Type: Collaborative Project Start/End Date: 01/03/2010 - 31/01/2013 Duration: 36 months
	Coordinator: Prof. Stein Sandven Nansen Environmental and Remote Sensing Center Thormøhlensgate 47, Bergen, Norway Tel.: +47 55 20 58 00 Fax. +47 55 20 58 01 E-mail: stein.sandven@nersc.no

Consortium

The NETMAR Consortium is comprised of:

- Nansen Environmental and Remote Sensing Center (NERSC), Norway (coordinator).
  Project Coordinator: Prof. Stein Sandven (stein.sandven@nersc.no)
  Deputy Coordinator: Dr. Torill Hamre (torill.hamre@nersc.no)
  Quality Control Manager: Mr. Lasse H. Pettersson (lasse.pettersson@nersc.no)
- British Oceanographic Data Centre (BODC), National Environment Research Council, United Kingdom
  Contact: Dr. Roy Lowry (rkl@bodc.ac.uk)
- Centre de documentation de recherche et d'expérimentations sur les pollutions accidentelles des eaux (Cedre), France.
  Contact: Mr. François Parthiot (Francois.Parthiot@cedre.fr)
- Coastal and Marine Resources Centre (CMRC), University College Cork, National University of Ireland, Cork, Ireland.
  Contact: Mr. Declan Dunne (d.dunne@ucc.ie)
- Plymouth Marine Laboratory (PML), United Kingdom.
  Contact: Mr. Steve Groom (sbg@pml.ac.uk)
- Institut français de recherche pour l'exploitation de la mer (Ifremer), France.
  Contact: Mr. Mickael Treguer (mickael.treguer@ifremer.fr)
- Norwegian Meteorological Institute (METNO), Norway.
  Contact: Mr. Øystein Torget (oysteint@met.no)

Author(s)

- Declan Dunne (d.dunne@ucc.ie)
- Adam Leadbetter (alead@bodc.ac.uk)
- Yassine Lassoued (y.lassoued@ucc.ie)

Document approval

- Document status: Revision 1
- WP leader approval: 2012-07-30
- Quality Manager approval: 2012-07-31
- Coordinator approval: 2012-07-31

Revision History

Issue	Date	Change records	Author(s)
Draft	2011-12-05	Added first draft of the "Understanding semantics" cookbook.	A. Leadbetter
Draft	2011-12-16	Added first draft of the "Connecting your Atlas" cookbook.	Y. Lassoued
Draft	2011-12-19	Revised the "Understanding semantics" cookbook according to comments from quality control.	A. Leadbetter
Draft	2011-12-20	Added first draft of the "Metadata" and "CSW metadata catalogue" cookbooks.	D. Dunne
Draft	2011-12-22	Revised the "Connecting your Atlas" cookbook according to comments from quality control.	D. Dunne
Draft	2011-12-23	Revised the "Metadata" and "and "CSW metadata catalogue" cookbooks according to comments from quality control.	D. Dunne
1	2011-12-29	Final release approved by coordinator.	T. Hamre
Draft	2012-07-18	Updated version of the "Understanding semantics" cookbook.	A. Leadbetter
Draft	2012-07-18	Updated version of the "Metadata" cookbook.	D. Dunne
Draft	2012-07-19	Updated version of the "CSW metadata catalogue" and "Connecting your Atlas" cookbook.	D. Dunne
Draft	2012-07-26	Revised the "Understanding semantics" cookbook according to comments from quality control.	A. Leadbetter
Draft	2012-07-26	Revised the "Connecting your Atlas" cookbook according to comments from quality control.	D. Dunne
Draft	2012-07-27	Revised the "CSW metadata catalogue" cookbook according to comments from quality control.	D. Dunne
Draft	2012-07-30	Revised the "Metadata" cookbook according to comments from quality control.	D. Dunne
Draft	2012-07-31	Revised the Executive Summary according to comments from quality control.	D. Dunne
2	2012-07-31	Final release approved by coordinator.	T. Hamre

Executive Summary

The NETMAR project will develop a pilot European Marine Information System (EUMIS) that will enable users to search, download and integrate satellite, in situ and model data from ocean and coastal areas. EUMIS will be a user-configurable system offering flexible service discovery, access and chaining facilities based on open and widely adopted web GIS standards. EUMIS will be a distributed system where a number of services and subsystems will be integrated and made interoperable by means of semantic technologies.

NETMAR's ICAN (International Coastal Atlas Network) pilot is scoping and implementing data interoperability approaches for distributed coastal web atlases (CWAs). The ICAN community wish to make their ongoing digital atlas developments more interoperable to better support data discovery, data visualisation, and data sharing across administrative, natural and thematic borders. Data discovery is central to this pilot as it enables subsequent data visualisation and data sharing. Data discovery includes smart search functionality utilising "semantic" resources.

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") using a common set of meaningful terms or phrases. These common phrases must be consistent throughout the full data system, 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", fulfilling the vision of Tim Berners-Lee to make a world wide web of data which may be queried in the same way as a traditional relational database system.

Atlas developers require guidelines to support improved semantic interoperability and connectivity between distributed atlases, both legacy and new. This document contains short documents or "cookbooks" which explain key information behind the NETMAR ICAN pilot. Central to these cookbooks is material to help atlas developers connect to the International Coastal Web Atlas (ICWA) prototype. It is a prototype atlas mediator which provides a common interface for accessing distributed local atlases, such as MIDA (Marine Irish Digital Atlas), OCA (Oregon Coastal Atlas), and Washington Coastal Atlas (WCA). It uses a knowledge organization system to improve data discovery by exploiting the semantics of keywords and allowing users to search data by "meaning" rather than by "mere keywords". The ICWA prototype version 3 is being funded by the NETMAR project.

In order for an atlas to connect to the ICWA prototype, the atlas developer must, at a minimum, install a CSW metadata catalogue. This CSW metadata catalogue is then registered with the ICWA Administrator to connect to the prototype. The ICWA requires CSW version 2.0.2 supporting the ISO Metadata Application Profile version 1.0.0, that is able to deliver metadata in Dublin Core and ISO 19115/19139. ISO 19119/19139 service metadata is also supported by the ICWA. There is no particular requirement as to the metadata profile used, but metadata records need to include keywords. However, to fully exploit semantics for data discovery, the atlas developer must use semantic resources for metadata keywords.

Four cookbooks are included in this document which covers key material concerning the technologies and standards utilised by the ICWA prototype:

1. Understanding Semantics.

2. Understanding Metadata.

3. Establishing a CSW metadata catalogue with GeoNetwork opensource.

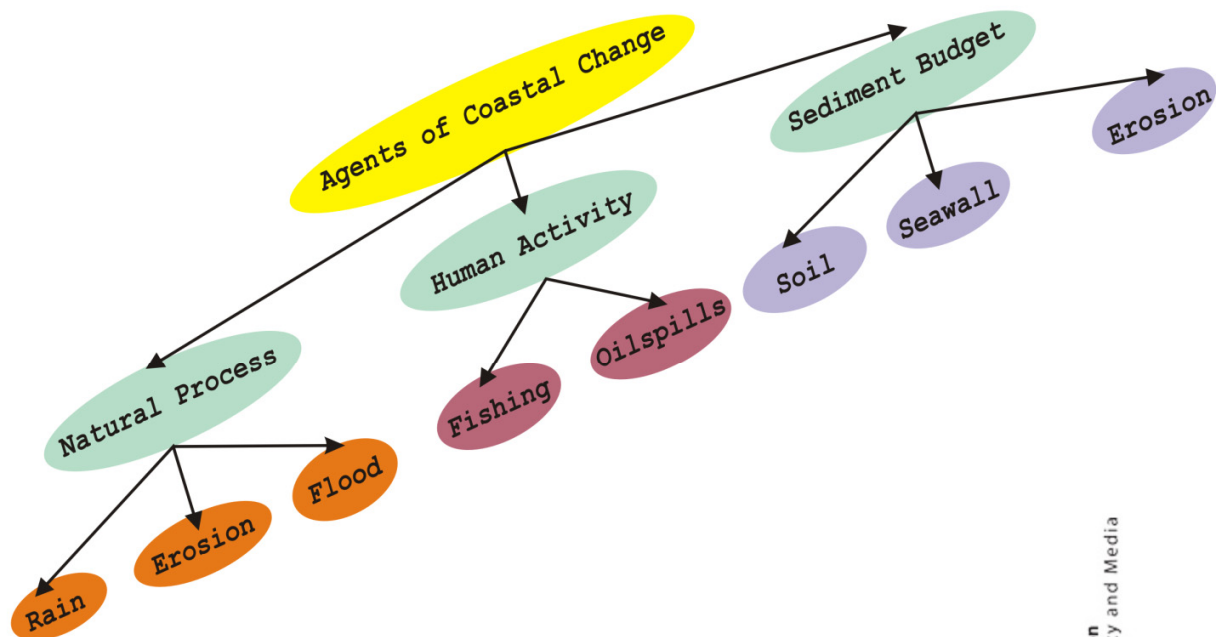4. Connecting your Atlas to the ICWA prototype.

The "Understanding Semantics" cookbook provides a tutorial for those who wish to investigate and make use of semantic web and knowledge organization system technologies. These technologies fall broadly into three groups: vocabularies, thesauri and ontologies. This cookbook includes material explaining how to deploy semantics which are required by the ICWA prototype, utilising the NERC Vocabulary Server (NVS) infrastructure.

The "Understanding Metadata" cookbook provides a tutorial for those who wish to understand metadata. Included is a description of metadata and why we need it, metadata standards in use today, description of different metadata hierarchy levels, and a list of some metadata editing tools available. The section also references example metadata records aimed at system developers who are familiar with ISO 19115/19119 metadata implemented in ISO 19139 XML. These metadata examples include semantic keywords which are registered in the NVS.

The "Establishing a CSW metadata catalogue with GeoNetwork opensource" cookbook provides a tutorial for those who wish to understand CSW (Catalog Services for the Web) metadata catalogues. Included is a description of a metadata catalogue, the CSW standard, and a list of selected CSW servers. The document also contains initial pointers to establishing a CSW server using GeoNetwork opensource and examples of selected CSW query operations aimed at system developers. GeoNetwork is recommended and used by the NETMAR project.

The "Connecting your Atlas to the ICWA prototype" cookbook provides a step-by-step guide explaining how to connect a local atlas as a node in the ICWA prototype utilising technologies and standards described in the other cookbooks. This cookbook specifies the ICWA connection requirements including metadata (ISO 19115/19119/19139), CSW metadata catalogues (CSW 2.0.2, ISO Metadata Application Profile version 1.0.0), and supported ISO 19139 encoding forms for semantic keywords.

International Coastal Atlas Network Cookbook: Understanding Semantics

Table of Contents

Introduction

"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"

Tim Berners-Lee[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") using a common set of meaningful terms or phrases. These common phrases must be consistent throughout the full data system, or there must be a means of translating between the phrases used at different points of the system, using common "semantics". Semantics is the study of meaning; it focuses on the relationships between words and what they stand for or mean. 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 document 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 environmental scientists and data managers.
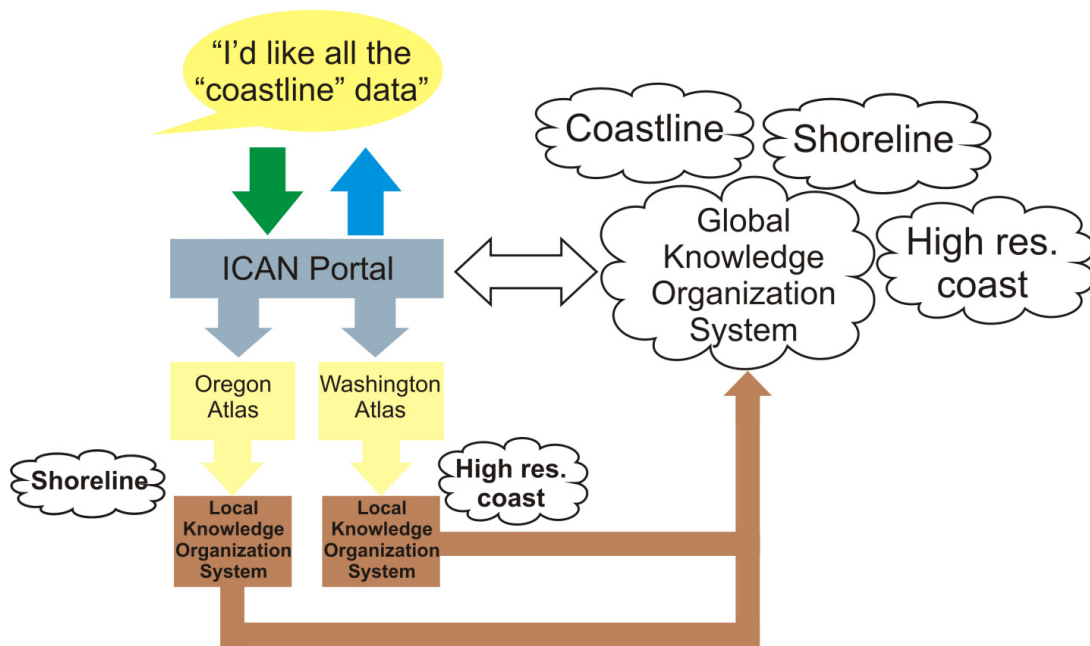
Why use a "knowledge organization system"?

One scenario for using knowledge organization systems in the International Coastal Atlas Network[2] (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"[3].

---

[1] Berners-Lee, T. (1999) *Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor*. Orion Business. ISBN-100752820907

[2] http://ican.science.oregonstate.edu/

[3] Latham, S. E.; Cramer, R.; Grant, M.; Kershaw, P.; Lawrence, B. N.; Lowry, R.; Lowe, D.; O'Neill, K.; Miller, P.; Pascoe, S.; Pritchard, M.; Snaith, H.; Woolf, A. (2009) The NERC DataGrid services. *Philosophical Transactions of the Royal Society A*, 367 (1890). 1015-1019.

A diagram illustrating one use for knowledge organization systems in the ICAN community.

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.

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 as illustrated in the diagram below.
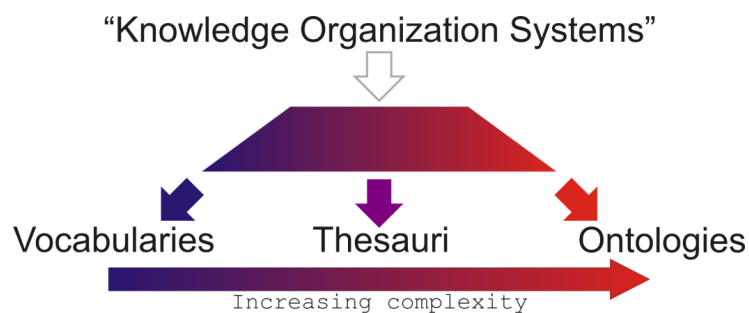
The "semantic spectrum" shows the increasing complexity of different forms of knowledge organization system. After McGuinness (2003)[4].

---

[4] Deborah L. McGuinness. (2003) Ontologies Come of Age. In Dieter Fensel, James Hendler, Henry Lieberman, and Wolfgang Wahlster (eds). *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. Massachusetts Institute of Technology Press.

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. A well known example of this form is the Yahoo! web directory[5] or the categorisation of auctions on the eBay homepage[6]. 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[7] can be seen to work in this way. For example, "Oceans" is at the top level, with "Coastal Processes" beneath it and terms such as "Beaches" and "Coastal Elevation" beneath that.

These more complex thesauri 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" 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.

A thesaurus may be expanded to an ontology by declaring a term to belong to a particular class; or the addition of property information to the term; or the restriction of values that data associated with the term may take. An ontology class is used to define a type which can be used to group related terms. 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".

---

[5] http://dir.yahoo.com/
[6] http://www.ebay.com/
[7] http://gcmd.nasa.gov/

How to discover existing knowledge organization systems?

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 (see page 13). Details of how to access an existing knowledge organization system relevant to the International Costal Atlas Network are provided on page 12 of this document.

It is also possible to extended existing resources by creating mappings between them and other resources. This activity is described on page 13, below.

Where might I find existing knowledge organization systems?

In order to reuse existing resources, it is essential to know where to find them and how to asses their quality. Existing resources which may be of interest can often be found in ontology registries, for instance the Marine Metadata Interoperability Ontology Registry and Repository[8] or the NERC Vocabulary Server[9]. The former has a search facility on its home page; the latter may be searched most easily through the SeaDataNet vocabulary pages hosted by Maris[10]. Both of these systems provide publication mechanisms for knowledge organization systems which may be created by a range of authorities, and the creating authority is acknowledged in the systems' output. An additional benefit of these systems is that they provide versioning of the content of the knowledge organization systems that they serve.

Other resources that are of interest to the Earth Sciences domain exist outside of these registry systems. These include the NASA's Global Change Master Directory[7] and Semantic Web for Earth and Environmental Terminology[11]; the European Environment Agency's General Multilingual Environmental Thesaurus (GEMET)[12]; the GeoSciML vocabularies[13]; and the United States Geological Survey thesaurus[14].

When considering the use of an existing knowledge organization system, the key things to look for are: an individual web address (or URL) to each term defined – this is how you will mark up your metadata; a well documented version control system; and an authoritative body in control of the content of the KOS.

---

[8] http://mmisw.org/orr/
[9] http://vocab.nerc.ac.uk/
[10] http://seadatanet.maris2.nl/v_bodc_vocab/welcome.aspx/
[11] http://sweet.jpl.nasa.gov/ontology/
[12] http://www.eionet.europa.eu/gemet/
[13] http://srvgeosciml.brgm.fr/eXist2010/brgm/client.html
[14] http://www.usgs.gov/science/about/

How to define the content of a knowledge organization system?

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.

Identifying the content

How narrow or broad should a term definition be?

The challenge of integrating data and information of different kinds at different levels of detail is well defined in computer science literature[15,16]. 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.

Linking term definitions together

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 and MEDATLAS Parameter Usage vocabularies are of similar granularity and are linked this way); and synonyms where two terms may be used interchangeably.

---

[15] Fonseca, F., Egenhofer, M., Davis, C., and Câmara, G. (2002) Semantic Granularity in Ontology-Driven Geographic Information Systems. *AMAI Annals of Mathematics and Artificial Intelligence - Special Issue on Spatial and Temporal Granularity* 36(1-2): 121-151.

[16] Yan, X., Lau, R.Y.K, Song, D., Li, X., Ma, J. (2011) Towards a Semantic Granularity Model for Domain Specific Information Retrieval. *ACM Transactions on Information Systems (TOIS)*. In press.
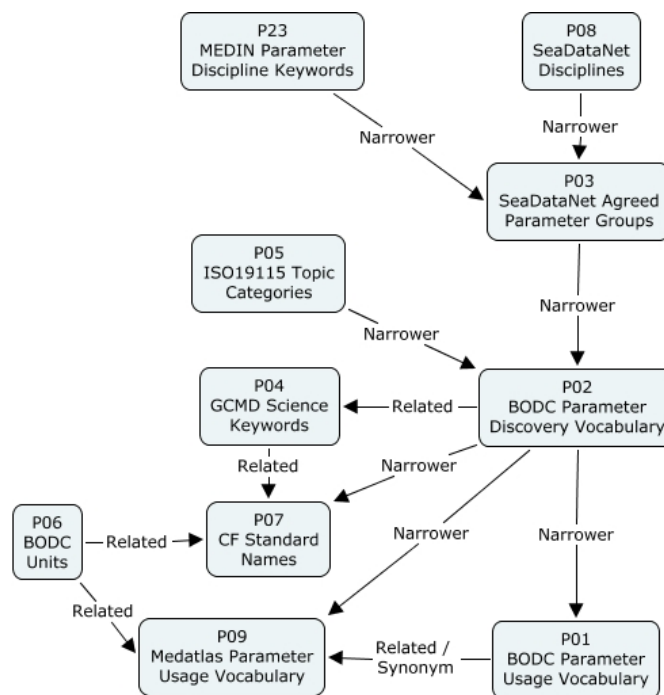
An example from the NERC Vocabulary Server[9] to show how identifying relationships between terms builds a network of parameter definitions.

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 Governance Group (SeaVoX)[18]; and the NETMAR ontology governance body[19]. The role of the content governance group is analogous to the International Organization for Standardization (ISO) definition of a "control body"[20].

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**Error! Bookmark not defined.** (NVS). The role of the technical governance group is analogous to the ISO definition of a "register manager"[20].

---

[17] http://cf-pcmdi.llnl.gov/
[18] https://www.bodc.ac.uk/data/codes_and_formats/seavox/
[19] http://netmar.nersc.no/
[20] http://www.dgiwg.org/Terminology/faq-other.php

Making the content available

Simple Knowledge Organization System

The NETMAR project's knowledge organization systems are built upon the World Wide Web Consortium's Simple Knowledge Organization System[21] (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 `topConcepts`, 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, as illustrated below. Therefore, concept schemes are a useful model for the publication of thesauri, for example the "ICAN coastal erosion thesaurus."
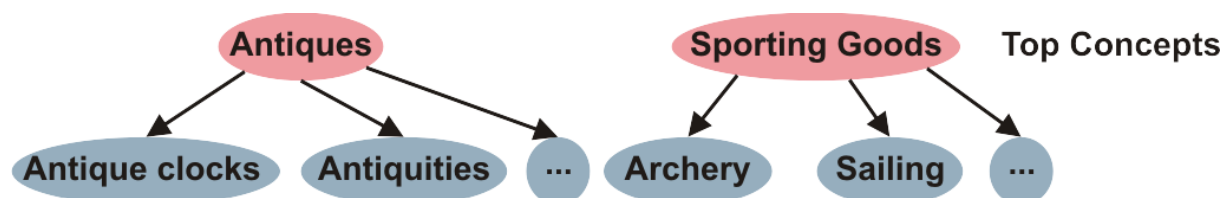

An illustrative example of top concepts in SKOS, and the first level of their associated narrower terms.

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 (see page 5). 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

---

[21] http://www.w3.org/2004/02/skos/

to be narrower than "Sporting Goods", which is not explicit in the non-transitive SKOS narrower relationship.
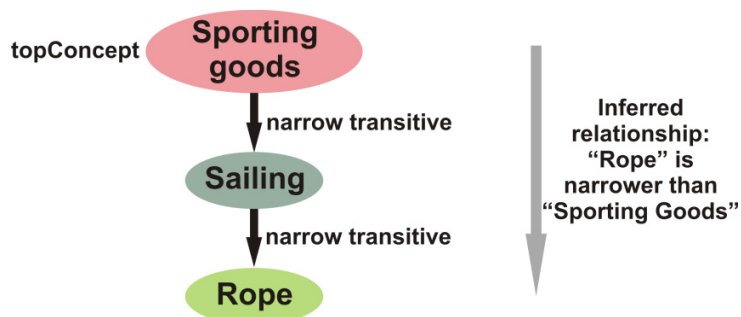
The differences between SKOS concept collections and concept schemes are very limited in the W3C's specification. The NETMAR project has chosen to use schemes as a discovery tool for concepts, and collections to store and publish concepts and for referencing their identifiers.

The NETMAR semantic framework has additionally extended the SKOS model to allow synonyms to be identified using the Web Ontology Language's[22] `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.

Deploying ICAN semantics in the NETMAR semantic framework

Incorporating a Knowledge Organization System

The simplest way for an ICAN community member to develop a new controlled vocabulary or thesaurus (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 below, should contain columns for

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

---

[22] http://www.w3.org/TR/owl2-overview/

Concept Key	Concept name and title	Concept alternative name	Concept definition
74PQ	Plymouth Quest	PQ	```{"title": "RV","callsign": "MEEU8", "platformClass": "research vessel", "commissioned": "2004-03-24","previous_name": "Sigurbjorg"}```

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 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. For example:

```
{"title": "RV", "callsign": "MEEU8", "platformClass": "research vessel",
"commissioned": "2004-03-24","previous_name": "Sigurbjorg"}
```

The second worksheet should contain three columns describing the relationship between concepts:

1. Subject
   - The subject of the sentence describing the relationship.
2. Relationship
   - Narrower, broader, related or sameAs mapping.
3. Object
   - The object of the sentence describing the relationship.

Subject	Relationship	Object
74PQ ("Plymouth Quest")	Is narrower than	http://vocab.nerc.ac.uk/collection/L06/current/31/ ("research vessel")
74PQ ("Plymouth Quest")	Is narrower than	http://vocab.nerc.ac.uk/collection/L19/current/SDNKG04 ("platform")

Once complete, the spreadsheet should be submitted to enquiries@bodc.ac.uk along 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[23].

Accessing the Knowledge Organization System

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[24] (URLs) to navigate the NVS. The base URL for the NVS is:

http://vocab.nerc.ac.uk

Catalogues of the SKOS concept collections and schemes hosted on the NVS can be accessed at:

http://vocab.nerc.ac.uk/collection/

http://vocab.nerc.ac.uk/scheme/

Once the identifier for an individual collections or schemes is known, it can then be accessed from:

http://vocab.nerc.ac.uk/collection/*collection_id*/current/

> e.g. http://vocab.nerc.ac.uk/collection/C17/current/ is the URL for the International Council for the Exploration of the Seas platform codes collection from which the example worksheets above were taken

http://vocab.nerc.ac.uk/scheme/*scheme_id*/current/

> e.g. http://vocab.nerc.ac.uk/scheme/ICANCOERO/current/ is the URL for the ICAN Coastal Erosion thesaurus

Finally, an individual concept can be accessed through this form of URL:

http://vocab.nerc.ac.uk/collection/*collection_id*/current/*concept_id*/

> e.g. http://vocab.nerc.ac.uk/collection/C17/current/74PQ/ gives access to the concept definition for "Plymouth Quest" which was described in the example worksheets above

---

[23] https://www.bodc.ac.uk/data/codes_and_formats/vocabulary_editor/
[24] http://en.wikipedia.org/wiki/Url

The collection URLs also provide a mechanism for accessing any concepts which have been removed from the collection (known as deprecation), or only those concepts which are currently accepted members of the collection or all the concepts which have ever been part of the collection (the default if neither deprecated, accepted or all is specified as a suffix to the collection URL):

http://vocab.nerc.ac.uk/collection/*collection_id*/current/deprecated/

http://vocab.nerc.ac.uk/collection/*collection_id*/current/accepted/

http://vocab.nerc.ac.uk/collection/*collection_id*/current/all/

The ../current/../ portion of the URLs given in this section is a shortcut to the most recent version of the collection or scheme. This can be replaced with an integer value in order to retrieve a given version of a collection or scheme.

In addition to this URL based access, application developers can make use of Simple Object Access Protocol (SOAP)[25] based access described in the associated Web Services Description Language (WSDL) document[26].

Bridging 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. The diagram on the next page shows how this web of data is highly interconnected.

A range of environmental science and geospatial knowledge organization systems exist that may be of interest for bridging a new knowledge organization system too. These include those stored in the NVS and the Marine Metadata Interoperability Ontology Registry and Repository[8]; the European Environment Agency General Multilingual Environmental Thesaurus[12]; and GeoNames[27]. Relationships between a concept in the NVS and any external concept can be specified in the same way as the internal mappings (see page 7) but with the NVS URL replaced by the URL of the external concept as the object of the relationship. For example:
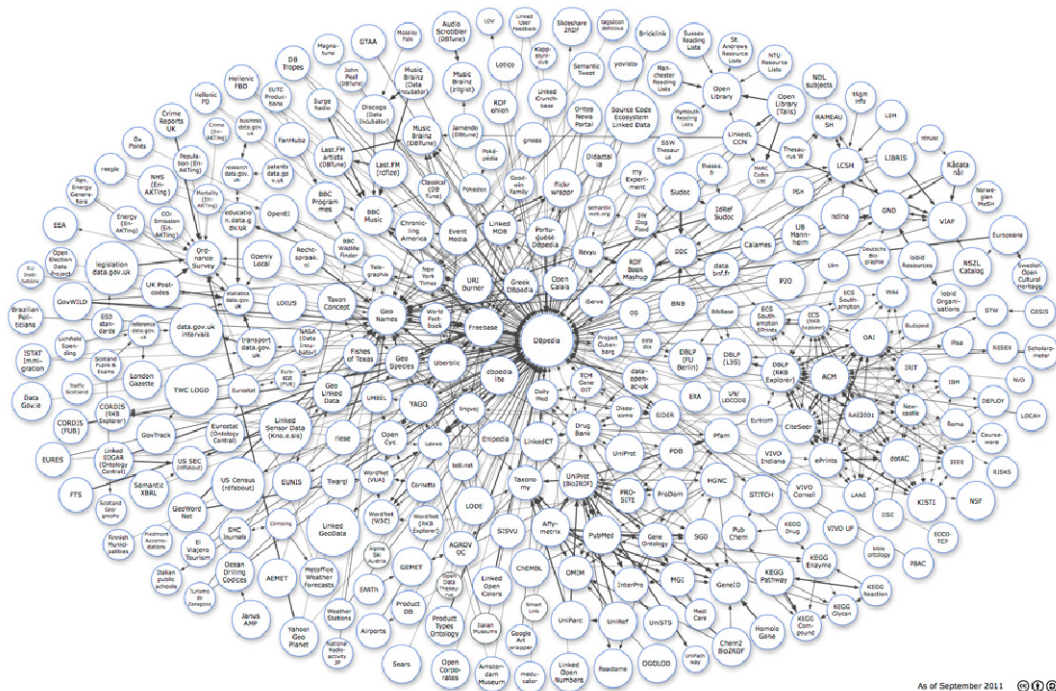
```
http://vocab.nerc.ac.uk/collection/P21/current/MS10360/ (sulphides)
"broader"
http://www.eionet.europa.eu/gemet/concept/4350 (inorganic substances)

http://vocab.nerc.ac.uk/collection/C19/current/3_1_2_1/ (Adriatic Sea)
"sameAs"
http://sws.geonames.org/3183462/
```

---

[25] http://en.wikipedia.org/wiki/SOAP
[26] http://vocab.nerc.ac.uk/v2.wsdl
[27] http://www.geonames.org/

The Linking Open Data project cloud[28].

Incorporating knowledge organization systems in ICAN metadata

This is described in detail in the accompanying cookbook: "Connecting Your Atlas." However, in overview, the web address (URL) of a term defined in a knowledge organization system should be incorporated within a metadata document, in the appropriate field. This may be as either a string, in an XML element such as `gco:CharacterString`, or as a reference from an anchor field, using the `xlink:href="http://..."` syntax.

Acknowledgements

---

[28] http://richard.cyganiak.de/2007/10/lod/imagemap.html

and readable. Further feedback on this document is welcomed, and may be provided by contacting the author whose details are below.

Document Information

Author	Adam Leadbetter, British Oceanographic Data Centre	
Contact	alead@bodc.ac.uk	
Version	2.1	
Date	2012 July 26	
Revisions	2.1	Responses to NETMAR internal review. Fixing spelling mistakes; adding new definitions; clarifying some existing definitions; additional concept scheme diagram.
	2.0	2012 July 17: Comments from ICAN community with respect to discovery of existing resources included
	1.0	2011 December 19

International Coastal Atlas Network Cookbook:

Understanding Metadata

Table of Contents

Introduction

This document provides a brief tutorial for those who wish to get an overview of metadata, with a focus on the ISO 19115 metadata standard. It is aimed specifically at members of the International Coastal Atlas Network (ICAN) community and more generally at scientists, data managers, and system developers. Aimed at scientists and data managers, this document includes a description of metadata and why we need it, metadata standards in use today, description of different metadata hierarchy levels, and a list of selected metadata editing tools available. Aimed at system developers, this document also includes information on metadata encoding using ISO 19139 XML and a reference to example metadata records located on the NETMAR WIKI. ISO 19115/19139 metadata is a requirement to connect to the International Coastal Web Atlas (ICWA) prototype.

What is metadata?

Geospatial metadata is "data about data". It contains information that documents the basic characteristics of a geospatial data resource. It can also document basic characteristics of geospatial applications or services. Metadata falls into broad categories where it answers the "what, why, when, who, where and how" questions about the resource. These questions include[1]:

- What: Title and description of the data.

- Why: A narrative summary detailing the reasons for data collection and its uses.

- When: When the data was created and the update cycles, if any.

- Who: Originator, data supplier, and possibly the intended audience.

- Where: The geographical extent based on latitude and longitude coordinates, geographical names or administrative areas.

- How: How the data was produced and how to access the data.

Why do you need metadata?

Metadata helps a user to find or discover the data that they need and, thereafter, evaluate whether this resource satisfies the user's requirements. Once a user has chosen the resource, usage metadata is then required to help fully understand and interpret the data. Metadata can be used in-house to help locate and use internal data resources. If a staff member leaves an organisation, important knowledge may also leave the organisation too. New staff members may have difficulty in taking up new responsibilities and fully understanding the organisation's data resources. Such undocumented data resources may lose

---

[1] Wilson, M., 2009, Chapter Three: Metadata -- Describing geospatial data, Spatial Data Infrastructure Cookbook.

value or cost time to relearn its value. Metadata can also be used to locate data resources published by other organisations, helping to minimise duplication of data collection and enabling more efficient and cost-effective use of this data. Specific examples of business cases for using metadata outlined by the Federal Geographic Data Committee (FGDC) include[2]:

Data Management:

- Preserve data history so that the data resource can be reused or adapted.

- Assess the age and character of data holdings to determine which data should be maintained, updated or deleted.

- Improve data accountability.

- Limit data liability by explicitly stating data limitations of use.

Project Management:

- Plan and document the data resources required for a project.

- Monitor data resources' development progress.

- Share data resources' development progress with project participants.

- Ability to access data characteristics for outsourced data production by ensuring metadata is a contract deliverable.

Metadata standards and profiles

In order for geospatial metadata to operate effectively between different organisations and data users, metadata must be compliant with international standards. Such standards provide a common structure and format to describe metadata. Standards enable improved metadata interoperability and integration, thus, facilitating more seamless sharing, searching, and discovery of metadata between organisations and users of geospatial data and services. Discovery metadata is the minimum amount of information that needs to be provided to help users find geospatial resources. Prominent metadata standards in use today include:

- ISO 19115 (Geographic information – Metadata)

- ISO 19119 (Geographic information – Services)

- ISO 19139 (Geographic information – Metadata – XML schema implementation)

- Dublin Core (ISO 15836)

---

[2] Business Case for Metadata (http://www.fgdc.gov/metadata/metadata-business-case)

- FGDC Content Standard for Digital Geospatial Metadata (CSDGM)

ISO 19115 / 19119 / 19139

The ISO 19115 standard provides a model or structure for describing geospatial data resources (e.g. digital datasets), while the ISO 19119 standard extends ISO 19115 to describe geospatial service resources (e.g. dataset view services). In turn, the ISO 19139 standard defines an XML schema for the physical implementation of these standards. As these are international standards, they are being widely adopted by governments and organisations around the world. While metadata entities and elements are well defined within the ISO standards, there is an extensive list of optional metadata elements on top of the mandatory metadata element set. It is left up to the system developer to define a specific information model or profile. There is no single metadata profile that fits all users' needs. Therefore, there is a need for metadata profiles to be adopted to support various regional, national, organisational, and communities' requirements.

The ISO 19115 standard is currently under formal review. The proposed ISO 19115-1 (Geographic information - Metadata - Part 1: Fundamentals) standard is expected to be approved in 2013, with the proposed XML schema implementation expected in 2015. The existing ISO standards are recommended until these revisions and associated implementations are formally approved and mature.

Dublin Core

The Dublin Core (ISO 15836) standard defines a cross-domain model or structure for describing web resources, typically general electronic documents. The Dublin Core Metadata Element Set contains fifteen properties. While Dublin Core can be successfully applied to describe geospatial resources, the ISO 19115 standard is more specialised in describing such geospatial resources. Since Dublin Core and 19115 are independent standards, therefore, a crosswalk is required to map from one standard to the other[3] [4]. Some metadata tools (e.g. GeoNetwork opensource), automatically provide a mapping from ISO 19115 to Dublin Core.

INSPIRE

Within Europe, the INSPIRE Directive[5] has defined a base metadata profile which is specified in European legislation. INSPIRE adopts the underlying ISO 19115 standard. However, full conformance to the ISO 19115 mandatory element set implies the provision of additional metadata elements which are not mandated by INSPIRE legislation. However, the INSPIRE metadata encoding guidelines respects these ISO 19115 mandatory elements by ensuring they are included as implementation recommendations. In addition, INSPIRE is defining thematic dataset specifications. Individually these dataset specifications have adapted some additional ISO 19115 elements on top of base INSPIRE metadata profile to help data evaluation.

---

[3] ftp://cenftp1.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa14857-00-2003-Nov.pdf
[4] OpenGIS Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile, Version 1.0.0, OGC 07-045
[5] http://inspire.jrc.ec.europa.eu

FGDC / NAP

Within the U.S., the Content Standard for Digital Geospatial Metadata (CSDGM) standard is the U.S. federal standard, while the North American Profile (NAP) of ISO 19115 is the U.S. national standard. The Federal Geographic Data Committee (FGDC) developed CSDGM in the 1990s for federal agencies. CSDGM is often referred as the 'FGDC metadata standard'. This standard preceded the ISO 19115 standard. The American National Standards Institute (ANSI), the U.S. member body of the ISO, adopted ISO 19115 in December of 2003. The U.S. and Canada have aligned national profile development efforts with the cooperative development of the NAP. An initial profile was adopted in 2009. In September 2010, the FGDC formally endorsed the NAP. However, once the NAP profile is fully developed, the FGDC will process it as a federal standard. Therefore, transition from CSDGM to NAP is an on-going process[6].

Metadata hierarchy levels

Metadata may exist at different levels of granularity. The most common implemented levels are "dataset" and "series". A dataset is defined by ISO 19115 as an: "identifiable collection of data" (e.g. a raster map). A dataset series is defined by ISO 19115 as a: "collection of datasets sharing the same product specification" (e.g. a collection of raster maps captured from a common series of paper maps). Metadata for which no hierarchy is listed are interpreted to be "dataset" metadata by default. Finer levels of granularity include feature and attribute metadata. Examples include: feature type (e.g. a tunnel), feature instance (e.g. the Mont Blanc Tunnel), attribute type (e.g. overhead clearance associated with a tunnel), and attribute instance (e.g. overhead clearance associated with the Mont Blanc Tunnel). These data hierarchical relationships are illustrated in Figure 1.

In addition, metadata can also be used to describe geospatial services, which typically includes information on how to access and invoke such services. Service metadata is defined by ISO 19119 as: "a service metadata record describes a service instance, including a description of the services operations and an 'address' to access the specific service instance". An example of a service metadata is a description of a Web Map Service (WMS) that enables viewing of a raster map collection.
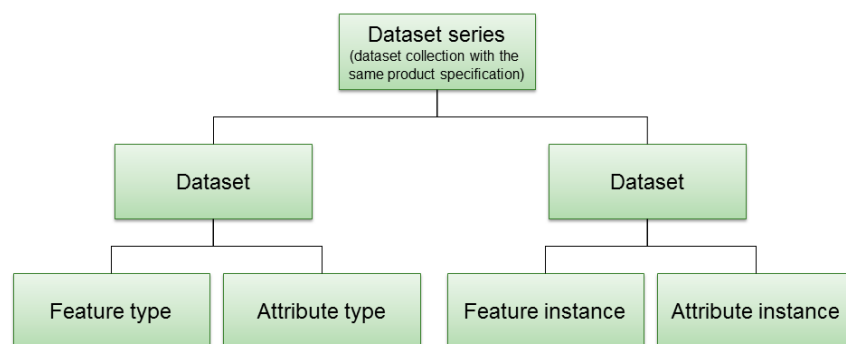

Figure 1: Metadata hierarchy[7]

---

[6] Preparing for International Metadata, Federal Geographic Data Committee, October 20, 2011
[7] ISO 19115:2003, Geographic information - Metadata

Metadata editing tools

A metadata editor is a program that is used for creating and editing metadata. It typically uses an intuitive graphical user interface which protects the user from the details of the underlying ISO 19139 XML document. A metadata editing tool may include functionalities such as:

- Creating, editing, deleting and viewing of metadata and metadata templates

- Metadata validation (support for XML schema and Schematron validation)

- Import and export of metadata

- Metadata search

- Automatic metadata generation

- Pre-processing and post-processing of metadata

- Extraction and transformation of metadata to different standards and formats

- Additional functionalities such as automatic selection of bounding box coordinates, thesaurus functions, etc.

Several metadata editing tools have been developed. Commonly used commercial metadata editing tools include:

1. ESRI ArcGIS Desktop (http://www.esri.com/)

2. Intergraph GeoMedia (http://www.intergraph.com/)

3. MapInfo Manager (http://www.pbinsight.com/)

Commonly used Opensource/freeware metadata editing tools include:

1. CatMDEdit (http://catmdedit.sourceforge.net/)

2. GeoNetwork opensource (http://geonetwork-opensource.org/)

3. M³Cat (http://www.intelec.ca/)

A screenshot of the GeoNetwork opensource metadata editor is shown in Figure 2. Example metadata fields illustrated include:

- title of the dataset (name in which the dataset is known)

- creation, publication and revision dates of the dataset

- identification code for the dataset ("SeaLevelRise" in the example)

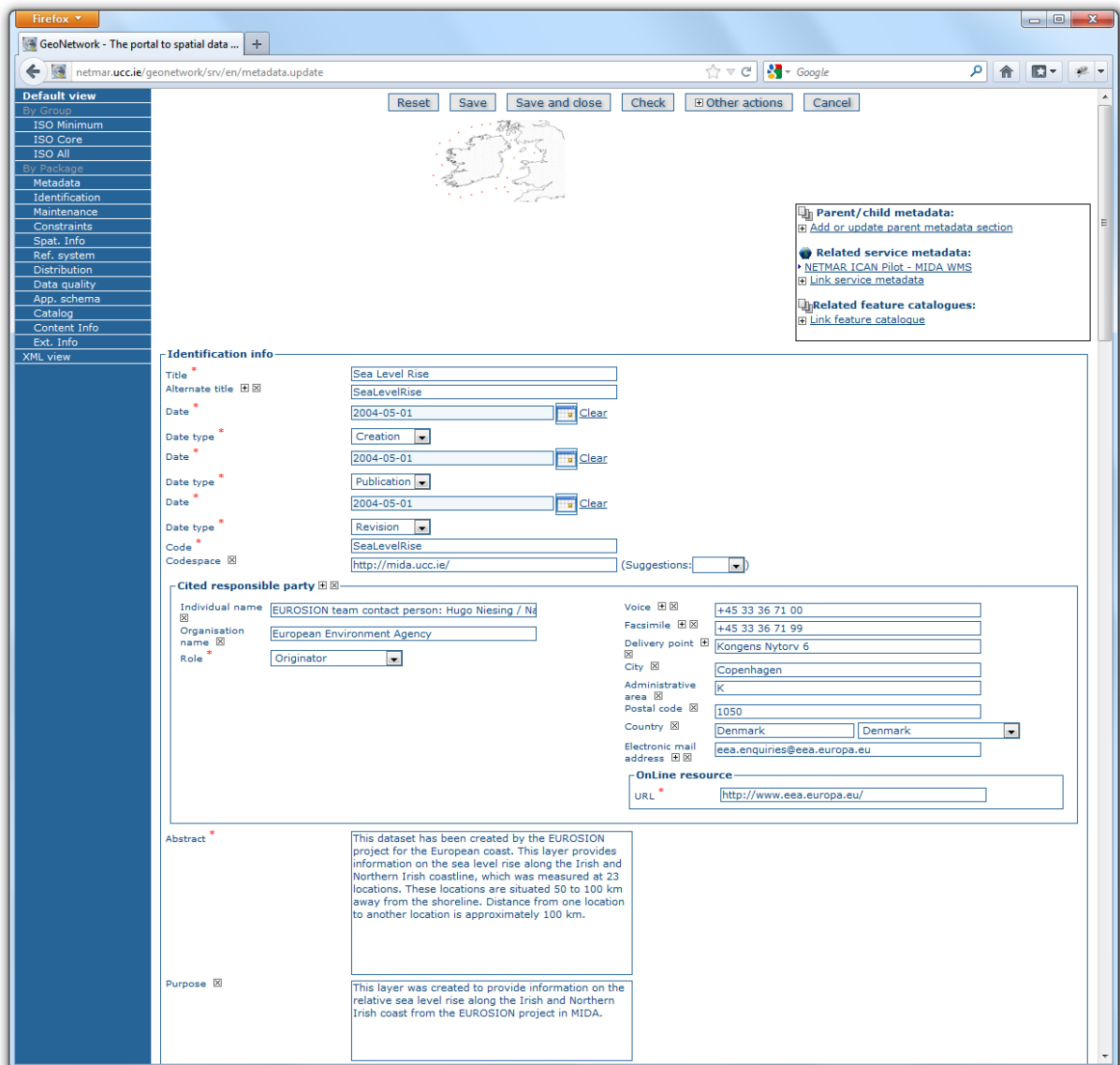- original owner of the dataset (orgainisation's name and contact details)

Figure 2: GeoNetwork opensource metadata editor

- dataset's abstract (narrative summary)
- purpose of dataset (summary of intentions)

An alternative to using geographic metadata editing tools is to directly edit the ISO 19139 XML document metadata using an XML editor (e.g. XMLSpy). However, this is only recommended for advanced users who have knowledge of ISO 19139 XML.

Metadata encoding

Metadata can be stored internally within a digital dataset or in a separate external file. Metadata can also be stored in a database to facilitate more efficient searching. The underlying metadata model (an

ISO 19115 profile) is the first important aspect to metadata interoperability. At the physical implementation level, ISO 19139 defines an XML implementation of the ISO 19115 metadata model using XSD (XML Schema Definition) schemas. The 19139 schemas define the structure of the XML metadata document. Therefore, an ISO 19139 XML implementation is the second important aspect to achieve interoperable metadata sharing and exchange between organisations and users of data. The details of the ISO 19139 encoding are not required to be understood by the typical metadata user. Most users typically use graphical based metadata search, view and editing tools. However, for system developers building spatial data infrastructures and who wish to understand more about the ISO 19139 encoding, some useful guidance resources include:

1. UK Gemini Encoding Guidance, version 1.2
   http://location.defra.gov.uk/wp-content/uploads/2010/04/UKGEMINI-Encoding-Guidance_20110505v1-2.pdf

2. Guidance notes for the production of discovery metadata for the Marine Environmental Data and Information Network (MEDIN), version 2.3.7
   http://www.oceannet.org/marine_data_standards/medin_approved_standards/documents/medin_schema_documentation_2_3_7_14mar12.pdf

ISO 19139 XSD schema sets for the full ISO 19115/19119 element set are available in the official ISO repository[8] and the OGC repository[9]. The unofficial EDEN repository[10] is a useful alternate as it has implemented patches to fix schema issues identified by users. It is recommended that metadata is validated against the ISO 19139 XML schema to ensure compliance with the standard. For metadata profiles which have specific requirements or constraints, then Schematron validation is also recommended in addition to XML schema validation. Schematron is a rule based validation language. Its strength is its ability to enforce additional constraints which XML schema cannot (e.g. specific attribute validation).

Metadata examples

For system developers who are familiar with XML, example metadata records can be downloaded from the NETMAR WIKI:

- http://eumis.nersc.no/en/wiki/-/wiki/Main/Metadata

Examples include 19115/19139 dataset metadata for vector and raster datasets, and ISO 19119/19139 service metadata examples for WMS and WPS. The details of this ISO 19139 XML are best described in the MEDIN and UK Gemini documents (referenced in the previous section).

Metadata keywords can be selected from a controlled vocabulary. A metadata user should decide if a controlled vocabulary should be used for their metadata keywords, and if so, whether an existing

---

[8] http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
[9] http://schemas.opengis.net/iso/19139/
[10] http://eden.ign.fr/xsd/isotc211/isofull/

vocabulary can be reused or a new vocabulary created. Please refer to the "*Understanding Semantics*" cookbook for future details regarding vocabularies, thesauri and ontologies.

Figure 3 illustrates an example segment of a MIDA (Marine Irish Digital Atlas) metadata record which uses thesauri for keywords. These thesauri are registered in the NERC Vocabulary Server[11] (NVS). This metadata example can be downloaded from the NETMAR WIKI.

Figure 3: GeoNetwork opensource metadata editor
(contains two "Descriptive keywords", each containing a keyword URI registered in NVS)

---

[11] http://vocab.nerc.ac.uk/

The two metadata keyword fields illustrated in Figure 3 are:

- **Keyword name:** Oceanographic geographical features
  NVS URI: http://vocab.nerc.ac.uk/collection/P22/current/28
  Thesaurus: INSPIRE themes

- **Keyword name:** Sea Level Change
  NVS URI: http://vocab.nerc.ac.uk/collection/A04/current/SeaLevelChange
  Thesaurus: MIDA Coastal Erosion Thesaurus

Other metadata fields illustrated in Figure 3 include:

- Dataset's use limitations (e.g. fitness for use) and access constraints (e.g. intellectual property rights)

- Dataset spatial type (i.e. "vector" dataset in the example)

- Dataset scale (i.e. "1:100000" in dataset in the example)

- High level thematic classification (i.e. "Oceans" in the example)

- Dataset's geographic area description, code identifier (i.e. "Northeast Atlantic" in the example) and geographic latitude/longitude coordinates

Metadata and the ICWA prototype

ISO 19115/19119/19139 metadata is required to connect to the International Coastal Web Atlas (ICWA) prototype. Profiles conforming to these standards should, in practice, connect to the ICWA. However, specific ICWA requirements regarding the encoding of semantic keywords are described in detail with snippet examples in the "*Connecting your Atlas to the ICWA prototype*" cookbook. In addition, a CSW (version 2.0.2) metadata catalogue server is required to publish and query metadata via the web. This is described in the "*Establishing a CSW metadata catalogue with GeoNetwork opensource*" cookbook.

Acknowledgements

Document Information

Author	Declan Dunne, Coastal and Marine Research Centre
Contact	d.dunne@ucc.ie
Version	2.1
Date	2012 July 30
Revisions	2.1 Responses to NETMAR internal review
	2.0 Comments from ICAN community
	1.0 2011 December 23

International Coastal Atlas Network Cookbook:

Establishing a CSW metadata catalogue
with GeoNetwork opensource

Table of Contents

Introduction

This document provides a tutorial for those who wish to understand CSW (Catalog Services for the Web) metadata catalogues, with a focus on the ISO Metadata Application Profile of CSW. It is aimed specifically at members of the International Coastal Atlas Network community and more generally at scientists, data managers, and system developers. Included in this document is a description of a metadata catalogue, the CSW standard, and a list of selected CSW severs. The document also contains initial pointers to establishing a CSW server using GeoNetwork opensource, and examples of selected CSW query operations aimed at system developers. GeoNetwork is recommended and used by the NETMAR project.

What is a metadata catalogue?

A metadata catalogue stores and publishes collections or sets of metadata records describing data, services, and related information resources. A user can search for and discover these resources within the metadata catalogue. Distributed metadata catalogues enable searching of metadata catalogues across the Internet between organisations.

What is CSW?

Standards are required to enable interoperable searching of distributed metadata catalogues between organisations. This is achieved using the CSW (Catalog Services for the Web) standard[1]. CSW is an OGC (Open Geospatial Consortium) specification that defines common interfaces and operations to query and retrieve metadata contained in metadata catalogues. It enables a client application to search or query metadata across organisational boundaries. Formal query languages supported by CSW are Filter encoding and CQL (Common Query Language). Filter Encoding is an XML encoded query language, while CQL is a text encoded query language which attempts to be more human readable. Filter encoding is a joint OGC[2] and ISO[3] standard. A simple Filter encoding example is outlined in a later section of this document.

CSW Application Profiles

While the common interfaces and operations of OGC catalogue services are well defined, it is left up to the system developer to define a specific information model for the catalogue service implementation. This includes mandatory and optional metadata elements to be incorporated in the catalogue, supported query languages, available search terms, results, etc. Experience has shown there is no single

---

[1] OpenGIS Catalogue Service: http://www.opengeospatial.org/standards/specifications/catalog
[2] OpenGIS Filter Encoding: http://www.opengeospatial.org/standards/filter
[3] ISO 19143:2010, Geographic information -- Filter encoding:
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=42137

solution for catalogue services that fits every user's needs. Therefore, there is a need for application profiles[4]. OGC has developed these profiles against CSW version 2.0.2:

1. ISO Metadata Application Profile (version 1.0.0)

2. ebRIM Profile (version 1.0.1)

3. OWL Application Profile (version 0.3)

Metadata catalogues that use ISO 19115 and ISO 19119 as its underlying information model should use the ISO Metadata Application Profile of CSW. The Filter encoding query language is mandatory for this profile, with CQL optional. Upon CSW client query requests, such CSW catalogues return ISO 19115 or ISO 19119 metadata records implemented using an ISO 19139 XML schema encoding. Also upon CSW client query requests, the return of Dublin Core encoded metadata is supported by the ISO Metadata Application Profile of CSW. The CSW specification defines the mappings between ISO 19115/19119 and Dublin Core.

CSW Servers

A CSW server is software that implements a metadata catalogue with a CSW interface. A CSW server may also implement ancillary metadata management functionality such as metadata editing tools, harvesting and synchronisation of metadata between distributed catalogues, group and user management, etc.

Several CSW servers have been developed. Commercial CSW servers include:

1. ESRI ArcGIS Server (http://www.esri.com/)

2. Intergraph GeoMedia (http://www.intergraph.com/)

3. MapInfo Manager (http://www.pbinsight.com/)

Opensource/freeware CSW servers include:

1. Constellation (http://constellation.codehaus.org/)

2. Degree (http://www.deegree.org/)

3. GeoNetwork opensource (http://geonetwork-opensource.org/)

4. GI-cat (http://essi-lab.eu/cgi-bin/twiki/view/GIcat/)

GeoNetwork opensource is recommended and used by the NETMAR project as a CSW server. It is open source, supports ISO 19115/19119 and ISO 19139, supports the OGC CSW ISO Metadata Application Profile, has a strong user community, and is used by many government agencies across Europe to support the implementation of the INSPIRE Directive. For static metadata it is recommended to use

---

[4] OpenGIS Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile, Version 1.0.0, OGC 07-045

GeoNetwork as an editing tool, unless specific organisations have special requirements which require the use of an alternative tool. For more dynamic metadata which needs to be automated in a processing chain, it is recommended that a data provider use alternative tools or scripting/programming languages to automate metadata production. However, for such dynamic metadata production environments, GeoNetwork can still be utilised as a CSW server. Guidance notes regarding the installation and configuring of GeoNetwork are outlined in the next section. Methods for metadata ingestion into GeoNetwork are also included.

Installing and configuring GeoNetwork 2.6.4

This section contains some guidance notes regarding the installation and configuring of GeoNetwork opensource version 2.6.4 aimed at system developers. GeoNetwork is an open source implementation of the CSW 2.0.2 ISO Metadata Application Profile 1.0.0 standard. It can run on Microsoft Windows, Linux and Mac OS X. The latest version can be downloaded from:

- http://sourceforge.net/projects/geonetwork/

A detailed user manual can be found on the GeoNetwork opensource website:

- http://geonetwork-opensource.org/manuals/2.6.4/users/

For software developers, a detailed developer manual can also be found on the GeoNetwork opensource website:

- http://geonetwork-opensource.org/manuals/2.6.4/developer/

By default GeoNetwork comes embedded with the Jetty Servlet container. However, Apache Tomcat (version 5.5+) is also supported via a custom install. GeoNetwork requires an RDBMS database in order to store metadata and ancillary information. The default database in GeoNetwork 2.6.4 is an embedded McKoiDB RDBMS database. However, other RDBMS databases including MySQL, PostGreSQL, Oracle and generic JDBC (Java Database Connectivity) connections are supported. It is recommended to use a standalone database instead of McKoiDB in a production environment. These installation instructions deal with configuring the MySQL database.

Installing GeoNetwork 2.6.4

Before installing GeoNetwork 2.6.4, please make sure you have a Java Runtime Environment (JRE 1.5.0+) installed. You can use the GeoNetwork Windows installer (.exe file) for the Windows platforms, or the platform independent installer (.jar file) for any platform (also works on Windows). For example, to start the graphical installation wizard in Windows double click on:

```
geonetwork-install-2.6.4-0.exe
```

To run the platform independent graphical installation wizard double click on:

```
geonetwork-install-2.6.4-0.jar
```

To run the platform independent graphical installation wizard from the command line type:

```
java -jar geonetwork-install-2.6.4-0.jar
```

Once the wizard is running, simply follow the onscreen instructions.

Configuring the database for GeoNetwork 2.6.4

After GeoNetwork 2.6.4 is installed we next need to configure the GeoNetwork database. For these installation instructions we are using MySQL. You can download the open source MySQL Community Server from the MySQL website: http://dev.mysql.com/downloads/. Alternatively on UNIX systems, you may be able to download and install MySQL Community Server via a UNIX distribution stream. Once you have successfully installed the MySQL server, next you can create a dedicated database for GeoNetwork 2.6.4, e.g. by using the following commands:

```
mysql -u root –p
  create database myDatabase;
  CREATE USER 'myUsername'@localhost IDENTIFIED BY 'myPassword';
  GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, INDEX, DROP ON *.*
     TO 'myUsername'@'localhost';
```

To start configuring the GeoNetwork database we need to run GeoNetwork's GAST software - GeoNetwork's Administrator Survival Tool. GAST is a standalone Java application. It performs administration tasks such as the configuration of the JDBC connection to the RDBMS database, database initialisation, etc. More detailed information on GAST can be found in the GeoNetwork user manual.

To start GAST on Windows, simply click:

```
Start GAST (under Start > Programs > GeoNetwork opensource)
```

GAST can also be started by double clicking on:

```
gast.jar (in the .../geonetwork/gast directory)
```

GAST can also be started from the command line by typing:

```
java -jar gast.jar (in .../geonetwork/gast directory)
```

Next, we need to configure and connect GeoNetwork to this new database. Once GAST is running, click the **DBMS** link (Figure 1). In this example, we configure GeoNetwork to connect to the MySQL database that we created previously. Then click the **Save** when finished.
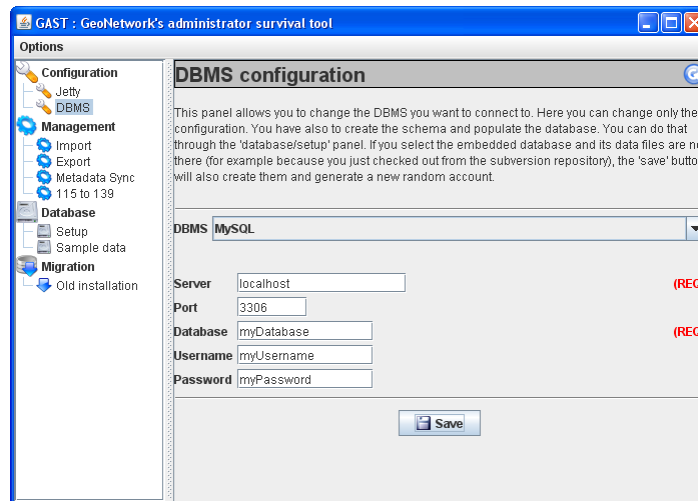
Figure 1 GeoNetwork GAST – DBMS configuration

Next, we need to initialise this new database for GeoNetwork 2.6.4 by creating tables that are required by GeoNetwork. Within the GAST, click the **Setup** link in the left menu, and then the **Setup** button within the main menu (Figure 2). This will recreate and reinitialise GeoNetwork's internal tables (note: this will overwrite and reinitialise existing tables if the database has previously been setup with GAST).
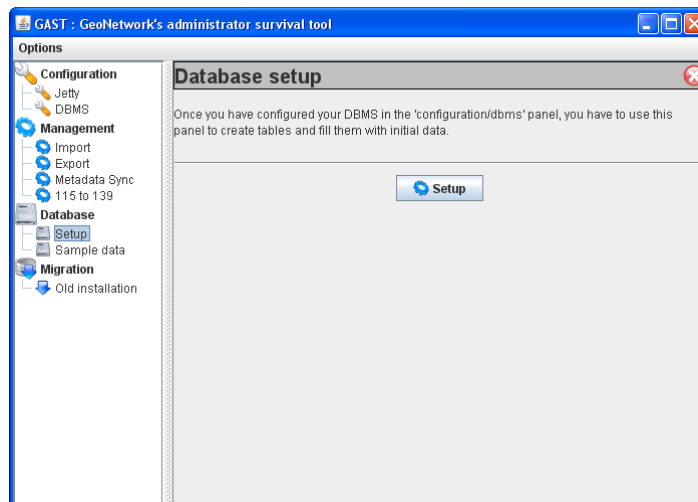

Figure 2 GeoNetwork GAST – Database setup

Note: in GeoNetwork 2.6.4 the following error message may appear in a dialog box when setting up the MySQL database for the first time via the GAST tool:

```
Error: Cyclic reference found:
[relations, categories, settings, languages, sources, isolanguages,
isolanguagesDec, Regions, RegionsDes, Users, Operations,
OperationsDes, Groups, GroupsDes, UsersGroups, CategorisesDes,
Metadata, MetadataCateg, OperationAllowed]
```

This error message can be ignored. Please click "OK" in this dialog box, and the GeoNetwork database setup will continue.

Configure GeoNetwork for Tomcat (optional)

If you wish to deploy GeoNetwork on Apache Tomcat as opposed to the embedded Jetty, then you have a number of options. If you installed GeoNetwork using the steps outlined in the previous sections, then you have two main options to configure for Tomcat:

1. Leave the existing GeoNetwork installation in the ".../geonetwork/web/geonetwork/" directory and point a Tomcat context file (named **geonetwork.xml**) to this directory. An example geonetwork.xml file is:

```
<Context docBase=".../geonetwork/web/geonetwork/"
         privileged="true"
         antiResourceLocking="false"
         antiJARLocking="false">
</Context>
```

The context file is usually placed in the $CATALINA_HOME/conf/Catalina/localhost directory. Make sure that the Tomcat user has the appropriate permissions to access the GeoNetwork installation directory.

2. An alternative option is to copy the installation from ".../geonetwork/web/geonetwork/" to Tomcat's webapps directory. Tomcat will then automatically pick up the GeoNetwork installation. Make sure that the Tomcat user has the appropriate permissions to access this copied installation directory.

If you did not install GeoNetwork using the steps outlined in the previous sections, then you have a third option to configure for Tomcat:

3. Download the **geonetwork.war** file from the GeoNetwork download website. This file can be placed in the Tomcat's webapps directory where it will be automatically deployed by GeoNetwork. However, if you are configuring this deployment with a MySQL database then you still need to create a MySQL database. You also need to install the GAST tool independently to initialise this MySQL database. Once the database is setup you then need to modify the **config.xml** file located in the ".../webapps/geonetwork/WEB-INF" directory.

The default McKoiDB RDBMS database must be removed from **config.xml**:

```
<resource enabled="true">
  <name>main-db</name>
  <provider>jeeves.resources.dbms.DbmsPool</provider>
  <config>
    <user>BayACrsQ</user>
    <password>Qf7Po9T0</password>
```

```
    <driver>com.mckoi.JDBCDriver</driver>
    <url>jdbc:mckoi://localhost:9157/</url>
    <poolSize>10</poolSize>
  </config>
  <activator class="org.fao.geonet.activators.McKoiActivator">
    <configFile>WEB-INF/db/db.conf</configFile>
  </activator>
</resource>
```

The following should be inserted in its place:

```
<resource enabled="true">
  <name>main-db</name>
  <provider>jeeves.resources.dbms.DbmsPool</provider>
  <config>
    <user>myUsername</user>
    <password>myPassword</password>
    <driver>com.mysql.jdbc.Driver</driver>
    <url>jdbc:mysql://localhost/myDatabase</url>
    <poolSize>10</poolSize>
  </config>
</resource>
```

This **config.xml** modification was tested in GeoNetwork 2.6.4. Please ensure Tomcat is shut down when making these changes.

Starting GeoNetwork 2.6.4 (required for Jetty)

If you installed GeoNetwork using the embedded Jetty and successfully configured the database, then you can start GeoNetwork. On Windows click:

```
Start Server (under Start > Programs > GeoNetwork opensource)
```

Or from the Windows command line by typing:

```
sh start-geonetwork.bat (in .../geonetwork/bin directory)
```

Under UNIX, you can start GeoNetwork from the command line by typing:

```
sh start-geonetwork.sh (in .../geonetwork/bin directory)
```

Once started (can take up to 1 minute), you can access GeoNetwork in your web browser at:

- http://localhost:8080/geonetwork/

Stopping GeoNetwork 2.6.4 (required for Jetty)

If you installed GeoNetwork using the embedded Jetty, you can stop GeoNetwork on Windows by clicking:

```
Start Server (under Start > Programs > GeoNetwork opensource)
```

Or from the Windows command line by typing:

```
sh stop-geonetwork.bat (in .../geonetwork/bin directory)
```

Under UNIX, you can stop GeoNetwork from the command line by typing:

```
sh stop-geonetwork.sh (in .../geonetwork/bin directory)
```

GeoNetwork 2.6.4 Administration

Once GeoNetwork is up and running, an important first time administration step is to change the default GeoNetwork administrator username/password which is admin/admin. The first time you start GeoNetwork, login as the administrator and select the "Administration" link in the tab bar. Then reset the administrator password via the "User management" link. For more information on all GeoNetwork 2.6.4 administration features, please consult the GeoNetwork manual.

Adding metadata records in GeoNetwork 2.6.4

There are a number of ways to add metadata records to the GeoNetwork 2.6.4 database. You must login with appropriate privileges to add metadata. These methods are documented in the user manual. A summary of these methods are outlined here:

1. **Creating a new metadata record with the GeoNetwork metadata editor:**
   You can choose this method by first selecting the "Administration" link in the tab bar, and then clicking "New metadata". Next, you need to select an appropriate metadata template. The ISO 19139 templates for vector and raster data are preferred. Once you select the template you can start editing metadata (e.g. Figure 1). Note, when setting up GeoNetwork, you may wish to add the default metadata templates as they are not activated by default. This can be done via "Add templates" under the "Administration" tab. You can also create your own template via the metadata editor. In this case, when saving the new metadata template in the editor you choose "Template" from the "Type" dropdown list.

2. **Uploading pre-existing metadata with the GeoNetwork metadata insert tool:**
   You can choose this method by first selecting the "Administration" link in the tab bar, and then clicking "Metadata insert". You can choose to upload the XML file directly or else to paste the XML contents into a textbox. This method is useful for uploading pre-existing metadata created by other tools or scripts. You can also choose to upload the XML as a metadata template by selecting "Template" from the "Type" dropdown list.
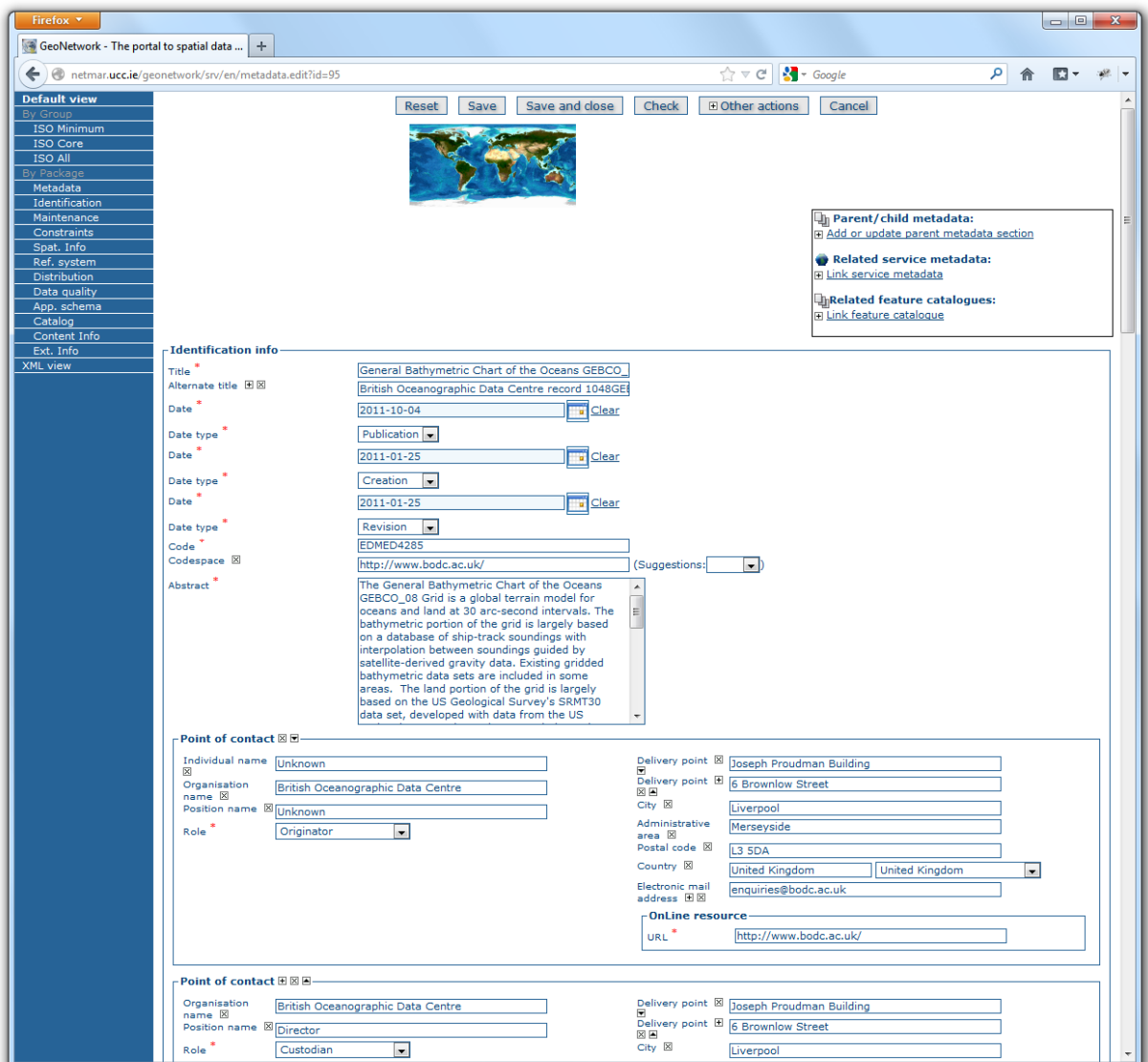
Figure 1: GeoNetwork opensource metadata editor

3. **Batch uploading pre-existing metadata from a server-side directory:**
You can choose this method by first selecting the "Administration" link in the tab bar, and then clicking "Batch Import". Next, you enter the full directory path located on the server's file system. Once activated, GeoNetwork will scan this directory and load all metadata records contained in the directory. This is a useful method where a user can batch upload numerous pre-existing metadata records in one go, where these metadata records have typically been created by another tool or script.

4. **Harvesting pre-existing metadata:**
You can choose this method by first selecting the "Administration" link in the tab bar, and then clicking "Harvesting management". Click "Add" to configure a new metadata harvesting task. GeoNetwork is able to harvest from a number of remote sources including: another

GeoNetwork node, a CSW server, a WebDAV server, an OAI-PMH server, a local file system, etc. You can decide to schedule the harvesting task (e.g. every hour, every week, etc.) or just run a once-off harvest. Harvesting is a useful method for collecting remote metadata and storing them locally for faster access. It is also a useful method for the automatic pulling of local file system metadata into the catalogue, where this metadata is being dynamically produced by an independent processing chain.

5. **XML services (advanced):**
    GeoNetwork provides access to several functionalities through the use of XML web services. These web services support the insertion of metadata records into the GeoNetwork database. There are two main methods: Metadata Service[5] and CSW Service[6]. With the Metadata Service, metadata can be added using the "metadata.insert" operation. With the CSW Service, metadata can be added using the CSW "Transaction" operation. The CSW Service is more standards compliant as it uses the OGC CSW 2.0.2 specification. These two metadata insertion methods are useful for the automatic pushing of metadata into the catalogue by a processing chain, where this metadata is being dynamically produced by the same processing chain. These two methods require software developer skills to implement. Further details, with examples, are documented in the developer manual.

Examples of CSW operations

This section gives examples of the CSW GetRecordById and GetRecords operation requests aimed at system developers who wish to understand better, explore and test the engineering details of CSW in more detail using examples. Full details of these operations can be referenced in CSW 2.0.2 - ISO Metadata Application Profile v1.0.0.

CSW GetRecordById operation request

This operation enables a client CSW application to request one or more specific metadata records from the CSW server using their metadata identifiers (mapped to the fileidentifier of an ISO 19139 document). This example, tested against GeoNetwork 2.6.4, requests a full ISO 19139 metadata record with an identifier of "64c8493d6bd95d93b7e04fb868fd568e":

```
http://netmar.ucc.ie/geonetwork/srv/en/csw?
    SERVICE=CSW&
    REQUEST=GetRecordById&
    VERSION=2.0.2&
    elementSetName=full&
    outputSchema=csw:IsoRecord&
    ID=64c8493d6bd95d93b7e04fb868fd568e
```

---

[5] http://www.geonetwork-opensource.org/manuals/2.6.4/developer/xml_services/metadata_xml_services.html
[6] http://www.geonetwork-opensource.org/manuals/2.6.4/developer/xml_services/csw_services.html

The example below contains the response with an extract of the requested ISO 19139 metadata record:

```
<csw:GetRecordByIdResponse>
  <gmd:MD_Metadata>
    <gmd:fileIdentifier>
      <gco:CharacterString>
        64c8493d6bd95d93b7e04fb868fd568e
      </gco:CharacterString>
    </gmd:fileIdentifier>
    .
    .
    .
  </gmd:MD_Metadata>
</csw:GetRecordByIdResponse>
```

CSW GetRecords operation request

This operation enables a client CSW application to search for metadata in the catalogue service. This example, tested against GeoNetwork 2.6.4, requests the CSW server to return full ISO 19139 metadata for any records which contain the word "GEBCO_08" (note: the XML Filter below needs to be URL encoded before pasting into a web browser):

```
http://netmar.ucc.ie/geonetwork/srv/en/csw?
    SERVICE=CSW&
    REQUEST=GetRecords&
    VERSION=2.0.2&
    resultType=results&
    elementSetName=full&
    outputSchema=http://www.isotc211.org/2005/gmd&
    typeNames=csw:Record&
    constraintLanguage=FILTER&
    constraint_language_version=1.1.0&
    constraint=
      <Filter xmlns="http://www.opengis.net/ogc"
              xmlns:gml="http://www.opengis.net/gml">
        <PropertyIsLike>
          <PropertyName>any</PropertyName>
          <Literal>GEBCO_08</Literal>
        </PropertyIsLike>
      </Filter>
```

The example below contains the response with an extract of the one ISO 19139 metadata record that was discovered:

```
<csw:GetRecordsResponse>
  <csw:SearchStatus timestamp="2012-07-26T17:07:29"/>
  <csw:SearchResults numberOfRecordsMatched="1"
                     numberOfRecordsReturned="1"
                     elementSet="full"
                     nextRecord="0">
    <gmd:MD_Metadata
```

```
      .
      .
      .
    <gmd:title>
      <gco:CharacterString>
        General Bathymetric Chart of the Oceans GEBCO_08 Grid
      </gco:CharacterString>
    </gmd:title>
      .
      .
      .
    </gmd:MD_Metadata>
  </csw:SearchResults>
</csw:GetRecordsResponse>
```

Acknowledgements

Document Information

Author	Declan Dunne, Coastal and Marine Research Centre
Contact	d.dunne@ucc.ie
Version	2.1
Date	2012 July 27
Revisions	2.1 Responses to NETMAR internal review
	2.0 Comments from ICAN community
	1.0 2011 December 23

International Coastal Atlas Network Cookbook:

Connecting your Atlas to the
ICWA prototype

Table of Contents

Introduction

This document is a step-by-step guide explaining how to connect your atlas as a node in the International Coastal Web Atlas (ICWA) prototype hosted at http://ican.ucc.ie/. This document is aimed specifically at members of the International Coastal Atlas Network community and more generally at scientists, data managers, and system developers.

The ICWA prototype

The ICWA prototype is developed by the Technical Working Group (TWG) of the International Coastal Atlas Network (ICAN), with the current prototype (version 3) being funded by the NETMAR project. It provides a common interface for accessing distributed local atlases, such as MIDA (Marine Irish Digital Atlas), OCA (Oregon Coastal Atlas), and Washington Coastal Atlas (WCA). The current version of ICWA only supports catalogue search, including "smart search". The ICAN TWG are currently investigating the extension of ICWA to support web mapping. ICWA version 3 is based on five components (Figure 1):

- Ontology Browser (OB)

- Discovery interface (Geo Finder)

- Metadata viewer (Meta Viewer)

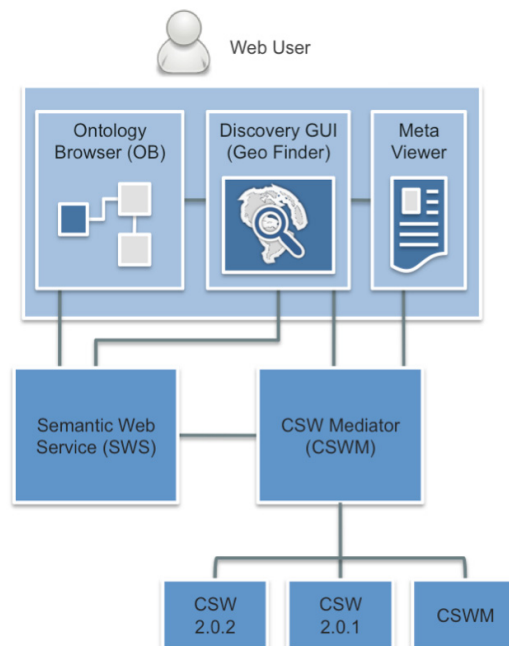- ICWA Mediator (CSW Mediator)

- Semantic Web Service (SWS)

Figure 1: Main components of the ICWA prototype

The first three components are frontend graphical user interfaces hosted at http://ican.ucc.ie/. A user can interact with these components in a web browser, and perform tasks such as ontology browsing, data search and metadata visualisation across distributed atlases.

The other components are backend web services. The ICWA mediator (CSW Mediator) communicates with local atlases through standard OGC CSW (Catalogue Service for the Web) interfaces. It queries the knowledge organisation system (KOS) through the semantic web service (SWS) interface. ICWA uses a knowledge organisation system to improve data discovery by exploiting the semantics of keywords and allowing users to search data by "meaning" rather than by "mere keywords".

For example, as illustrated in Figure 2, a user arrives at the ICAN portal and requests "coastline" data. The ICWA mediator is connected to a global knowledge organisation system that is aware that "coastline" is related to both "shoreline" and "high resolution coastline". The user request, together with this information from the global knowledge organisation system, is then passed on to the local atlases that search for "coastline", "shoreline" and "high resolution coastline". Each local atlas then returns the relevant metadata results to the ICWA mediator. In turn, these individual metadata results are aggregated and passed to the graphical user interface. This is an implementation of the so-called "smart-search"[1].
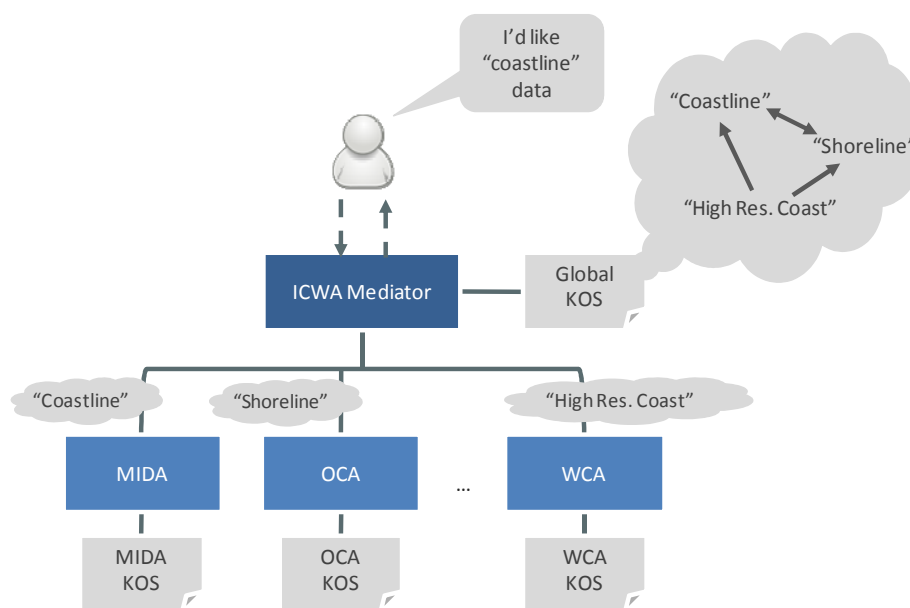

Figure 2: Illustrating the use for knowledge organisation systems in the ICWA

---

[1] Latham, S. E.; Cramer, R.; Grant, M.; Kershaw, P.; Lawrence, B. N.; Lowry, R.; Lowe, D.; O'Neill, K.; Miller, P.; Pascoe, S.; Pritchard, M.; Snaith, H.; Woolf, A. (2009) The NERC DataGrid services. *Philosophical Transactions of the Royal Society A*, 367 (1890). 1015-1019.

Connection Requirements

In order to connect your atlas as a node in the ICWA, you need to:

1. Provide access to your metadata through a CSW[2] interface supporting the CSW ISO Metadata Application Profile[2], which supports both Dublin Core[3] and the ISO 19115[4]/19119[5]/19139[6] metadata standards.

2. Optionally, provide your local KOS that defines the terms used as keywords in your metadata records and specifies how they semantically relate to the ICAN global KOS terms.

The subsequent sections explain these two requirements. Any feedback regarding these requirements and technology compatibility to implement and test these requirements is welcome, and may be provided by contacting the author whose details are listed at the end of this document.

CSW Requirements

The ICWA mediator supports CSW 2.0.2 fully and CSW 2.0.1 partially. Therefore, we recommend that you use CSW 2.0.2 for delivering your metadata to the ICWA mediator.

The ICWA mediator requires CSW 2.0.2 supporting the ISO Metadata Application Profile version 1.0.0, that is able to deliver metadata in Dublin Core and ISO 19115/19139. ISO 19119/19139 service metadata is also supported by the ICWA mediator.

The ICWA mediator has been tested with GeoNetwork 2.4 and later versions. GeoNetwork automatically provides a mapping from ISO 19115/19119 to Dublin Core. The mediator should support other CSW 2.0.2 implementations. Currently, all ICAN CSW nodes use GeoNetwork, and ICWA functionality is tested against this. Other CSW servers will be tested in future, but please contact the author if you have a particular CSW server that requires testing.

There is no particular requirement as to the ISO 19115/19119 metadata profile used, but metadata records need to include keywords expressed in one of the following four encoding forms:

1. As character strings (gco:CharacterString), each containing the URI of a concept from your local KOS (if any). This encoding form is currently the favoured recommendation as it uses a direct URI link to the semantic keyword concept. An example:

```
<gmd:descriptiveKeywords>
  <gmd:MD_Keywords>
    <gmd:keyword>
      <gco:CharacterString>
        http://vocab.nerc.ac.uk/collection/A03/current/Beach
      </gco:CharacterString>
    </gmd:keyword>
    <gmd:thesaurusName>
```

---

[2] http://www.opengeospatial.org/standards/cat/
[3] http://www.dublincore.org/
[4] http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020
[5] http://www.iso.org/iso/catalogue_detail.htm?csnumber=39890
[6] http://www.iso.org/iso/catalogue_detail.htm?csnumber=32557

```
        <gmd:CI_Citation>
          <gmd:title>
            <gco:CharacterString>
              Oregon Coastal Atlas Coastal Erosion Thesaurus discovery terms
            </gco:CharacterString>
          </gmd:title>
          <gmd:date>
            <gmd:CI_Date>
              <gmd:date>
                <gco:Date>2011-08-19</gco:Date>
              </gmd:date>
              <gmd:dateType>
                <gmd:CI_DateTypeCode
codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
resources/Codelist/gmxCodelists.xml#CI_DateTypeCode" codeListValue="revision">
                  Revision
                </gmd:CI_DateTypeCode>
              </gmd:dateType>
            </gmd:CI_Date>
          </gmd:date>
        </gmd:CI_Citation>
      </gmd:thesaurusName>
  </gmd:MD_Keywords>
</gmd:descriptiveKeywords>
```

2. As character strings, each containing the label of a concept from your local KOS (if any). However, this does not support a direct link to the concept. Instead, the mediator must search all the local KOS for the relevant concept based on the label. Therefore, this encoding form is not preferred. An example:

```
<gmd:descriptiveKeywords>
  <gmd:MD_Keywords>
    <gmd:keyword>
      <gco:CharacterString>Beach</gco:CharacterString>
    </gmd:keyword>
    <gmd:thesaurusName>
      <gmd:CI_Citation>
        <gmd:title>
          <gco:CharacterString>
            Oregon Coastal Atlas Coastal Erosion Thesaurus discovery terms
          </gco:CharacterString>
        </gmd:title>
        <gmd:date>
          <gmd:CI_Date>
            <gmd:date>
              <gco:Date>2011-08-19</gco:Date>
            </gmd:date>
            <gmd:dateType>
              <gmd:CI_DateTypeCode
codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
resources/Codelist/gmxCodelists.xml#CI_DateTypeCode" codeListValue="revision">
                Revision
              </gmd:CI_DateTypeCode>
            </gmd:dateType>
          </gmd:CI_Date>
        </gmd:date>
      </gmd:CI_Citation>
    </gmd:thesaurusName>
  </gmd:MD_Keywords>
</gmd:descriptiveKeywords>
```

3. As anchor elements (gmx:Anchor), each containing both the URI and label of a concept from your local KOS (if any). This encoding form is currently not supported by GeoNetwork (version

2.6.4 and earlier) and, therefore, has not been tested by ICWA. In the long term this will be the preferred encoding form as it supports both a direct URI link to the concept and a label of this concept. An example:

```
<gmd:descriptiveKeywords>
  <gmd:MD_Keywords>
    <gmd:keyword>
      <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/A03/current/Beach">
        Beach
      </gmx:Anchor>
    </gmd:keyword>
    <gmd:thesaurusName>
      <gmd:CI_Citation>
        <gmd:title>
          <gco:CharacterString>
            Oregon Coastal Atlas Coastal Erosion Thesaurus discovery terms
          </gco:CharacterString>
        </gmd:title>
        <gmd:date>
          <gmd:CI_Date>
            <gmd:date>
              <gco:Date>2011-08-19</gco:Date>
            </gmd:date>
            <gmd:dateType>
              <gmd:CI_DateTypeCode
codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
resources/Codelist/gmxCodelists.xml#CI_DateTypeCode" codeListValue="revision">
                Revision
              </gmd:CI_DateTypeCode>
            </gmd:dateType>
          </gmd:CI_Date>
        </gmd:date>
      </gmd:CI_Citation>
    </gmd:thesaurusName>
  </gmd:MD_Keywords>
</gmd:descriptiveKeywords>
```

4. As character strings containing free text keywords. This is the simplest encoding form. It is not recommended because of its lack of a KOS, but it is supported to enable non-semantic metadata catalogues to be readily connected to the ICWA. Therefore, smart search functionality will not be supported for this encoding form.

```
<gmd:descriptiveKeywords>
  <gmd:MD_Keywords>
    <gmd:keyword>
      <gco:CharacterString>Beach</gco:CharacterString>
    </gmd:keyword>
  </gmd:MD_Keywords>
</gmd:descriptiveKeywords>
```

KOS Requirements

Local knowledge organisation systems are used in ICWA to define the semantics of metadata keywords used within a local atlas and how these relate semantically to the global KOS concepts. A local KOS is a SKOS thesaurus organised into one or many concept schemes and collections. It defines:

- The concepts represented by keywords (themes, parameters, instruments, etc.), their labels (terms), and definitions (free text explaining the meaning of the concepts).

- The semantic relationships between the local concepts.

- The semantic relationships between the local concepts and the global ones. These relationships are referred to as mappings.

You are not required to define your metadata keywords in a KOS. However, you are recommended to do so as this would improve data search as explained earlier in this document.

The ICWA mediator has a built-in SWS connector for querying global and local KOS. Therefore, for your KOS to be accessed and used by the mediator, it must be delivered through SWS. You can implement your own SWS (source code available from CMRC), but this means an extra maintenance load for you. BODC are hosting and maintaining an operational vocabulary server (NERC Vocabulary Server, NVS) that feeds the NETMAR SWS hosted by CMRC. You are strongly recommended to profit from this existing infrastructure and provide BODC with your KOS for inclusion in the NVS. BODC have made available two worksheets in an Excel document for defining your concepts, relationships, and mappings in a very intuitive way, avoiding the hassle of RDF, SKOS and ontology editing tools. Please refer to the "*Understanding Semantics*" cookbook for examples of these worksheets.

Once you have finished filling out the worksheets, BODC can load it onto the NVS.

What is Next?

If your atlas fills the requirements above (CSW and KOS), adding it as a node in ICWA is a very easy process. All you need to do is provide the ICWA administrator (CMRC) with the following details:

1. URL of your CSW server;

2. Login and password for accessing metadata through your CSW server (if metadata records are not public);

3. Optionally, but ideally, a logo of your atlas (250x250 PNG image);

4. Method used for encoding keywords (c.f., methods 1 to 4 listed in the CSW Requirements section above);

5. If you are using method 1, 2 or 3 for keyword encoding then you will need to provide the ICWA with the following information:

   a. URL of the SWS responsible for delivering your KOS if other than the NVS and the CMRC SWS;

   b. URI of the concept schemes containing the concepts used as metadata keywords.

Once provided with this information, the ICWA administrator can add your atlas in the ICWA configuration file and it becomes an active node of ICWA.

Acknowledgements

Document Information

Author	Yassine Lassoued, Coastal and Marine Research Centre
Contact	y.lassoued@ucc.ie
Version	2.1
Date	2012 July 30
Revisions	2.1 Responses to NETMAR internal review
	2.0 Comments from ICAN community
	1.0 2011 December 22