



Project No. 249024

NETMAR

Open service network for marine environmental data

<b>Instrument:</b> <i>Please tick</i>	CA	<input checked="" type="checkbox"/> STREP	IP	NOE
--	----	---	----	-----

**ICT - Information and Communication Technologies Theme**

**D7.9.1 ICAN semantic interoperability pilot cookbooks**

Reference: D7.9.1\_ICAN\_semantic\_cookbooks\_r1\_20111229

Due date of deliverable (as in Annex 1): M0 + 23  
Actual submission date: 29 December 2011

Start date of project: 1 February 2010

Duration: 3 years

Coastal and Marine Resources Centre (CMRC)

Revision 1

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
<b>PU</b>	Public	X
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission Services)	
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	



 	<p><b>NETMAR</b>  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</p>
	<p>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: <a href="mailto:stein.sandven@nersc.no">stein.sandven@nersc.no</a></p>

### Acknowledgements

The work described in this report has been partially funded by the European Commission under the Seventh Framework Programme, Theme ICT 2009.6.4 ICT for environmental services and climate change adaptation.

### 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](mailto:stein.sandven@nersc.no))  
Deputy Coordinator: Dr. Torill Hamre ([torill.hamre@nersc.no](mailto:torill.hamre@nersc.no))  
Quality Control Manager: Mr. Lasse H. Petterson ([lasse.petterson@nersc.no](mailto:lasse.petterson@nersc.no))
- British Oceanographic Data Centre (BODC), National Environment Research Council, United Kingdom  
Contact: Dr. Roy Lowry ([rkl@bodc.ac.uk](mailto: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](mailto: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](mailto:d.dunne@ucc.ie))
- Plymouth Marine Laboratory (PML), United Kingdom.  
Contact: Mr. Steve Groom ([sbg@pml.ac.uk](mailto:sbg@pml.ac.uk))
- Institut français de recherche pour l'exploitation de la mer (Ifremer), France.  
Contact: Mr. Mickael Treguer ([mickael.treguer@ifremer.fr](mailto:mickael.treguer@ifremer.fr))
- Norwegian Meteorological Institute (METNO), Norway.  
Contact: Mr. Øystein Tøtger ([oysteint@met.no](mailto:oysteint@met.no))

### Author(s)

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

### Document approval

- Document status: Revision 1
- WP leader approval: 23 December 2011
- Quality Manager approval: 29 December 2011
- Coordinator approval: 29 December 2011

## Revision History

<b>Issue</b>	<b>Date</b>	<b>Change records</b>	<b>Author(s)</b>
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

## 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.

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, 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.

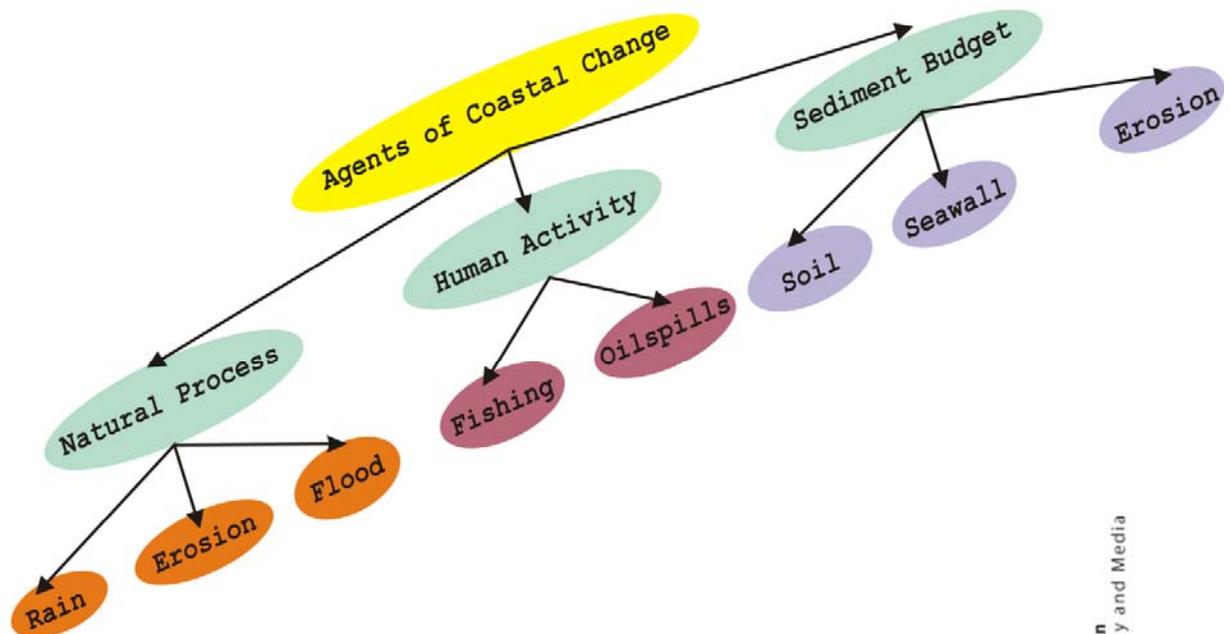
The section of this document called “Understanding Semantics” provides a tutorial for those who wish to investigate and make use of these technologies, aimed specifically at members of the International Coastal Atlas Network community and more generally at scientists and data managers.

The section of this document called “Understanding Metadata” 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 contains an example metadata record aimed at system developers who are familiar with XML.

The section of this document called “Establishing a CSW metadata catalogue with GeoNetwork opensource” 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 some CSW servers available. The document also contains initial pointers to establishing a CSW server using GeoNetwork opensource and examples of some simple CSW query operations aimed at system developers. GeoNetwork is recommended and used by the NETMAR project.

The section of this document called “Connecting your Atlas” provides a step-by-step guide explaining how to connect a local atlas as a node in the International Coastal Web Atlas (ICWA). ICWA 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). ICWA uses a knowledge organisation system (KOS) to improve data discovery by exploiting the semantics of keywords and allowing users to search data by “meaning” rather than by “mere keywords.”

## International Coastal Atlas Network Cookbook: Understanding Semantics



European Commission  
Information Society and Media

## Table of Contents

Introduction .....	3
Why use a “knowledge organization system”? .....	3
What are vocabularies, thesauri and ontologies?.....	4
How to define the content of a knowledge organization system? .....	6
Can I reuse existing resources? .....	6
What is the scope of the knowledge organization system? .....	6
Identifying the content .....	6
How narrow or broad should a term definition be? .....	6
Linking term definitions together .....	6
Ensuring the quality of the content of the Knowledge Organization System .....	7
Making the content available.....	8
Simple Knowledge Organization System .....	8
Deploying ICAN semantics in the NETMAR semantic framework .....	9
Bridging to existing Knowledge Organization Systems.....	12
Acknowledgements .....	13
Document Information .....	13

## 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<sup>1</sup>*

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, 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 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 scientists and data managers.

### Why use a "knowledge organization system"?

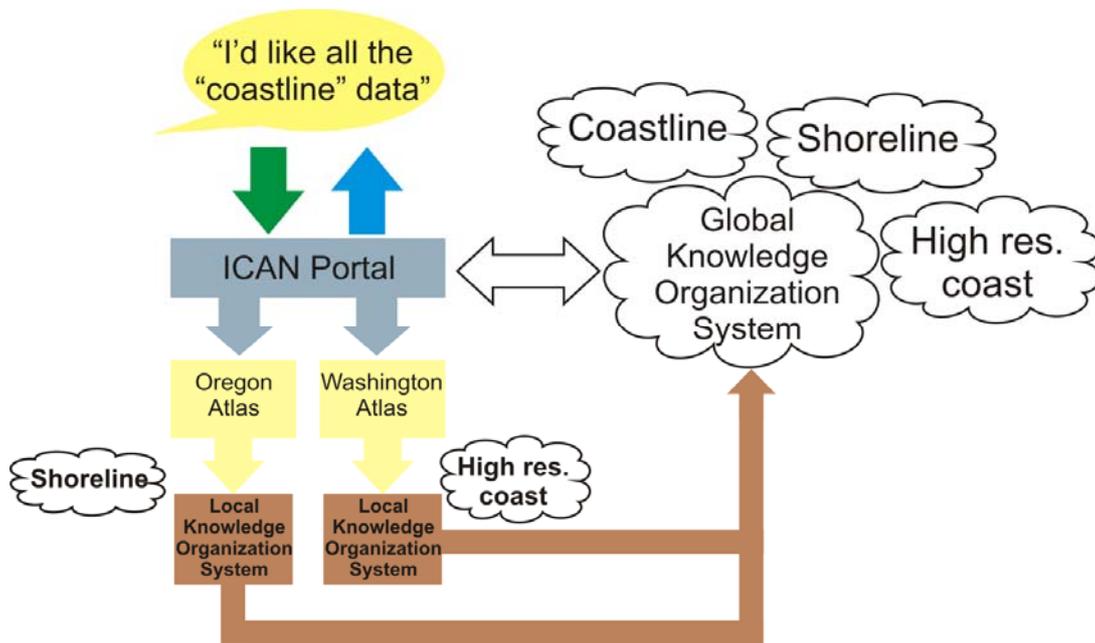
One scenario for using knowledge organization systems in the International Coastal Atlas Network<sup>2</sup> (ICAN) is to search through the local atlases for a given data keyword from a central portal. For example, as illustrated below, a user arrives at the ICAN portal and request "coastline" data. 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"<sup>3</sup>.

---

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

<sup>2</sup> <http://ican.science.oregonstate.edu/>

<sup>3</sup> 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.

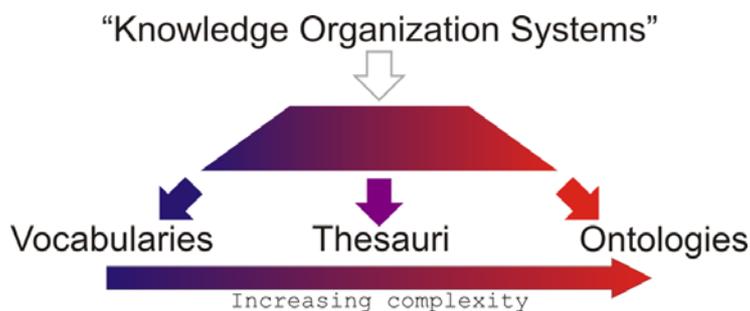


A diagram illustrating one use for knowledge organization systems in the ICAN community.

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)<sup>4</sup>.

<sup>4</sup> 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. This provides the simplest form of a formal ontology. A well known example of this form is the Yahoo! web directory<sup>5</sup> or the categorisation of auctions on the eBay homepage<sup>6</sup>. 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<sup>7</sup> 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.

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”.

---

<sup>5</sup> <http://dir.yahoo.com/>

<sup>6</sup> <http://www.ebay.com/>

<sup>7</sup> <http://gcmd.nasa.gov/>

## How to define the content of a knowledge organization system?

### 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 12). Details of how to access an existing knowledge organization system relevant to the International Coastal Atlas Network are provided on page 11 of this document.

### 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<sup>8,9</sup>. 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

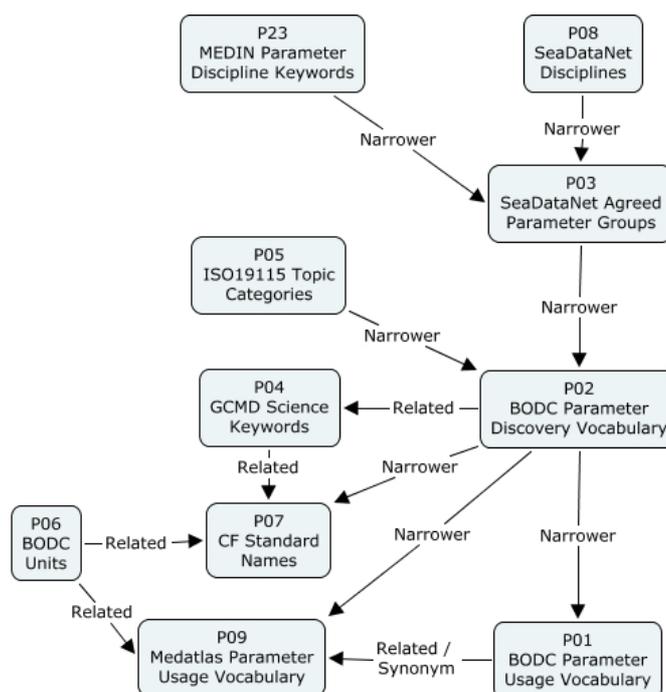
As described above, the definition of terms by themselves is useful but the impact of the work can be greatly extended by providing relationships which link the terms together to form networks of

---

<sup>8</sup> 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.

<sup>9</sup> 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.

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.



An example from the NERC Vocabulary Server<sup>10</sup> 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<sup>11</sup> netCDF metadata conventions group; the SeaDataNet and MarineXML Vocabulary Content Governance Group (SeaVoX)<sup>12</sup>; and the NETMAR ontology governance body<sup>13</sup>. The

<sup>10</sup> <http://vocab.nerc.ac.uk/>

<sup>11</sup> <http://cf-pcmdi.llnl.gov/>

<sup>12</sup> [https://www.bodc.ac.uk/data/codes\\_and\\_formats/seavox/](https://www.bodc.ac.uk/data/codes_and_formats/seavox/)

<sup>13</sup> <http://netmar.nerc.no/>

role of the content governance group is analogous to the International Organization for Standardization (ISO) definition of a “control body”<sup>14</sup>.

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<sup>10</sup> (NVS). The role of the technical governance group is analogous to the ISO definition of a “register manager”<sup>14</sup>.

## 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<sup>15</sup> (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 also a part of the concept scheme, so it is a useful model for the publication of thesauri, for example the “ICAN coastal erosion thesaurus.”

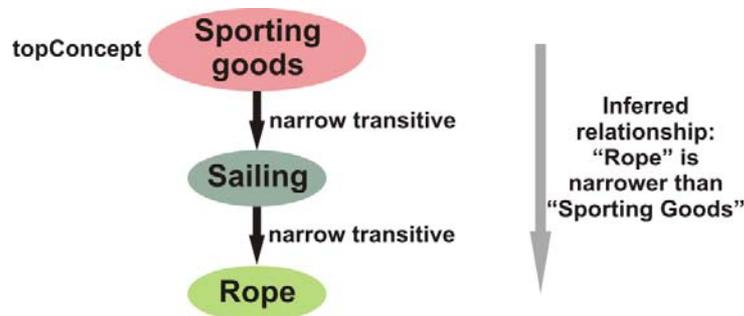
SKOS also defines three forms of relationship between concepts. A concept may be broader or narrower than another concept, or related to another concept. The related attribute allows the loose mapping of one concept to another, allowing the resource to become orthogonal (see page 5). The broader and 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

---

<sup>14</sup> <http://www.dgiwg.org/Terminology/faq-other.php>

<sup>15</sup> <http://www.w3.org/2004/02/skos/>

to be narrower than “Sporting Goods”, which is not explicit in the non-transitive SKOS narrower relationship.



An illustration of transitive relations in SKOS using terms from the eBay classification of auctions.

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<sup>16</sup> `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.

---

<sup>16</sup> <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 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. 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	<a href="http://vocab.nerc.ac.uk/collection/L06/current/31/">http://vocab.nerc.ac.uk/collection/L06/current/31/</a> ("research vessel")
74PQ ("Plymouth Quest")	Is narrower than	<a href="http://vocab.nerc.ac.uk/collection/L19/current/SDNKG04">http://vocab.nerc.ac.uk/collection/L19/current/SDNKG04</a> ("platform")

Once complete, the spreadsheet should be submitted to [enquiries@bodc.ac.uk](mailto: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<sup>17</sup>.

### *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<sup>18</sup> (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/](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/](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/](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

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/deprecated/)

---

<sup>17</sup> [https://www.bodc.ac.uk/data/codes\\_and\\_formats/vocabulary\\_editor/](https://www.bodc.ac.uk/data/codes_and_formats/vocabulary_editor/)

<sup>18</sup> <http://en.wikipedia.org/wiki/Url>



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<sup>22</sup>; the European Environment Agency General Multilingual Environmental Thesaurus<sup>23</sup>; and GeoNames<sup>24</sup>. Relationships between a concept in the NVS and any external concept can be specified in the same way as the internal mappings (see page 6) 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/
```

## Acknowledgements

This cookbook was written 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.

This document has been reviewed by, and incorporates comments from, Jennifer Andrew and Roy Lowry of the British Oceanographic Data Centre; Torill Hamre of the Nansen Environmental and Remote Sensing Center; Yassine Lassoued of the Coastal and Marine Research Centre, University College Cork; and François Parthiot of CEDRE. Thanks go to the reviewers for their help in making the document clear and readable. Further feedback on this document is welcomed, and may be provided by contacting the author whose details are below.

## Document Information

<b>Author</b>	Adam Leadbetter, British Oceanographic Data Centre
<b>Contact</b>	alead@bodc.ac.uk
<b>Version</b>	1.0
<b>Date</b>	2011 December 16
<b>Revisions</b>	

---

<sup>21</sup> <http://richard.cyganiak.de/2007/10/lod/imagemap.html>

<sup>22</sup> <http://mmisw.org/orr/>

<sup>23</sup> <http://www.eionet.europa.eu/gemet>

<sup>24</sup> <http://www.geonames.org/>

**International Coastal Atlas Network Cookbook:  
Understanding Metadata**



## Table of Contents

Introduction .....	3
What is metadata? .....	3
Why do you need metadata?.....	3
Metadata standards and profiles .....	4
Metadata hierarchy levels .....	5
Metadata editing tools .....	6
Metadata encoding guidance .....	7
Dataset metadata example.....	8
Acknowledgements .....	20
Document Information .....	20

## Introduction

This document provides a tutorial for those who wish to understand metadata. 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 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 document also contains an example metadata record aimed at system developers who are familiar with XML.

## 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<sup>1</sup>:

- What: Title and description of the data.
- Why: Abstract detailing reasons for the 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 the user 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 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

---

<sup>1</sup> Wilson, M., 2009, Chapter Three: Metadata -- Describing geospatial data, Spatial Data Infrastructure Cookbook.

enabling more efficient and cost-effective use of this data. Specific examples of business cases for using metadata outlined by Federal Geographic Data Committee (FGDC) include<sup>2</sup>:

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 resource development progress.
- Share data resource 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)
- FGDC Content Standard for Digital Geospatial Metadata (CSDGM)

---

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

The Dublin Core Metadata Element Set contains fifteen properties capable of describing a wide range of general web resources. While Dublin Core can be successfully applied to describe geospatial resources, the ISO 19115 standard is more specialised in describing such geospatial resources. In turn, the ISO 19119 standard extends the ISO 19115 to describe geospatial service resources. Therefore, governments and organisations are widely adapting the ISO 19115 and ISO 19119 standards using the ISO 19139 XML implementation. 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 adapted to support various regions, nations, organisations, and communities' needs.

Within Europe, the INSPIRE Directive has defined a base metadata profile, which builds on the underlying ISO 19115 standard. However, full conformance to the ISO 19115 mandatory elements implies the provision of additional metadata elements which are not required by INSPIRE legislation. However INSPIRE metadata encoding guidelines respects these ISO 19115 mandatory elements. In addition, INSPIRE is defining thematic dataset specifications. Individually these dataset specifications have defined a small additional number of mandatory and optional elements on top of base metadata profile to help data evaluation<sup>3</sup>.

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<sup>4</sup>.

## 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.

---

<sup>3</sup> <http://inspire.jrc.ec.europa.eu>

<sup>4</sup> Preparing for International Metadata, Federal Geographic Data Committee, October 20, 2011

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.

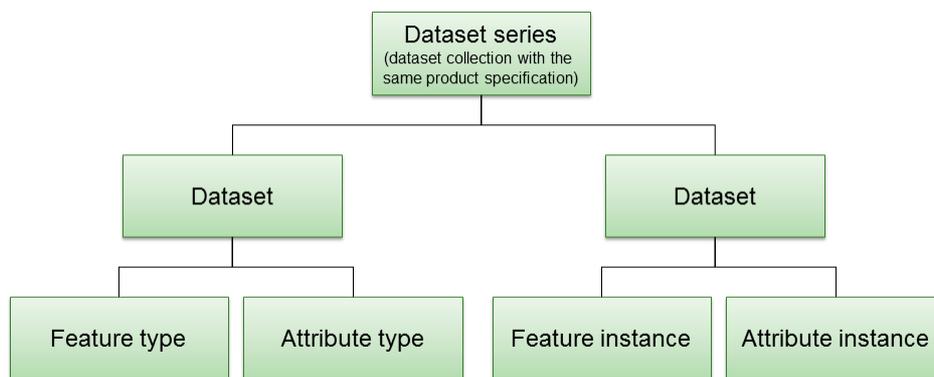


Figure 1: Metadata hierarchy<sup>5</sup>

## Metadata editing tools

A metadata editor is a program that is used for creating and editing metadata. It typically uses a graphical user interface, as opposed to direct editing of the XML document. A metadata editing tool may include functionalities such as:

- Creating, editing, deleting and viewing of metadata and metadata templates
- Metadata 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. Commercial metadata editing tools include:

---

<sup>5</sup> ISO 19115:2003, Geographic information - Metadata

1. ESRI ArcGIS Desktop (<http://www.esri.com/>)
2. Intergraph GeoMedia (<http://www.intergraph.com/>)
3. MapInfo Manager (<http://www.pbinsight.com/>)

Opensource/freeware metadata editing tools include:

1. CatMDEdit (<http://catmdedit.sourceforge.net/>)
2. GeoNetwork opensource (<http://geonetwork-opensource.org/>)
3. M<sup>3</sup>Cat (<http://www.intelec.ca/>)

A screenshot of the GeoNetwork opensource metadata editor is shown below:

Figure 2: GeoNetwork opensource metadata editor

## Metadata encoding guidance

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

19139 schemas define the structure of the metadata XML document. 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 1.0, September 2010  
<http://location.defra.gov.uk/wp-content/uploads/2010/09/UK-GEMINI-Encoding-Guidance-20100930-v1-01.pdf>
2. Guidance notes for the production of discovery metadata for the Marine Environmental Data and Information Network (MEDIN), version 2.3.4  
[http://www.oceannet.org/marine\\_data\\_standards/medin\\_approved\\_standards/documents/medin\\_schema\\_doc\\_2\\_3\\_4\\_30nov2010.pdf](http://www.oceannet.org/marine_data_standards/medin_approved_standards/documents/medin_schema_doc_2_3_4_30nov2010.pdf)

## Dataset metadata example

For system developers who are familiar with XML, an example of a full dataset metadata instance is shown below. The details of this ISO 19139 XML schema implementation are better described in the MEDIN and UK Gemini documents referenced in the previous section. Also, Figure 2 illustrates part of this same metadata record through the graphical GeoNetwork opensource metadata editor.

```
<?xml version="1.0" encoding="UTF-8"?>
<gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd"
  xmlns:gco="http://www.isotc211.org/2005/gco"
  xmlns:gmx="http://www.isotc211.org/2005/gmx"
  xmlns:gml="http://www.opengis.net/gml/3.2"
  xmlns:xlink="http://www.w3.org/1999/xlink">
  <gmd:fileIdentifier>
    <gco:CharacterString>64c8493d6bd95d93b7e04fb868fd568e</gco:CharacterString>
  </gmd:fileIdentifier>
  <gmd:language>
    <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/php/code_list.php" codeListValue="eng">English</gmd:LanguageCode>
  </gmd:language>
  <gmd:hierarchyLevel>
    <gmd:MD_ScopeCode
      codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/gmxCodetlists.xml#MD_ScopeCode"
      codeListValue="dataset">dataset</gmd:MD_ScopeCode>
    </gmd:hierarchyLevel>
  <gmd:contact>
    <gmd:CI_ResponsibleParty>
      <gmd:individualName>
        <gco:CharacterString>Pauline Weatherall</gco:CharacterString>
      </gmd:individualName>
      <gmd:organisationName>
        <gco:CharacterString>British Oceanographic Data Centre</gco:CharacterString>
      </gmd:organisationName>
      <gmd:contactInfo>
        <gmd:CI_Contact>
          <gmd:address>
            <gmd:CI_Address>
              <gmd:deliveryPoint>
```

```

        <gco:CharacterString>Joseph Proudman Building</gco:CharacterString>
    </gmd:deliveryPoint>
    <gmd:deliveryPoint>
        <gco:CharacterString>6 Brownlow Street</gco:CharacterString>
    </gmd:deliveryPoint>
    <gmd:city>
        <gco:CharacterString>Liverpool</gco:CharacterString>
    </gmd:city>
    <gmd:administrativeArea>
        <gco:CharacterString>Merseyside</gco:CharacterString>
    </gmd:administrativeArea>
    <gmd:postalCode>
        <gco:CharacterString>L3 5DA</gco:CharacterString>
    </gmd:postalCode>
    <gmd:country>
        <gco:CharacterString>United Kingdom</gco:CharacterString>
    </gmd:country>
    <gmd:electronicMailAddress>
        <gco:CharacterString>enquiries@bodc.ac.uk</gco:CharacterString>
    </gmd:electronicMailAddress>
</gmd:CI_Address>
</gmd:address>
<gmd:onlineResource>
    <gmd:CI_OnlineResource>
        <gmd:linkage>
            <gmd:URL>http://www.bodc.ac.uk/</gmd:URL>
        </gmd:linkage>
    </gmd:CI_OnlineResource>
</gmd:onlineResource>
</gmd:CI_Contact>
</gmd:contactInfo>
<gmd:role>
    <gmd:CI_RoleCode
codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
resources/Codelist/gmxCodelists.xml#CI_RoleCode"
codeListValue="pointOfContact">pointOfContact</gmd:CI_RoleCode>
    </gmd:role>
</gmd:CI_ResponsibleParty>
</gmd:contact>
<gmd:dateStamp>
    <gco:DateTime>2011-10-03T16:56:09</gco:DateTime>
</gmd:dateStamp>
<gmd:metadataStandardName>
    <gco:CharacterString>MEDIN Discovery Metadata Standard</gco:CharacterString>
</gmd:metadataStandardName>
<gmd:metadataStandardVersion>
    <gco:CharacterString>Version 2.3.5</gco:CharacterString>
</gmd:metadataStandardVersion>
<gmd:referenceSystemInfo>
    <gmd:MD_ReferenceSystem>
        <gmd:referenceSystemIdentifier>
            <gmd:RS_Identifier>
                <gmd:code>
                    <gco:CharacterString>urn:ogc:def:crs:EPSG::4326</gco:CharacterString>
                </gmd:code>
                <gmd:codeSpace>
                    <gco:CharacterString>OGP</gco:CharacterString>
                </gmd:codeSpace>
            </gmd:RS_Identifier>
        </gmd:referenceSystemIdentifier>
    </gmd:MD_ReferenceSystem>
</gmd:referenceSystemInfo>
<gmd:identificationInfo>

```

```

    <gmd:MD_DataIdentification>
      <gmd:citation>
        <gmd:CI_Citation>
          <gmd:title>
            <gco:CharacterString>General Bathymetric Chart of the Oceans GEBCO_08
Grid</gco:CharacterString>
          </gmd:title>
          <gmd:alternateTitle>
            <gco:CharacterString>British Oceanographic Data Centre record
1048GEBCO_08</gco:CharacterString>
          </gmd:alternateTitle>
          <gmd:date>
            <gmd:CI_Date>
              <gmd:date>
                <gco:Date>2011-10-03</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="publication">publication</gmd:CI_DateTypeCode>
                </gmd:dateType>
              </gmd:CI_Date>
            </gmd:date>
            <gmd:date>
              <gmd:CI_Date>
                <gmd:date>
                  <gco:Date>2011-01-25</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="creation">creation</gmd:CI_DateTypeCode>
                  </gmd:dateType>
                </gmd:CI_Date>
              </gmd:date>
            <gmd:date>
              <gmd:CI_Date>
                <gmd:date>
                  <gco:Date>2011-01-25</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:identifier>
              <gmd:RS_Identifier>
                <gmd:code>
                  <gco:CharacterString>EDMED4285</gco:CharacterString>
                </gmd:code>
                <gmd:codeSpace>
                  <gco:CharacterString>http://www.bodc.ac.uk/</gco:CharacterString>
                </gmd:codeSpace>
              </gmd:RS_Identifier>
            </gmd:identifier>
          </gmd:CI_Citation>
        </gmd:citation>
      <gmd:abstract>

```

```

    <gco:CharacterString>The General Bathymetric Chart of the Oceans GEBCO_08
    Grid is a global terrain model for oceans and land at 30 arc-second intervals. The
    bathymetric portion of the grid is largely based on a database of ship-track
    soundings with interpolation between soundings guided by satellite-derived gravity
    data. Existing gridded bathymetric data sets are included in some areas. The land
    portion of the grid is largely based on the US Geological Survey's SRMT30 data set,
    developed with data from the US National Aeronautics and Space Administration (NASA)
    Shuttle Radar Topographic Mission (SRTM). For the area around Antarctica, the land
    data are taken from the Geoscience Laser Altimeter System (GLAS)/ Ice, Cloud and
    Land Elevation Satellite (ICESat) laser altimetry 500m digital elevation model. The
    grid is accompanied by a Source Identifier (SID) Grid which identifies which cells in
    the GEBCO_08 Grid are based on soundings or existing grids and which have been
    interpolated. The data sets are updated as new bathymetric compilations are made
    available. Both grids are freely available to download, in netCDF form, from the web.
    Free software is available for viewing and accessing data from the grids in netCDF
    and ASCII formats. This includes the option to export the data in an ASCII format
    suitable for conversion to an ESRI raster file. The grids are also included as part
    of the GEBCO Digital Atlas DVD. </gco:CharacterString>
  </gmd:abstract>
  <gmd:pointOfContact>
    <gmd:CI_ResponsibleParty>
      <gmd:individualName>
        <gco:CharacterString>Unknown</gco:CharacterString>
      </gmd:individualName>
      <gmd:organisationName>
        <gco:CharacterString>British Oceanographic Data
Centre</gco:CharacterString>
      </gmd:organisationName>
      <gmd:positionName>
        <gco:CharacterString>Unknown</gco:CharacterString>
      </gmd:positionName>
      <gmd:contactInfo>
        <gmd:CI_Contact>
          <gmd:address>
            <gmd:CI_Address>
              <gmd:deliveryPoint>
                <gco:CharacterString>Joseph Proudman
Building</gco:CharacterString>
              </gmd:deliveryPoint>
              <gmd:deliveryPoint>
                <gco:CharacterString>6 Brownlow Street</gco:CharacterString>
              </gmd:deliveryPoint>
              <gmd:city>
                <gco:CharacterString>Liverpool</gco:CharacterString>
              </gmd:city>
              <gmd:administrativeArea>
                <gco:CharacterString>Merseyside</gco:CharacterString>
              </gmd:administrativeArea>
              <gmd:postalCode>
                <gco:CharacterString>L3 5DA</gco:CharacterString>
              </gmd:postalCode>
              <gmd:country>
                <gco:CharacterString>United Kingdom</gco:CharacterString>
              </gmd:country>
              <gmd:electronicMailAddress>
                <gco:CharacterString>enquiries@bodc.ac.uk</gco:CharacterString>
              </gmd:electronicMailAddress>
            </gmd:CI_Address>
          </gmd:address>
          <gmd:onlineResource>
            <gmd:CI_OnlineResource>
              <gmd:linkage>
                <gmd:URL>http://www.bodc.ac.uk/</gmd:URL>
              </gmd:linkage>
            </gmd:CI_OnlineResource>
          </gmd:onlineResource>
        </gmd:CI_Contact>
      </gmd:contactInfo>
    </gmd:CI_ResponsibleParty>
  </gmd:pointOfContact>

```

```

        </gmd:linkage>
        </gmd:CI_OnlineResource>
    </gmd:onlineResource>
    </gmd:CI_Contact>
</gmd:contactInfo>
<gmd:role>
    <gmd:CI_RoleCode
codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
resources/Codelist/gmxCodelists.xml#CI_RoleCode"
codeListValue="originator">originator</gmd:CI_RoleCode>
    </gmd:role>
</gmd:CI_ResponsibleParty>
</gmd:pointOfContact>
<gmd:pointOfContact>
    <gmd:CI_ResponsibleParty>
        <gmd:organisationName>
            <gco:CharacterString>British Oceanographic Data
Centre</gco:CharacterString>
        </gmd:organisationName>
        <gmd:positionName>
            <gco:CharacterString>Director</gco:CharacterString>
        </gmd:positionName>
        <gmd:contactInfo>
            <gmd:CI_Contact>
                <gmd:address>
                    <gmd:CI_Address>
                        <gmd:deliveryPoint>
                            <gco:CharacterString>Joseph Proudman
Building</gco:CharacterString>
                        </gmd:deliveryPoint>
                        <gmd:deliveryPoint>
                            <gco:CharacterString>6 Brownlow Street</gco:CharacterString>
                        </gmd:deliveryPoint>
                        <gmd:city>
                            <gco:CharacterString>Liverpool</gco:CharacterString>
                        </gmd:city>
                        <gmd:administrativeArea>
                            <gco:CharacterString>Merseyside</gco:CharacterString>
                        </gmd:administrativeArea>
                        <gmd:postalCode>
                            <gco:CharacterString>L3 5DA</gco:CharacterString>
                        </gmd:postalCode>
                        <gmd:country>
                            <gco:CharacterString>United Kingdom</gco:CharacterString>
                        </gmd:country>
                        <gmd:electronicMailAddress>
                            <gco:CharacterString>enquiries@bodc.ac.uk</gco:CharacterString>
                        </gmd:electronicMailAddress>
                    </gmd:CI_Address>
                </gmd:address>
            </gmd:CI_Contact>
            <gmd:onlineResource>
                <gmd:CI_OnlineResource>
                    <gmd:linkage>
                        <gmd:URL>http://www.bodc.ac.uk/</gmd:URL>
                    </gmd:linkage>
                </gmd:CI_OnlineResource>
            </gmd:onlineResource>
        </gmd:CI_Contact>
    </gmd:contactInfo>
</gmd:role>
    <gmd:CI_RoleCode
codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
resources/Codelist/gmxCodelists.xml#CI_RoleCode"

```

```

codeListValue="custodian">custodian</gmd:CI_RoleCode>
  </gmd:role>
</gmd:CI_ResponsibleParty>
</gmd:pointOfContact>
<gmd:resourceMaintenance>
  <gmd:MD_MaintenanceInformation>
    <gmd:maintenanceAndUpdateFrequency>
      <gmd:MD_MaintenanceFrequencyCode
codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
resources/Codelist/gmxCodelists.xml#MD_MaintenanceFrequencyCode"
codeListValue="asNeeded">asNeeded</gmd:MD_MaintenanceFrequencyCode>
      </gmd:maintenanceAndUpdateFrequency>
    </gmd:MD_MaintenanceInformation>
  </gmd:resourceMaintenance>
<gmd:graphicOverview>
  <gmd:MD_BrowseGraphic>
    <gmd:fileName>

<gco:CharacterString>http://netmar.ucc.ie/geonetwork/srv/en/resources.get?id=52&f
name=gebco_2000x1000_s.png&access=public</gco:CharacterString>
    </gmd:fileName>
    <gmd:fileDescription>
      <gco:CharacterString>thumbnail</gco:CharacterString>
    </gmd:fileDescription>
    <gmd:fileType>
      <gco:CharacterString>png</gco:CharacterString>
    </gmd:fileType>
  </gmd:MD_BrowseGraphic>
</gmd:graphicOverview>
<gmd:graphicOverview>
  <gmd:MD_BrowseGraphic>
    <gmd:fileName>

<gco:CharacterString>http://netmar.ucc.ie/geonetwork/srv/en/resources.get?id=52&f
name=gebco_2000x1000.png&access=public</gco:CharacterString>
    </gmd:fileName>
    <gmd:fileDescription>
      <gco:CharacterString>large_thumbnail</gco:CharacterString>
    </gmd:fileDescription>
    <gmd:fileType>
      <gco:CharacterString>png</gco:CharacterString>
    </gmd:fileType>
  </gmd:MD_BrowseGraphic>
</gmd:graphicOverview>
<gmd:descriptiveKeywords>
  <gmd:MD_Keywords>
    <gmd:keyword>
      <gmx:Anchor
xlink:href="http://vocab.nerc.ac.uk/collection/P01/current/MBHTGB08">Sea-floor height
(above mean sea level) {bathymetric height} in the water body by derivation from
GEBCO_08 30 arc-second global grid</gmx:Anchor>
      </gmd:keyword>
    <gmd:thesaurusName>
      <gmd:CI_Citation>
        <gmd:title>
          <gco:CharacterString>SeaDataNet PDV</gco:CharacterString>
        </gmd:title>
        <gmd:date>
          <gmd:CI_Date>
            <gmd:date>
              <gco:Date>2011-10-04</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>
<gmd:descriptiveKeywords>
        <gmd:MD_Keywords>
        <gmd:keyword>
        <gmx:Anchor
xlink:href="vocab.nerc.ac.uk/collection/L13/current/UK">unknown</gmx:Anchor>
        </gmd:keyword>
        <gmd:thesaurusName>
        <gmd:CI_Citation>
        <gmd:title>
        <gco:CharacterString>Vertical Coverages</gco:CharacterString>
        </gmd:title>
        <gmd:date>
        <gmd:CI_Date>
        <gmd:date>
        <gco>Date>2011-10-04</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>
<gmd:descriptiveKeywords>
        <gmd:MD_Keywords>
        <gmd:keyword>
        <gmx:Anchor
xlink:href="http://vocab.nerc.ac.uk/collection/N01/current/NDGO0001">Marine
Environmental Data and Information Network</gmx:Anchor>
        </gmd:keyword>
        <gmd:thesaurusName>
        <gmd:CI_Citation>
        <gmd:title>
        <gco:CharacterString>NDG Data Providers</gco:CharacterString>
        </gmd:title>
        <gmd:date>
        <gmd:CI_Date>
        <gmd:date>
        <gco>Date>2011-10-04</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>
    <gmd:descriptiveKeywords>
        <gmd:MD_Keywords>
            <gmd:keyword>
                <gmx:Anchor
xlink:href="vocab.nerc.ac.uk/collection/P22/current/10">Elevation</gmx:Anchor>
                </gmd:keyword>
                <gmd:keyword>
                    <gmx:Anchor
xlink:href="vocab.nerc.ac.uk/collection/P22/current/163">Hydrography</gmx:Anchor>
                </gmd:keyword>
            </gmd:thesaurusName>
            <gmd:CI_Citation>
                <gmd:title>
                    <gco:CharacterString>INSPIRE themes</gco:CharacterString>
                </gmd:title>
                <gmd:date>
                    <gmd:CI_Date>
                        <gmd:date>
                            <gco>Date>2011-10-04</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>
    <gmd:resourceConstraints>
        <gmd:MD_LegalConstraints>
            <gmd:useLimitation>
                <gco:CharacterString>Usage restrictions are specified in the terms of the
licence</gco:CharacterString>
            </gmd:useLimitation>
            <gmd:accessConstraints>
                <gmd:MD_RestrictionCode
codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
resources/Codelist/gmxCodelists.xml#MD_RestrictionCode"
codeListValue="otherRestrictions">otherRestrictions</gmd:MD_RestrictionCode>
                </gmd:accessConstraints>
                <gmd:otherConstraints>
                    <gco:CharacterString>Usage restrictions are specified in the terms of the
licence</gco:CharacterString>
                </gmd:otherConstraints>
            </gmd:MD_LegalConstraints>
        </gmd:resourceConstraints>
        <gmd:spatialResolution>
            <gmd:MD_Resolution>
                <gmd:distance gco:nilReason="inapplicable"/>
            </gmd:MD_Resolution>
        </gmd:spatialResolution>
        <gmd:language>
            <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-
2/php/code_list.php" codeListValue="eng">English</gmd:LanguageCode>

```

```

</gmd:language>
<gmd:topicCategory>
  <gmd:MD_TopicCategoryCode>oceans</gmd:MD_TopicCategoryCode>
</gmd:topicCategory>
<gmd:extent>
  <gmd:EX_Extent>
    <gmd:geographicElement>
      <gmd:EX_GeographicBoundingBox>
        <gmd:westBoundLongitude>
          <gco:Decimal>-180</gco:Decimal>
        </gmd:westBoundLongitude>
        <gmd:eastBoundLongitude>
          <gco:Decimal>180</gco:Decimal>
        </gmd:eastBoundLongitude>
        <gmd:southBoundLatitude>
          <gco:Decimal>-90</gco:Decimal>
        </gmd:southBoundLatitude>
        <gmd:northBoundLatitude>
          <gco:Decimal>90</gco:Decimal>
        </gmd:northBoundLatitude>
      </gmd:EX_GeographicBoundingBox>
    </gmd:geographicElement>
    <gmd:geographicElement>
      <gmd:EX_GeographicDescription>
        <!--Extent - by Identifier-->
        <gmd:geographicIdentifier>
          <gmd:MD_Identifier>
            <gmd:authority>
              <gmd:CI_Citation>
                <gmd:title>
                  <gco:CharacterString>SeaVoX water
bodies</gco:CharacterString>
                </gmd:title>
              </gmd:CI_Citation>
            </gmd:authority>
            <gmd:code>
              <gco:CharacterString>World</gco:CharacterString>
            </gmd:code>
          </gmd:MD_Identifier>
        </gmd:geographicIdentifier>
      </gmd:EX_GeographicDescription>
    </gmd:geographicElement>
    <gmd:temporalElement>
      <gmd:EX_TemporalExtent>
        <gmd:extent>
          <gml:TimePeriod gml:id="medinMEDIN01">
            <gml:beginPosition>2009</gml:beginPosition>
            <gml:endPosition>2009</gml:endPosition>
          </gml:TimePeriod>
        </gmd:extent>
      </gmd:EX_TemporalExtent>
    </gmd:temporalElement>
  </gmd:EX_Extent>
</gmd:extent>
codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
resources/Codelist/gmxCodelists.xml#CI_DateTypeCode"
codeListValue="revision">revision</gmd:CI_DateTypeCode>
</gmd:CI_DateTypeCode>
</gmd:CI_Date>
</gmd:date>
</gmd:CI_Citation>
</gmd:authority>
</gmd:code>
</gmd:code>
</gmd:MD_Identifier>
</gmd:geographicIdentifier>
</gmd:EX_GeographicDescription>
</gmd:geographicElement>
</gmd:temporalElement>
</gmd:EX_TemporalExtent>
</gmd:extent>
</gmd:EX_Extent>
</gmd:extent>

```



```

        </gmd:CI_Contact>
    </gmd:contactInfo>
    <gmd:role>
        <gmd:CI_RoleCode
codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
resources/Codelist/gmxCodelists.xml#CI_RoleCode"
codeListValue="distributor">distributor</gmd:CI_RoleCode>
        </gmd:role>
    </gmd:CI_ResponsibleParty>
</gmd:distributorContact>
</gmd:MD_Distributor>
</gmd:distributor>
<gmd:transferOptions>
    <gmd:MD_DigitalTransferOptions>
        <gmd:onLine>
            <gmd:CI_OnlineResource>
                <gmd:linkage>
                    <gmd:URL>https://www.bodc.ac.uk/data/online_delivery/gebco/</gmd:URL>
                </gmd:linkage>
                <gmd:name>
                    <gco:CharacterString>BODC online data delivery
service</gco:CharacterString>
                </gmd:name>
                <gmd:description>
                    <gco:CharacterString>Access the gridded bathymetric (GEBCO) data
delivery service. This allows grid files for a user-defined or global area to be
auto-delivered for each of the GEBCO grids.</gco:CharacterString>
                </gmd:description>
                <gmd:function>
                    <gmd:CI_OnLineFunctionCode
codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
resources/Codelist/gmxCodelists.xml#CI_OnLineFunctionCode"
codeListValue="download">download</gmd:CI_OnLineFunctionCode>
                    </gmd:function>
                </gmd:CI_OnlineResource>
            </gmd:onLine>
        </gmd:MD_DigitalTransferOptions>
    </gmd:transferOptions>
<gmd:transferOptions>
    <gmd:MD_DigitalTransferOptions>
        <!--WMS online resource-->
        <gmd:onLine>
            <gmd:CI_OnlineResource>
                <gmd:linkage>
                    <gmd:URL>http://www.gebco.net/data_and_products/gebco_web_services/web_map_service/ma
pserv?</gmd:URL>
                </gmd:linkage>
                <gmd:protocol>
                    <!--The protocol value should be taken from the GeoNetwork controlled
vocabulary-->
                    <gco:CharacterString>OGC:WMS-1.1.1-http-get-map</gco:CharacterString>
                </gmd:protocol>
                <gmd:name>
                    <!--Layer name-->
                    <gco:CharacterString>GEBCO_08_Grid</gco:CharacterString>
                </gmd:name>
                <gmd:description>
                    <!--Description of the layer delivered by the WMS service, etc.-->
                    <gco:CharacterString>Global map of height above sea level (depths
negative) taken from the GEBCO-08 30 arc second grid</gco:CharacterString>
                </gmd:description>
            </gmd:CI_OnlineResource>
        </gmd:onLine>
    </gmd:MD_DigitalTransferOptions>
</gmd:transferOptions>

```

```

        </gmd:onLine>
    </gmd:MD_DigitalTransferOptions>
</gmd:transferOptions>
</gmd:MD_Distribution>
</gmd:distributionInfo>
<gmd:dataQualityInfo>
  <gmd:DQ_DataQuality>
    <!--Scope - Required by ISO 19115 constraint-->
    <gmd:scope>
      <gmd:DQ_Scope>
        <gmd:level>
          <gmd:MD_ScopeCode
codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
resources/Codelist/gmxCodelists.xml#MD_ScopeCode"
codeListValue="dataset">dataset</gmd:MD_ScopeCode>
          </gmd:level>
        </gmd:DQ_Scope>
      </gmd:scope>
      <!--Lineage-->
      <gmd:lineage>
        <gmd:LI_Lineage>
          <gmd:statement>
            <gco:CharacterString>unknown</gco:CharacterString>
          </gmd:statement>
        </gmd:LI_Lineage>
      </gmd:lineage>
    </gmd:DQ_DataQuality>
  </gmd:dataQualityInfo>
</gmd:MD_Metadata>

```

## Acknowledgements

This cookbook was written 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.

Thanks to Roy Lowry of the British Oceanographic Data Centre for supplying baseline GEBCO metadata example which has been tuned to the NETMAR project.

This document has been reviewed by, and incorporates comments from, Pete Walker of Plymouth Marine Laboratory. Thanks to the reviewers for their help in making the document clear and readable. Further feedback on this document is welcomed, and may be provided by contacting the author whose details are below.

## Document Information

<b>Author</b>	Declan Dunne, Coastal and Marine Research Centre
<b>Contact</b>	d.dunne@ucc.ie
<b>Version</b>	1.0
<b>Date</b>	2011 December 23
<b>Revisions</b>	

**International Coastal Atlas Network Cookbook:  
Establishing a CSW metadata catalogue  
with GeoNetwork opensource**



## Table of Contents

Introduction .....	3
What is a metadata catalogue?.....	3
What is CSW? .....	3
CSW Application Profiles.....	3
CSW Servers .....	4
Installing and configuring GeoNetwork 2.6.4 .....	4
Installing GeoNetwork 2.6.4.....	5
Configuring the database for GeoNetwork 2.6.4 .....	5
Starting GeoNetwork 2.6.4 .....	7
Stopping GeoNetwork 2.6.4.....	7
GeoNetwork 2.6.4 Administration .....	7
Adding metadata records in GeoNetwork 2.6.4.....	8
Examples of CSW operations .....	9
CSW GetRecordById operation request.....	9
CSW GetRecordByMd operation request.....	10
Acknowledgements .....	11
Document Information .....	11

## Introduction

This document provides a tutorial for those who wish to understand CSW (Catalog Services for the Web) metadata catalogues. 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 some CSW servers available. The document also contains initial pointers to establishing a CSW server using GeoNetwork opensource, and examples of some 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. 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. Query languages supported by CSW are CQL (Common Query Language) and OGC FILTER.

## 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 solution for catalogue services that fits every user's needs. Therefore, there is a need for application profiles<sup>1</sup>. 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)

---

<sup>1</sup> OpenGIS Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile, Version 1.0.0, OGC 07-045

Metadata catalogues that use the ISO 19115 and ISO 19119 as its underlying information model should use the ISO Metadata Application Profile of CSW. 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 also supported by the ISO Metadata Application Profile of CSW.

## 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/>)

## 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 standard. It can run on Microsoft Windows, Linux and Mac OS X. GeoNetwork is recommended and used by the NETMAR project. The latest version can be downloaded from: <http://sourceforge.net/projects/geonetwork/>. A more detailed user manual can be found on the GeoNetwork opensource website: <http://geonetwork-opensource.org/manuals/2.6.4/users/>.

By default GeoNetwork comes embedded with the Jetty Servlet container. However, Apache Tomcat (version 5.5+) is also supported via a custom install. However, these installation instructions deal with the default Jetty installation. 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

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 can 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 example 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 configuration tasks such as 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: <http://geonetwork-opensource.org/manuals/2.6.4/users/>

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.

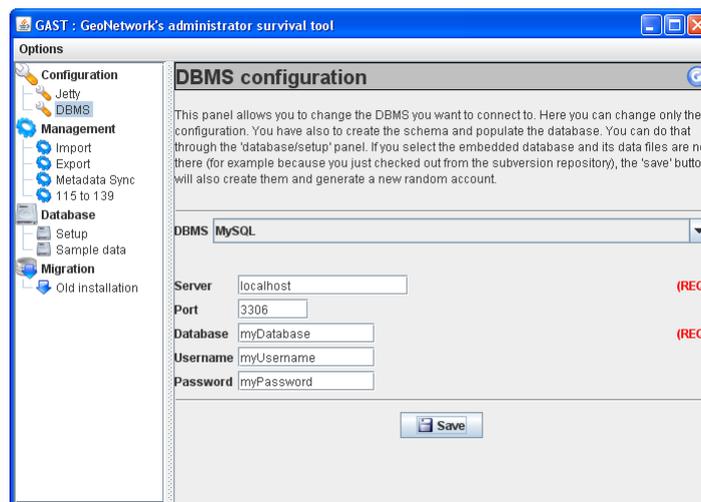


Figure 1 GeoNetwork GAST – DBMS configuration

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 internal tables if the database has previously been setup with GAST).

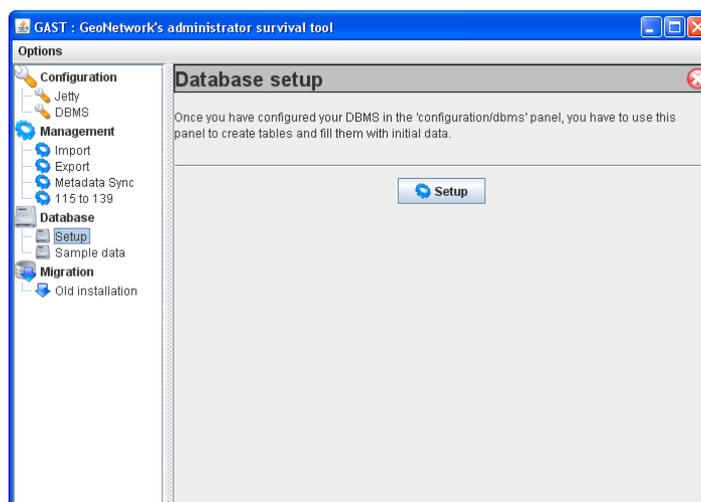


Figure 2 GeoNetwork GAST – Database setup

Note: in GeoNetwork 2.6.4 the following error message may appear in a dialog box when setting up the MySQL database 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.

## Starting GeoNetwork 2.6.4

Once the database is initialised, 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

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

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

## 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. Most of these methods are documented in the user manual: <http://geonetwork-opensource.org/manuals/2.6.4/users/>. 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 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. The method is useful for uploading existing metadata created by other GIS applications. You can also choose to upload the XML as a metadata template by selecting “Template” from the “Type” dropdown list.

### 3. **Batch loading 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.

### 4. **Metadata harvesting:**

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 process for collecting remote metadata and storing them locally for faster access.

### 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 to the GeoNetwork database. There are two main methods: Metadata Service and CSW Service. 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. This XML services option for adding metadata requires software developer skills to implement. Further details are documented in the developer manual: <http://geonetwork-opensource.org/manuals/2.6.4/developer/>.

Figure 1: GeoNetwork opensource metadata editor

## Examples of CSW operations

This section gives examples of the CSW GetRecordById and GetRecords operation requests aimed at system developers. Full details of these operations can be referenced in CSW 2.0.2 - ISO Metadata Application Profile v1.0.0<sup>2</sup>.

### 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”:

<sup>2</sup> OpenGIS Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile, Version 1.0.0, OGC 07-045

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

## CSW GetRecordById 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" (note: the XML Filter below needs to be URL encoded to work in 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&
typeName=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</Literal>
    </PropertyIsLike>
  </Filter>
```

## Acknowledgements

This cookbook was written 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.

This document has been reviewed by, and incorporates comments from, Pete Walker of Plymouth Marine Laboratory. Thanks to the reviewers for their help in making the document clear and readable. Further feedback on this document is welcomed, and may be provided by contacting the author whose details are below.

## Document Information

<b>Author</b>	Declan Dunne, Coastal and Marine Research Centre
<b>Contact</b>	d.dunne@ucc.ie
<b>Version</b>	1.0
<b>Date</b>	2011 December 23
<b>Revisions</b>	

**International Coastal Atlas Network Cookbook:  
Connecting your Atlas**



**Table of Contents**

Introduction ..... 3

CSW Requirements.....4

KOS Requirements..... 6

What is Next? ..... 7

Acknowledgements ..... 8

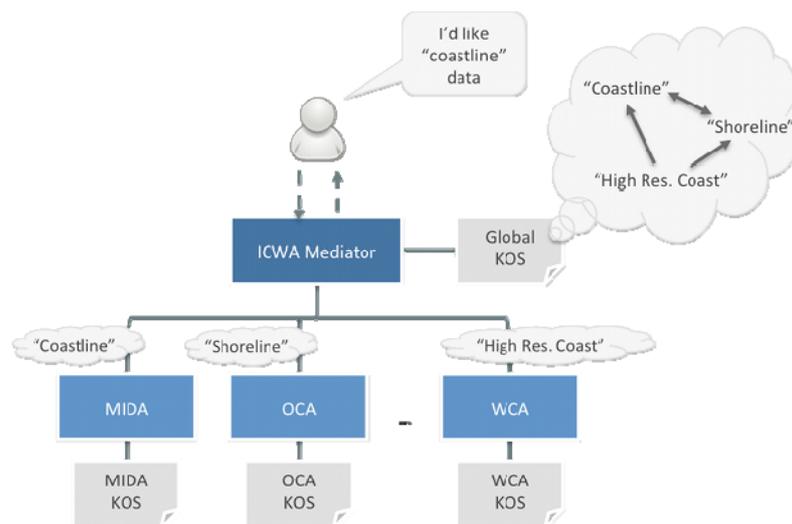
Document Information ..... 8

## Introduction

This document is a step-by-step guide explaining how you connect your atlas as a node in the International Coastal Web Atlas (ICWA). It is aimed specifically at members of the International Coastal Atlas Network community and more generally at scientists, data managers, and system developers.

ICWA is a prototype atlas mediator developed by the Technical Working Group (TWG) of the International Coastal Atlas Network (ICAN). 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. The ICAN TWG are currently investigating the extension of ICWA to support web mapping.

ICWA uses a knowledge organisation system (KOS) 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 below, 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 and this information from the global knowledge organisation system are then passed on to the local atlases that search on “coastline”, “shoreline” and “high resolution coastline.” The local atlases then return the relevant data to the ICWA and then to the user. This is an implementation of the so-called “smart-search”<sup>1</sup>.



A diagram illustrating the use for knowledge organisation systems in the ICWA

<sup>1</sup> 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.

ICWA communicates with local atlases through standard OGC CSW (Catalogue Service for the Web) interfaces. It queries the KOS through a semantic web service (SWS) interface.

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

1. Provide access to your metadata through a CSW interface supporting both Dublin core and ISO-19139 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 the above two requirements.

## CSW Requirements

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

The ICWA mediator requires a CSW 2.0.2 that is able to deliver metadata in Dublin Core and ISO-19139.

The ICWA mediator has been tested with GeoNetwork 2.4 and later versions. It should support other CSW 2.0.2 implementations but this has not been tested yet.

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

1. As character strings (`gco:CharacterString`), each containing the URI of a concept from your local KOS (if any). For 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>
      <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>
```

```

    </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). For example:

```

<gmd:descriptiveKeywords>
  <gmd:MD_Keywords>
    <gmd:keyword>
      <gco:CharacterString>Beach</gco:CharacterString>
    </gmd:keyword>
  </gmd:MD_Keywords>
  <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/gmxCodetlists.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). For 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:MD_Keywords>
  <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. For example:

```

<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 you metadata keywords in a KOS. However, you are recommended to do so as this would improve data search as explained in the introductory section above.

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 should 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;
2. Login and password for accessing metadata through your CSW (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

This cookbook was written 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.

This document has been reviewed by, and incorporates comments from Declan Dunne of the Coastal and Marine Research Centre, University College Cork. Thanks to the reviewers for their help in making the document clear and readable. Further feedback on this document is welcomed, and may be provided by contacting the author whose details are below.

## Document Information

**Author** Yassine Lassoued, Coastal and Marine Research Centre

**Contact** y.lassoued@ucc.ie

**Version** 1.0

**Date** 2011 December 22

**Revisions**