



Project No. 249024

NETMAR

Open service network for marine environmental data

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

ICT - Information and Communication Technologies Theme

D3.3 Review of available ontologies and their interfaces

Reference: D_3_3_review_existing_semantic_resources_r1

Due date of deliverable (as in Annex 1): 31 July 2010
Actual submission date: 28 September 2010

Start date of project: 1 February 2010

Duration: 3 years

British Oceanographic Data Centre (BODC), National Environment Research Council, United Kingdom

Revision 1

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



European Commission
Information Society and Media

Working document with status draft pending formal review and approval by EC Services

 	<p>NETMAR Open service network for marine environmental data Project Reference: 249024 Contract Type: Collaborative Project Start/End Date: 01/03/2010 - 31/01/2013 Duration: 36 months</p>
	<p>Coordinator: Prof. Stein Sandven Nansen Environmental and Remote Sensing Center Thormøhlensgate 47, Bergen, Norway Tel.: +47-55205800 Fax. +47 55 20 58 01 E-mail: stein.sandven@nersc.no</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)
Deputy Coordinator: Dr. Torill Hamre (torill.hamre@nersc.no)
Quality Control Manager: Mr. Lasse H. Pettersson (lasse.pettersson@nersc.no)
- British Oceanographic Data Centre (BODC), National Environment Research Council, United Kingdom
Contact: Dr. Roy Lowry (rkl@bodc.ac.uk)
- Centre de documentation de recherche et d'expérimentations sur les pollutions accidentelles des eaux (Cedre), France.
Contact: Mr. François Parthiot (Francois.Parthiot@cedre.fr)
- Coastal and Marine Resources Centre (CMRC), University College Cork, National University of Ireland, Cork, Ireland.
Contact: Mr. Declan Dunne (d.dunne@ucc.ie)
- Plymouth Marine Laboratory (PML), United Kingdom.
Contact: Mr. Steve Groom (sbg@pml.ac.uk)
- Institut français de recherche pour l'exploitation de la mer (Ifremer), France
Contact: Mr. Mickael Treguer (mickael.treguer@ifremer.fr)
- Norwegian Meteorological Institute (METNO), Norway.
Contact: Mr. Jan Ivar Pladsen (janip@met.no)

Author(s)

- Dr Roy Lowry, BODC. rkl@bodc.ac.uk
- Dr Adam Leadbetter, BODC alead@bodc.ac.uk
- Mr Olly Clements, BODC daol@bodc.ac.uk

Document approval

- Document status: Release 1
- WP leader approval: 2010-09-28
- Quality Manager approval: 2010-09-28
- Coordinator approval: 2010-09-28

Revision History

Issue	Date	Change records	Author(s)
Draft	2010-08-20	First draft of the report.	Roy Lowry Olly Clements Adam Leadbetter
Draft 2	2010-09-17	Document amended for comments from review undertaken by Roy Lowry. Examples of API payloads added & summary tables included in conclusions.	Adam Leadbetter
1	2010-09-28	Final release, incorporating comments from Advisory Board and NERSC	Adam Leadbetter

Executive Summary

The Open Service Network for Marine Environmental Data (NETMAR) aims to develop a pilot European Marine Information System (EUMIS) for searching, downloading and integrating satellite, in situ and model data from ocean and coastal areas. EUMIS will use a semantic framework coupled with ontologies for identifying and accessing distributed data, such as near-real time, forecast and historical data. This report is a review of existing semantic resources which aim to describe concepts from the environmental sciences domain in general, and the marine sciences domain in particular.

The resources reviewed and a summary of the findings are presented below:

The GEMET concept thesaurus is a rich multilingual resource with a suite of simple to use access methods that should be included in part if not in its entirety in the NETMAR semantic infrastructure. The only relevance of the INSPIRE spatial data themes, also served through the GEMET infrastructure, to NETMAR is as a source of theme Uniform Resource Identifiers, if semantic linkages to INSPIRE are required within NETMAR services.

The NERC Vocabulary Server holds a very large amount of English language content (for example a sea area gazetteer; a hierarchy of sensors and sampling devices; and a hierarchy of parameters for both discovery and usage metadata) and resource bridges that could be subsumed developed and extended through ontology extension to form the core of the NETMAR semantic resource.

The Marine Metadata Interoperability Ontology Registry and Repository is a resource that is of significant interest to NETMAR as it is the preferred repository for the US International Coastal Atlas Network community, which is one of the targeted user communities in NETMAR. This resource also provides an ideal test bed for ontology extension experiments and as a mechanism for exposure of semantic resources to NETMAR that are out of scope for ingestion into the NERC Vocabulary Service.

The Global Change Master Directory (GCMD) keyword lists, particularly the Science Keywords, have significant potential for NETMAR. However, investment in the currently public content, particularly mapping work is considered unwise as, in the past, there have been major changes to the content of the lists without consultation of the wider community or warning that changes would be implemented. Should a (planned) GCMD keyword server become available during the NETMAR project then this assessment may need to be revised.

The NASA SWEET ontologies represent a rich resource of knowledge of relevance to the NETMAR semantic framework, encoded in machine readable form. They were initially populated from the GCMD Science Keywords, but the content has undergone significant subsequent development. The network is designed as an upper-level ontology for Earth system science. Significant emphasis has been placed on the development of concepts that may be combined to produce additional concepts. For example, the ontologies include substances such as 'water', 'air' and 'blood' and properties such as 'temperature' rather than pre-combined concepts like 'air temperature'. This facilitates discovery searches either by substance or by property. Every effort should be made to incorporate them.

There may be ICES vocabularies of interest to NETMAR such as the sea area classification hierarchy; coastal region area codes; plankton sampling methodologies; qualifying terms vocabulary for expressing uncertainty in the identification of biots; and the list of coastal station types. For example, it may be useful for an oil spill management team to be able to communicate geographically with the fisheries community using geographic terms that the latter understands.

The USGS thesaurus has the potential to be a valuable source of content for NETMAR as it defines concepts of scientific disciplines, of scientific methodology, of scientific themes, of information within a resource (such as map or dataset), of geological time period and of institutional structures and activities. The mixture between US and UK English needs to be watched in a multi-lingual environment.

The InterRisk ontologies were developed for a NETMAR forerunner project covering two of the NETMAR example scenarios for oil spill and harmful algal bloom monitoring and forecasting. Their relevance to NETMAR is therefore beyond question, both as a source of content and as technical guidance to the requirements of the NETMAR semantic framework. Consequently, they will be further developed during NETMAR. Extension of the phytoplankton ontology to utilise the knowledge encoded in the World Register of Marine Species is a particularly attractive idea.

There is little subject area overlap between NETMAR and the GeoSciML vocabularies unless the International Coastal Atlas Network scenario incorporates coastal geology layers. However, the programmatic methods used to access the semantic resource are amongst the most comprehensive available and should be considered during the design phase of the access methods to the NETMAR semantic resource.

WoRMS is a comprehensive marine biological information resource that totally covers the NETMAR information requirements in this domain.

NETMAR has a requirement to be able to document service chain interfaces. Some of the required semantic information is contained within Quantities, Units, Dimensions and Types ontology. In particular, concepts such as dimensionality and its relationship to measured phenomena and units of measure are covered, which provides the basis for a mechanism to ensure that a pressure is not delivered to a service when a temperature is expected. However, it is only part of the answer to this issue. Both nitrate concentration and chlorophyll concentration can be expressed in terms of the same dimensionality, but are obviously significantly different.

There are two potential semantic resources for the NETMAR project from within the USGS Information Bank: the Field Activities Catalog and the Marine Realms Information Bank (MRIB). The Field Activities Catalog suffers because the granularity of the vocabularies is not well defined. The MRIB vocabularies present a much better possibility for integration into the NETMAR semantic resource, but they would require integration into an ontology server in order to be served in a semantic web standard format and be easily into the project's resource.

Contents

EXECUTIVE SUMMARY	II
1 INTRODUCTION.....	1
1.1 BACKGROUND.....	1
1.2 OBJECTIVE OF THIS REPORT	1
1.3 TERMINOLOGY	1
1.4 ORGANISATION OF THIS REPORT	1
2 GENERAL MULTILINGUAL ENVIRONMENTAL THESAURUS (GEMET).....	2
2.1 CONTENTS	2
2.1.1 <i>Concepts</i>	2
2.1.2 <i>INSPIRE Themes</i>	3
2.2 GEMET GOVERNANCE	3
2.3 GEMET ACCESS AND METHODS.....	3
2.4 GEMET CONCLUSIONS	5
3 NERC VOCABULARY SERVER	6
3.1 CONTENTS AT BRITISH OCEANOGRAPHIC DATA CENTRE (BODC).....	6
3.2 ADDITIONAL MATERIAL AT THE BRITISH GEOLOGICAL SURVEY (BGS)	8
3.3 NERC VOCABULARY SERVER GOVERNANCE	8
3.4 NERC VOCABULARY SERVER ACCESS AND METHODS.....	9
3.5 NERC VOCABULARY SERVER CONCLUSIONS.....	11
4 MARINE METADATA INTEROPERABILITY (MMI) ONTOLOGY REPOSITORY & REGISTRY (ORR).....	12
4.1 MMI ORR CONTENTS	12
4.2 MMI ORR GOVERNANCE.....	13
4.3 MMI ORR ACCESS AND METHODS	13
4.4 MMI ORR CONCLUSIONS	14
5 GLOBAL CHANGE MASTER DIRECTORY (GCMD)	15
5.1 GCMD CONTENTS	15
5.2 GCMD GOVERNANCE	15
5.3 GCMD ACCESS AND METHODS.....	16
5.4 GCMD CONCLUSIONS	16
6 NASA SEMANTIC WEB FOR EARTH AND ENVIRONMENTAL TERMINOLOGY (SWEET)	17
6.1 NASA SWEET ONTOLOGY CONTENTS.....	17
6.2 SWEET GOVERNANCE.....	18
6.3 SWEET ACCESS AND METHODS	18
6.4 SWEET CONCLUSIONS	18
7 INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEAS (ICES).....	19
7.1 ICES CONTENTS	19
7.2 ICES GOVERNANCE.....	19
7.3 ICES ACCESS AND METHODS	19
7.4 ICES CONCLUSIONS	19
8 UNITED STATES GEOLOGICAL SURVEY (USGS) THESAURUS	20
8.1 USGS THESAURUS CONTENTS	20
8.2 USGS THESAURUS GOVERNANCE.....	20
8.3 USGS THESAURUS ACCESS AND METHODS	20
8.4 USGS THESAURUS CONCLUSIONS	22
9 INTERRISK.....	23

9.1	INTERRISK CONTENTS	23
9.2	INTERRISK GOVERNANCE	23
9.3	INTERRISK ACCESS AND METHODS	23
9.4	INTERRISK CONCLUSIONS	24
10	GEOSCIENCE MARKUP LANGUAGE (GEOSCIML) VOCABULARIES	25
10.1	GEOSCIML VOCABULARIES CONTENT	25
10.2	GEOSCIML VOCABULARIES GOVERNANCE	25
10.3	GEOSCIML VOCABULARIES ACCESS AND METHODS	25
10.4	GEOSCIML VOCABULARIES CONCLUSIONS	28
11	WORLD REGISTER OF MARINE SPECIES (WORMS) APHIA RECORDS	29
11.1	WoRMS CONTENT	29
11.2	WoRMS GOVERNANCE	29
11.3	WoRMS APHIA RECORDS ACCESS AND METHODS	29
11.4	WoRMS CONCLUSIONS	30
12	QUANTITIES, UNITS, DIMENSIONS AND TYPES (QUDT)	31
12.1	QUDT CONTENTS	31
12.2	QUDT GOVERNANCE	31
12.3	QUDT ACCESS	31
12.4	QUDT CONCLUSIONS	31
13	USGS INFORMATION BANK	32
13.1	FIELD ACTIVITIES CATALOG CONTENTS	32
13.2	MARINE REALMS INFORMATION BANK CONTENTS	32
13.3	USGS INFORMATION BANK GOVERNANCE	33
13.4	USGS INFORMATION BANK ACCESS	33
13.5	USGS INFORMATION BANK CONCLUSIONS	33
14	CONCLUSIONS	34
15	REFERENCES	37
APPENDICES	38	
APPENDIX A.	LIST OF ABBREVIATIONS	38
APPENDIX B.	COMPLETE CATALOGUE OF VOCABULARY LISTS SERVED FROM BODC THROUGH THE NERC VOCABULARY SERVER	40
APPENDIX C.	EXAMPLE MMI ORR LIST RETURN	50
APPENDIX D.	EXAMPLE USGS THESAURUS XML PAYLOAD	53
APPENDIX E.	EXAMPLE WORMS APHIA RECORDS	55

1 Introduction

1.1 Background

The Open Service Network for Marine Environmental Data (NETMAR) (<http://www.netmar-project.eu/>) aims to develop a pilot European Marine Information System (EUMIS) for searching, downloading and integrating satellite, in situ and model data from ocean and coastal areas. It will be a user-configurable system offering flexible service discovery, access and chaining facilities using Open Geospatial Consortium (OGC), Open-source Project for a Network Data Access Protocol (OPeNDAP) and World Wide Web Consortium (W3C) standards. It will use a semantic framework coupled with ontologies for identifying and accessing distributed data, such as near-real time, forecast and historical data. EUMIS will also enable further processing of such data to generate composite products and statistics suitable for decision-making in diverse marine application domains. Figure 1-1 illustrates how observations, derived parameters and predictions are retrieved from a distributed service network through standard protocols, and delivered through the EUMIS portal using ontologies and semantic frameworks to select suitable products and where new products can be generated dynamically using chained processing services.

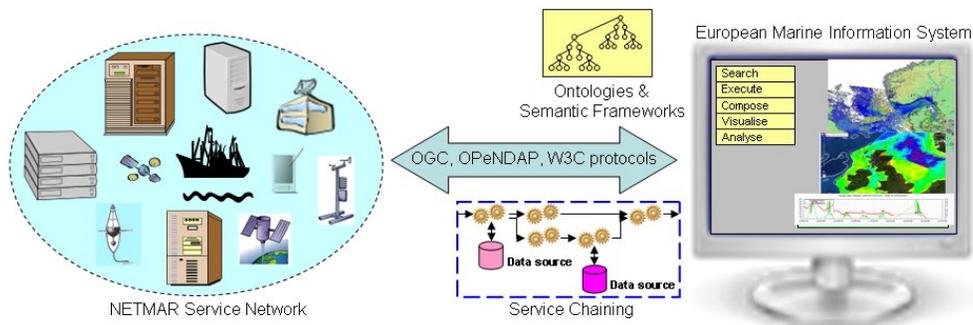


Figure 1-1 The NETMAR Service Network.

1.2 Objective of this report

The objective of this report is to provide a description of the content of existing semantic resources relevant to the NETMAR project. Also under discussion will be the governance models of those semantic resources and the ways in which a computer programmer can access the individual semantic resources and make use of the concepts that they define.

1.3 Terminology

The phrase “semantic resource” is used to describe a published list of concepts from a given domain which may be accessed through the World Wide Web.

The term “ontology” is widely used in this report. In this context, an ontology is the formal representation of a body of knowledge through the declaration of concepts from a given domain and defining the relationships between those concepts. As such ontologies form a distinct subset of “semantic resources” and an ontology can be used to describe and infer knowledge about a given domain.

“Ontology extension” or “ontology mapping” are the terms used to describe the mapping of concepts represented in one ontology to those concepts represented in a second ontology.

1.4 Organisation of this report

This report is broken down into chapters, with each chapter reviewing the content, governance and access methods of a given existing semantic resource which describes an aspect of the environmental sciences domain. The table of contents shows the semantic resource described in each chapter. A final chapter offers the report’s conclusions as to which semantic resources should be integrated through ontology extension into the NETMAR semantic resource.

2 GEneral Multilingual Environmental Thesaurus (GEMET)

2.1 Contents

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

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

2.1.1 Concepts

The GEMET concept thesaurus has the following list of themes as its top level, which is accessible at http://www.eionet.europa.eu/gemet/index_html?langcode=en in English (replace 'en' by 'fr' for French, 'no' for Norwegian or 'ga' for Gaelic):

administration	legislation
agriculture	materials
air	military aspects
animal husbandry	natural areas, landscape, ecosystems
biology	natural dynamics
building	noise, vibrations
chemistry	physics
climate	pollution
disasters, accidents, risk	radiations
economics	research
energy	resources
environmental policy	social aspects, population
fishery	soil
food, drinking water	space
forestry	tourism
general	trade, services
geography	transport
human health	urban environment, urban stress
industry	waste
information	water

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

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

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

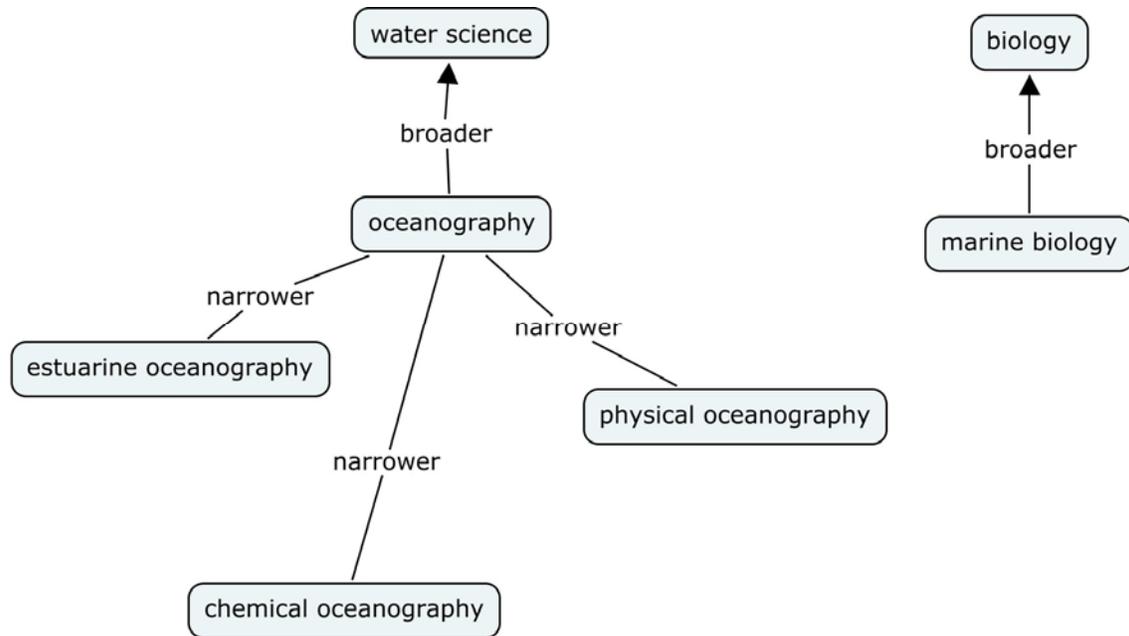


Figure 2-1 An example of the incomplete mapping within GEMET. Note the how the "marine biology" concept is divorced from the "oceanography" concept.

In spite of imperfections such as this, the thesaurus carries a very large number of concepts of relevance to NETMAR in 29 languages with significant number of semantic linkages.

2.1.2 INSPIRE Themes

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

2.2 GEMET Governance

The GEMET concept thesaurus falls under the content governance of EIONET, but as it is the merger of 8 published semantic resources (listed at <http://www.eionet.europa.eu/gemet/about>) the content is considered static.

The INSPIRE themes fall under the content governance of the INSPIRE Drafting Teams. They form the backbone of the INSPIRE development strategy and consequently change is considered unlikely.

2.3 GEMET Access and Methods

At the time of writing, the Application Programming Interface (API) access to the GEMET thesaurus is being revised, and must therefore be considered 'at risk'. The GEMET Thesaurus is available as both HyperText Markup Language (HTML) and Resource Description Framework (RDF) files from the GEMET website [GE10a]. At the time of writing, the GEMET thesaurus can be accessed as three thesauri covering concepts, super groups and INSPIRE themes.

However, the API specification is under development, and this can be publicly accessed [GE10b]. The GEMET API is based on the 'concept'. GEMET defines a 'concept' as a Uniform Resource Identifier (URI) which gives the definition of an environmental term. The payload from the API is a JavaScript Object Notation (JSON) document. The following methods are described within the API specification:

- getTopmostConcepts (also fetchThemes, fetchGroups)

- This method provides an entry point into the thesaurus for concept browsing by returning the concepts at the top level of the thesaurus; that is those concepts with the broadest descriptions.
- http://www.eionet.europa.eu/gemet/getTopmostConcepts?thesaurus_uri=http://www.eionet.europa.eu/gemet/group/&language=en
- The results of the `getTopmostConcepts` method are duplicated by the `fetchThemes` and `fetchGroups` methods.
- `getAllConceptRelatives`
 - For a given concept, this method returns a list of all the directly related concepts.
 - http://www.eionet.europa.eu/gemet/getAllConceptRelatives?concept_uri=http://www.eionet.europa.eu/gemet/group/234
- `getRelatedConcepts`
 - Where `getAllConceptRelatives` returns all the related concepts to a given term, `getRelatedConcepts` only returns a those concepts which are related by a specified link to the given term.
 - http://www.eionet.europa.eu/gemet/getRelatedConcepts?concept_uri=http://www.eionet.europa.eu/gemet/concept/913&relation_uri=http://www.w3.org/2004/02/skos/core%23broader&language=fr
- `getConcept`
 - This method returns the full details of the given concept in the language requested by the method call.
 - http://www.eionet.europa.eu/gemet/getConcept?concept_uri=http://www.eionet.europa.eu/gemet/concept/95&language=en
- `getAllTranslationsForConcept`
 - GEMET is a multilingual thesaurus, and this method supplies all of the available translations of the concept.
 - http://www.eionet.europa.eu/gemet/getAllTranslationsForConcept?concept_uri=http://www.eionet.europa.eu/gemet/concept/10126&property_uri=http://www.w3.org/2004/02/skos/core%23prefLabel
- `getConceptsMatchingKeyword`
 - This method is one of the two concept searching methods provided for GEMET. In this method, a Structured Query Language (SQL)-like syntax is used to query one of the GEMET thesauri for matches to a concept keyword. The language is specified in the method call both to define the language the keyword is in and the language to return the search results in.
 - http://www.eionet.europa.eu/gemet/getConceptsMatchingKeyword?keyword=air&search_mode=0&thesaurus_uri=http://www.eionet.europa.eu/gemet/concept/&language=en
- `getConceptsMatchingRegexByThesaurus`

- The second search method available in the GEMET API, this method returns the results of searching a given thesaurus using a regular expression. The language is specified in the method call both to define the language the regular expression is in and the language to return the search results in.
- [http://www.eionet.europa.eu/gemet/getConceptsMatchingRegexByThesaurus ? regex=air&thesaurus_uri=http://www.eionet.europa.eu/gemet/concept/&language=en](http://www.eionet.europa.eu/gemet/getConceptsMatchingRegexByThesaurus?regex=air&thesaurus_uri=http://www.eionet.europa.eu/gemet/concept/&language=en)
- **getAvailableLangs**
 - This method returns the languages a given concept's preferred label is available in. In GEMET, a concept must have a preferred label in a given language before any other property of the concept can be given in that language.
 - [http://www.eionet.europa.eu/gemet/getAvailableLanguages ?concept_uri=http://inspire.jrc.it/theme/9](http://www.eionet.europa.eu/gemet/getAvailableLanguages?concept_uri=http://inspire.jrc.it/theme/9)
- **getAvailableThesauri**
 - This method returns each of the thesauri URIs available to the GEMET API.
 - <http://www.eionet.europa.eu/gemet/getAvailableThesauri>

An example of the payload from a call to the GEMET API (in this case getConcept example call from above) is given below:

```
{
  'preferredLabel':
    {
      'string': 'administration',
      'language': 'en'
    },
  'definition':
    {
      'string': 'The management or direction of the
affairs of a public or private office, business
or organization.\r\n(Source: RHW)',
      'language': 'en'
    },
  'uri':
    'http://www.eionet.europa.eu/gemet/concept/95',
  'thesaurus':
    'http://www.eionet.europa.eu/gemet/concept/'
}
```

2.4 GEMET Conclusions

The GEMET concept thesaurus is a rich multilingual resource with a suite of simple to use access methods that should be included in part if not in its entirety in the NETMAR semantic infrastructure.

The only relevance of the INSPIRE themes to NETMAR is as a source of theme URIs, if semantic linkages to INSPIRE are required for NETMAR services.

3 NERC Vocabulary Server

3.1 Contents at British Oceanographic Data Centre (BODC)

The NERC Vocabulary Server (NVS) is a collection of controlled vocabularies that cover many subjects and facets of the oceanographic domain. At the time of writing there were some 135 publicly accessible lists. Those deemed of potential interest to NETMAR are given in the table below. The full table is presented in Appendix B. All together there were a total of 104788 concepts in the 135 lists, and approximately 45% of these are in the list subset deemed to be of interest to NETMAR.

Table 3-1 A summary of the NVS lists of potential interest within NETMAR.

IDENTIFIER	NAME	DEFINITION	GOVERNANCE
C191	SeaVoX salt and fresh water body gazetteer	Terms specified by the SeaVoX vocabulary governance to describe coherent regions of the hydrosphere. Includes land masses enclosing freshwater bodies.	SeaVoX
C320	International Standards Organisation countries	ISO country codes from ISO3166-1 list taken from www.iso.org on 22/08/2007.	ISO
C321	International Standards Organisation deprecated country codes	Deprecated ISO country codes from the ISO3166-3 list.	ISO
C381	Ports Gazetteer	Geographic locations from which a cruise may begin or end	SeaDataNet
C382	Ports Gazetteer Deprecated Entries	Entries that have been defined in the C381 cruise start and end point gazetteer, but have been deprecated.	SeaDataNet
L051	SeaDataNet sample collector categories	Terms used to describe groups of devices that collect water, suspended matter, sediment, rock, air or biota samples.	SeaDataNet
L052	SeaDataNet sample processor categories	Terms used to describe groups of devices that are used to modify water, suspended matter, sediment, rock, air or biota samples prior to analysis.	SeaDataNet
L053	SeaDataNet sample analyser categories	Terms used to describe groups of laboratory instruments that are used to analyse water, suspended matter, sediment, rock, air or biota samples.	SeaDataNet
L054	SeaDataNet sensor and instrument package categories	Terms used to describe groups of sensors or packages that make physical, chemical, biological or geological measurements in place in the field.	SeaDataNet
L055	SeaDataNet remote sensor categories	Terms used to describe groups of devices that take images of the environment or make measurements in the field without making contact with the measured entity.	SeaDataNet
L056	SeaDataNet positioning system categories	Terms used to describe groups of devices that determine the spatial location of a platform	SeaDataNet
L061	SeaVoX Platform Classes	Grouping terms used for vehicles, objects, structures or organisms capable of bearing instruments or tools for the collection of physical, chemical, geological or biological samples and data.	SeaVoX
L062	SeaVoX Platform Class Categories	Classification terms for the grouping terms contained in the SeaVoX Platform Classes vocabulary	SeaVoX
L131	SeaVoX Vertical Co-ordinate Coverages	Terms used to describe data coverage over the vertical (z) co-ordinate.	SeaVoX

L211	SeaDataNet device category types	High-level classification of devices based on function and mode of deployment.	SeaDataNet
L221	SeaVoX Device Catalogue	Terms for distinct sampling or measuring devices that may be identified in the real world in terms of manufacturer and model number.	SeaVoX
P011	BODC Parameter Usage Vocabulary	Terms built using the BODC parameter semantic model designed to describe individual measured phenomena. May be used to mark up sets of data such as a NetCDF array or spreadsheet column.	BODC
P021	SeaDataNet Parameter Discovery Vocabulary	Terms describing fine-grained related groups of measurement phenomena designed to be used in dataset discovery interfaces.	SeaDataNet
P022	SeaDataNet Parameter Discovery Vocabulary deprecates	Deprecated terms describing fine-grained related groups of measured phenomena designed to be used in dataset discovery interfaces.	SeaDataNet
P031	SeaDataNet Agreed Parameter Groups	Terms agreed within the EU SeaDataNet community to describe coarse-grained groupings of related measurement phenomena.	SeaDataNet
P032	SeaDataNet Agreed Parameter Group deprecates	Deprecated terms agreed within the EU SeaDataNet community to describe coarse-grained groupings of related measurement phenomena.	SeaDataNet
P051	International Standards Organisation ISO19115 Topic Categories	Terms defined by ISO describing data themes. Used to populate the mandatory ISO19115 topic category field.	ISO
P061	BODC data storage units	Terms used by BODC to describe the measurement units for data held in its repositories.	BODC
P071	Climate and Forecast Standard Names	Terms used for definitive but not necessarily comprehensive descriptions of measured phenomena in the CF conventions (a content standard for data stored in NetCDF).	CF Standard Names Committee
P072	Climate and Forecast deprecated Standard Names	Terms used for definitive descriptions of measured phenomena in the CF conventions (NetCDF content standard) that have been replaced by other terms.	CF Standard Names Committee
P081	SeaDataNet Parameter Disciplines	Terms used to classify SeaDataNet Agreed Parameter Groups to provide topic/theme level terms in a hierarchical parameter discovery interface	SeaDataNet
P211	Marisaurus thesaurus terms	Base terms (all hierarchical levels) interlinked by a relationship network to build the Marisaurus thesaurus.	BODC
P231	MEDIN Parameter Discipline Keywords	Coarse grained grouping terms designed by MEDIN to provide the top level for a parameter discovery thesaurus	MEDIN Standards Group

The list entries are linked together by a set of relationships that approximate to the Simple Knowledge Organisation System (SKOS) model. Some of the linkages are between members of a single list. Others are linkages between members of different lists. Both types of linkages are currently implemented as mappings, which deviates from the SKOS standard. This and other issues such as the semantic paucity of the relationships and the inability to map to URIs held in other servers will be addressed as part of the NETMAR infrastructure development. The relationship set currently used is exactMatch, broadMatch, narrowMatch and minorMatch. The minorMatch relationship is used for any semantic relationship that isn't an exact fit to one of the other three. Some remapping of this category will be required.

The mappings of interest to NETMAR, including some mappings between concepts within the same vocabularies, are:

C191 mapped to C191

This is a self-contained water body gazetteer based on International Hydrographic Bureau (IHB) (1953) Sea Areas, but significantly extended by SeaDataNet and MarineXML Vocabulary Content Governance Group (SeaVoX) implemented as a hierarchical thesaurus. The top node is the term 'world'.

L051 mapped to L051 L051 mapped to L211 L052 mapped to L211 L053 mapped to L211
 L054 mapped to L054 L054 mapped to L211 L054 mapped to L221 L055 mapped to L056
 L055 mapped to L211 L056 mapped to L055 L056 mapped to L211

This group of mappings forms a hierarchical thesaurus (top level terms in L211) classification of sensors and sample collection devices from many domains of marine interest.

L061 mapped to L062

This mapping forms a simple classification hierarchy (L062 is the upper level) for platforms (vehicles, objects, structures or organisms capable of bearing instruments or tools for the collection of physical, chemical, geological or biological samples and data).

P011 mapped to P011 P011 mapped to P021 P021 mapped to P021 P021 mapped to P022
 P021 mapped to P031 P022 mapped to P021 P032 mapped to P031 P031 mapped to P081
 P031 mapped to P231 P022 mapped to P031

This group of mappings forms a 4-level hierarchical thesaurus of oceanographic parameters. There are alternative top level lists (P081 preferred by SeaDataNet and P231 preferred by the UK Marine Environmental Data Network (MEDIN)), beneath which are P031/P032 (P032 contains deprecated terms), P021/P022 and P011.

P021 mapped to P041 P022 mapped to P041 P021 mapped to P051 P022 mapped to P051
 P021 mapped to P071 P022 mapped to P071 P021 mapped to P072 P061 mapped to P071
 P061 mapped to P072 P021 mapped to P220 P022 mapped to P220

This group of mappings provides the basis for semantic interoperability for parameter descriptions at the discovery level. The core vocabulary is the SeaDataNet Parameter Discovery Vocabulary (P021), which is mapped to Global Change Master Directory (GCMD) Science Keywords (as of March 2007), Climate and Forecast (CF) Standard Names, ISO Topic Categories and INSPIRE themes. There is also a BODC-maintained units and dimensionality vocabulary (P061) linked through the CF Standard Names.

P211 mapped to P211

This mapping is a self-contained marine environmental thesaurus (12371 concepts) known as 'Marisaurus' developed in the Netherlands in the 1980s and donated to BODC for utilisation and further development. The top node is the term 'environment'.

3.2 Additional Material at the British Geological Survey (BGS)

There is a planned expansion of the NVS to incorporate 248 controlled vocabularies served by the BGS Vocabulary Server¹ as a single harmonised resource delivering SKOS Reference documents.

Many of these vocabularies are internal BGS code lists or specific to the geological domain, such as rock type or fold geometry descriptive terms. However, there is a 6000-term geoscience thesaurus, a vocabulary of measurement precision terms and a world gazetteer that may be of interest to NETMAR.

3.3 NERC Vocabulary Server Governance

The NVS is based on the ISO19135 governance model with designated bodies responsible for the management of individual bodies. These are included in the table of lists served. The majority of the

¹ <http://www.bgs.ac.uk/data/services/vocabulary/1.0/home.html>

content falls either directly under BODC or under active bodies in which BODC has significant influence (SeaDataNet, SeaVoX and MEDIN Standards Group). This may prove beneficial should vocabulary modifications be required for NETMAR.

The BGS content is largely under BGS governance, although a small number of the published lists, such as ISO country codes, have external governance.

3.4 NERC Vocabulary Server Access and Methods

The NERC vocabularies are backed up by several web service methods. They can be accessed either by a SOAP interface or a HyperText Transfer Protocol (HTTP)-Plain Old XML (HTTP-POX) interface (HTTP GET request through a web browser). The WSDL is located at http://vocab.ndg.nerc.ac.uk/1.1/VocabServerAPI_dl.wsdl which can provide client access to all of the functions. The HTTP-POX interface uses the same methods, but constructs the query as a URL.

Table 3-2 NVS API methods. Inputs marked with * are optional

Method name	Inputs	Outputs
getList	String recordKey URI dateTime earliestRecord*	Returns XML containing any matching terms, or all terms from a list if listKey given as input
getMap	String SubjectList URI Integer Predicate String objectList URI* String inference (true or false)	Returns an XML document containing the mappings between items of the subject list and the object list or all mappings for the subject list if no object list is supplied
getRelatedRecordByCriteria	String subjectList* String subjectField* String subjectText Boolean caseSensitivity Boolean partialMatch Integer predicate String objectList* String inference(true or false)	Returns an XML document containing subject terms matching the given criteria and any object terms matching the given predicate
getRelatedRecordByTerm	String subjectTerm Integer predicate String objectList* String inferences(true or false)	Returns an XML document containing all of the matched terms, as specified by the predicate, for the given subjectTerm. Will only match to objectList if supplied
searchVocab	String listKey* String searchTerm String termType*	Returns an XML document containing terms which match the given searchTerm, limited to the termType if supplied
verifyTerm	String listKey String entryTerm String termType*	Returns an true or false depending on whether the entry term is found within the list given. Can be specialised to a type of term, i.e. long or short
whatLists	String categoryKey*	Returns an XML document containing information on all list or limited to a particular category if supplied
whatListsCat	NONE	Returns XML document containing the terms used for defining list categories, for use with the whatLists method

The following URL returns the concept "O005" from the list P031:
<http://vocab.ndg.nerc.ac.uk/term/P031/17/O005>

as the following SKOS payload:

```
<?xml version="1.0" ?>
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:skos="http://www.w3.org/2004/02/skos/core#"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
    <skos:Concept rdf:about='http://vocab.ndg.nerc.ac.uk/term/P031/17/O005'>
      <skos:externalID>
        SDN:P031:17:O005
      </skos:externalID>
      <skos:prefLabel>
        Fluxes
      </skos:prefLabel>
      <skos:altLabel>
      </skos:altLabel>
      <skos:definition>
        This group contains parameter codes that relate to the
        rate of transfer of material across real or virtual
        interfaces. The current (May 2003) BODC Parameter
        Dictionary includes two types of flux: Vertical
        fluxes of mass, chemical species and biological
        entities through the water column (sediment trap
        fluxes). Fluxes of chemical species between the water
        column and the underlying sediment, including sediment
        oxygen demand. However, other categories, such as the
        estuary discharge fluxes, are envisaged in the future.
        Note that this APG has been given 'priority'
        status to enable modellers to easily ascertain the
        extent of rate variable availability. Flux parameters
        are included in this APG in preference to any other APG
        that might seem appropriate.
      </skos:definition>
      <dc:date>
        2010-09-08T13:40:18.300+0000
      </dc:date>
      <skos:minorMatch
        rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P231/1/MD006" />
      <skos:minorMatch
        rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P231/1/MD003" />
      <skos:minorMatch
        rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P231/1/MD005" />
      <skos:broadMatch
        rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P081/3/DS06" />
      <skos:narrowMatch
        rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/A01X" />
      <skos:narrowMatch
        rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/ATDP" />
      <skos:narrowMatch
        rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/ATEM" />
      <skos:narrowMatch
        rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/AXFX" />
      <skos:narrowMatch
        rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/CBFX" />
      <skos:narrowMatch
        rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/CFTM" />
      <skos:narrowMatch
        rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/FDMS" />
      <skos:narrowMatch
        rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/FPFX" />
      <skos:narrowMatch
        rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/HFLX" />
      <skos:narrowMatch
```

```
        rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/LIFX" />
    <skos:narrowMatch
      rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/MLFX" />
    <skos:narrowMatch
      rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/MSFX" />
    <skos:narrowMatch
      rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/OPFX" />
    <skos:narrowMatch
      rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/PCFX" />
    <skos:narrowMatch
      rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/PNFX" />
    <skos:narrowMatch
      rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/RVDS" />
    <skos:narrowMatch
      rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/SAMO" />
    <skos:narrowMatch
      rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/SCDO" />
    <skos:narrowMatch
      rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/SICO" />
    <skos:narrowMatch
      rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/TCFX" />
    <skos:narrowMatch
      rdf:resource="http://vocab.ndg.nerc.ac.uk/term/P021/54/AAFX" />
  </skos:Concept>
</rdf:RDF>
```

3.5 NERC Vocabulary Server Conclusions

The NVS holds a very large amount of English language content and resource bridges that could be subsumed developed and extended through ontology extension to form the core of the NETMAR semantic resource.

4 Marine Metadata Interoperability (MMI) Ontology Repository & Registry (ORR)

4.1 MMI ORR Contents

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

Note that MMI-ORR is a registry as well as a repository. For example, there is an entry for the INSPIRE glossary with the URL <http://inspire-registry.jrc.ec.europa.eu/registers/GLOSSARY/> that is served from the INSPIRE registry at the EC Joint Research Centre in Italy.

At the time of review (2010-09-09) the newly released beta version of the framework contained 155 resources, of which the following were considered of possible relevance to NETMAR.

- BIOMAR Classification of Marine Biotopes (<http://marinemetadata.org/2005/08/biomar>)
- CF Standard Names (<http://mmisw.org/ont/cf/parameter>). Parameter name terms used by the Climate and Forecast community that are being used to label a large number of datasets, particularly data encoded in NetCDF. These are also served through the NERC Vocabulary Server.
- INSPIRE Glossary (<http://inspire-registry.jrc.ec.europa.eu/registers/GLOSSARY/>). The authoritative source of terms to be used to tag INSPIRE-compliant data and metadata.
- OCA-ICAN Matting (demo) (<http://mmisw.org/ont/oca/ocaican>). A demonstration (i.e. sparsely populated with <10 terms) mapping between the Oregon Coastal Atlas and the International Coastal Atlas Network (ICAN) global term list.
- OCA Thesaurus (<http://mmisw.org/ont/oca/thesaurus>). The terms from the Oregon Coastal Atlas that was used in the demonstration mapping are presented in a thesaurus structure.
- ICAN thesaurus (demo) (<http://mmisw.org/ont/ican/thesaurus>). The global term list used to provide the overarching ontology for the ICAN pilot demonstrator mappings.
- Coastal Marine Ecological Classification Standard: A Habitat Classification Framework (<http://mmisw.org/ont/noaa/cmecs>)
- OGC Def Resource (as of 2009-05-19) (<http://mmisw.org/ont/ogc/def>). A snapshot of OGC terms of general interest to the semantic expression of spatial data presented as URNs plus descriptions.
- A Marine Instrument and Parameter Ontology (http://mmisw.org/ont/compass/instruments_v3). This is the primary semantic deliverable from the JISC-funded COastal Marine Perception Application Profile for Scientific Scholarship project (COMPASS) project that ran from December 2007 until June 2009. It provides a framework holding the orthogonal facets of what was measured and how it was measured for part of the marine domain.
- Oregon Coastal Atlas Draft Erosion Ontology (<http://mmisw.org/orr/#http://www.coastalatlant.net/ont/20080701/theme/theme.owl>). A well populated collection of terms relevant to coastal spatial datasets.
- Coastal Erosion Topics (http://ww.coastalatlant.net/ont/coastal_erosion.owl). Classes representing coastal erosion topics used by the ICAN community.

- IOOS Vocabulary Version 1 (http://mmisw.org/ont/ioos/identifier_parameter_program). A list of terminology used to describe backbone programmes and core variables in the US Integrated Ocean Observing System (IOOS) project.
- MarineOrganism Vocabulary (<http://mmisw.org/ont/mmi/MarineOrganism>). A list of taxonomic names used by the Monterey Bay Aquarium Research Institute (MBARI), each of which has a taxonomy represented as a comma-separated string. There are no references to international taxonomic standards such as WoRMS or ITIS and no indication of the provenance of the taxonomy.

4.2 MMI ORR Governance

MMI-ORR is a semantic framework into which external users can upload and maintain resources. The governance of each resource is therefore at the behest of the poster. Each resource is labelled with extensive metadata, which includes fields that allow the governance to be identified providing that they have been populated.

4.3 MMI ORR Access and Methods

The MMI ORR supports programmatic access to its ontologies and provides a Java package for including in software [MMI09].

Full lists can be addressed returning an Web Ontology Language (OWL) eXtensible Markup Language (XML) payload, from these full lists the individual terms can be addressed returning a list of RDF predicates and objects [W3C04] which describe the term. Search is enabled using both keywords and a SPARQL Protocol and RDF Query Language (SPARQL) endpoint. Keyword search returns an HTML list of RDF triples in the form subject-object-predicate. The SPARQL interface returns HTML by default, but this can be overridden to return a comma separated value file by adding `form=csv` to the Uniform Resource Locator (URL).

An example list URL for the MMI ORR returning an OWL/XML payload is:

- <http://mmisw.org/ont?form=rdf&uri=http://mmisw.org/ont/calcofi/bottleqc>
- The return from this call is shown in appendix C.

An example term URL for the MMI ORR returning a list of RDF predicate-object pairs is:

- http://mmisw.org/orr/#http://mmisw.org/ont/calcofi/bottleqc/T_qual
- This example call returns the following:

Term: http://mmisw.org/ont/calcofi/bottleqc/T_qual

Predicate	Object
http://mmisw.org/ont/calcofi/bottleqc/ Type	Integer
http://www.w3.org/2000/01/rdf-schema#label	T_qual
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://mmisw.org/ont/calcofi/bottleqc/Bottle_Qual
http://mmisw.org/ont/calcofi/bottleqc/Name	T_qual
http://mmisw.org/ont/calcofi/bottleqc/Description	Temperature quality
http://www.w3.org/1999/02/	http://mmisw.org/ont/calcofi/

An example URL for a SPARQL query of the MMI ORR is:

- Query:

- Select the creator of each MMI ORR list:
- PREFIX dc: <http://purl.org/dc/elements/1.1/> SELECT ?s ?creator WHERE { ?s dc:creator ?creator. }
- URL
 - `http://mmisw.org/ont/?sparql=PREFIX+dc%3A+%3Chttp%3A%2F%2Fpurl.org%2Fdc%2Felements%2F1.1%2F%3E%0D%0ASELECT++%3Fs++%3Fcreator+%0D%0AWHERE+{+%3Fs+dc%3Acreator+%3Fcreator.+}%0D%0A`
- Example payload (first 15 responses shown).

s	creator
http://mmisw.org/ont/mmi/devontssn	Carlos Rueda
http://mmisw.org/ont/mmitest/qualityFlag	Carlos Rueda
http://mmisw.org/ont/mmitest/qualityflag_map	Carlos Rueda
http://mmisw.org/ont/mmi/resourcetype	Carlos Rueda
http://mmisw.org/ont/mmi/systemtype	Carlos Rueda
http://mmisw.org/ont/oca/ocaican	Carlos Rueda
http://mmisw.org/ont/mmi/om	Carlos Rueda
http://mmisw.org/ont/oca/thesaurus	Carlos Rueda
http://mmisw.org/ont/drdc/parameter	Anthony W. Isenor
http://mmisw.org/ont/univmemphis/sensor	David Russomano, PhD, Univ. of Memphis
http://www.coastalatlant.net/ont/20080701/theme/theme.owl	Tanya Haddad
http://mmisw.org/ont/gcoos/parameter	Matt Howard
http://mmisw.org/ont/mbarissds/eqtorientation	John Graybeal, Monterey Bay Aquarium Research Institute
http://mmisw.org/ont/mmitest/qualityflag_map	carlos rueda
http://mmisw.org/ont/mbari/parameter	John Graybeal

4.4 MMI ORR Conclusions

This is a resource that is of significant interest to NETMAR as it is the preferred repository for the US ICAN community. It also provides an ideal test bed for ontology extension experiments and as a mechanism for exposure of semantic resources to NETMAR that are out of scope for ingestion into the NERC Vocabulary Service.

5 Global Change Master Directory (GCMD)

5.1 GCMD Contents

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

Table 5-1 The controlled vocabularies of the GCMD

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

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

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

The Science Keywords are an extensive list of parameter terms aimed at data discovery (i.e. metadata keyword population) over a wide range of science domains. This can be clearly seen from the 14 topics that form the top level of the keyword hierarchy.

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

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

5.2 GCMD Governance

The governance of the GCMD keyword lists is internal to the organisation and has been somewhat anarchic with terms deleted resulting in user system breakages and unannounced dramatic semantic changes such as the replacement of the Science Keyword topic 'Hydrosphere' by 'Terrestrial Hydrosphere' in May 2007.

GCMD only came to realise that their vocabularies had developed into a de-facto international standard when the way changes were implemented attracted severe criticism, particularly at a Joint World Meteorological Organization (WMO) - Intergovernmental Oceanographic Commission (IOC) Technical Commission on Oceanography and Marine Meteorology (JCOMM) International Standards Development meeting. The reaction was to freeze the published version of the keywords with the following disclaimer on the web site:

“We are planning to release the new vocabularies to the users with the release of [Master Directory] MD 10. We did not include in our estimates the depth and breadth of the impact the changes and additions on those who currently maintain their metadata using the GCMD keywords outside of the directory itself. Therefore, it appeared to be prudent to build a well-structured keyword server to handle the mappings from the current version of the keywords to the new and extended sets. The keyword server is targeted to save valuable time in translations for everyone involved and is being built now.”

It is clear that the keyword lists are continuing to evolve internally and that the plan is to publish them as a reliably governed standard supported by semantic technology at some stage in the future.

5.3 GCMD Access and Methods

The GCMD keyword vocabularies are published through the NASA Goddard Space Flight Centre web pages as Portable Document Format (PDF) documents [NASA07]. There is no API access available to the keywords and they are not available as XML.

5.4 GCMD Conclusions

The GCMD keyword lists, particularly the Science Keywords, have significant potential for NETMAR. However, investment in the currently public content, particularly mapping work is considered unwise. Should the promised keyword server become available during NETMAR then this assessment may need to be revised.

6 NASA Semantic Web for Earth and Environmental Terminology (SWEET)

6.1 NASA SWEET Ontology Contents

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

The modules are arranged by subject as shown in the diagram below that was taken from the SWEET web site:

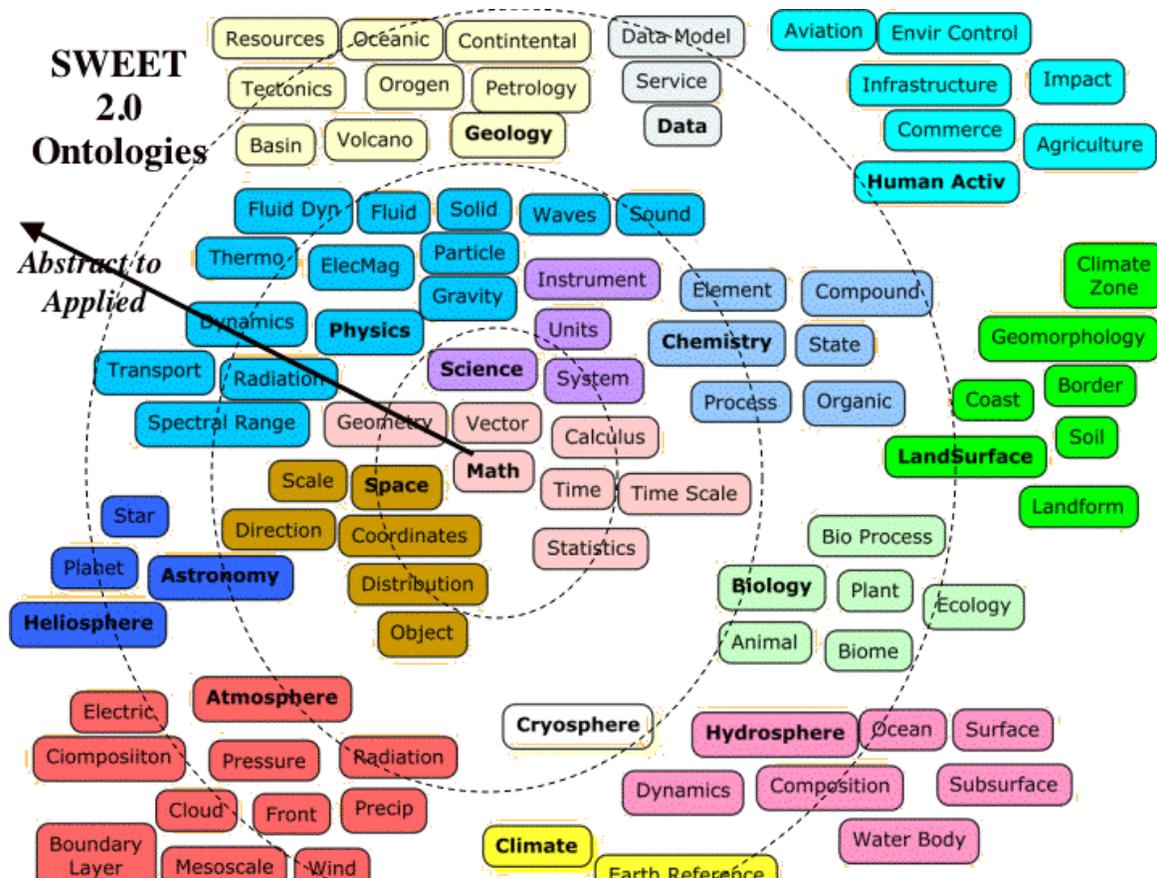


Figure 6-1 SWEET ontology modules²

The content is heavily skewed towards the atmospheric and space domains. However, there are ontology modules holding significant concept numbers in the Cryosphere, Hydrosphere, Biology, LandSurface, Chemistry and possibly HumanActivity elements of the network have obvious relevance to NETMAR.

² <http://sweet.jpl.nasa.gov/ontology/>

6.2 SWEET Governance

The SWEET ontology contents are governed internally by their management group at NASA's Jet Propulsion Laboratory under guidance from the ontology user community.

6.3 SWEET Access and Methods

There is no available API to the NASA SWEET ontology; as such the only access to the semantic resource is through the published OWL files [NASA10a].

6.4 SWEET Conclusions

The NASA SWEET ontologies represent a rich resource of knowledge encoded in machine readable form of relevance to the NETMAR semantic framework. Every effort should be made to incorporate them.

7 International Council for the Exploration of the Seas (ICES)

7.1 ICES Contents

The ICES Reference Code (RECO) system serves nearly 230 different controlled vocabulary lists that cover every aspect of the physical oceanographic, chemical contaminant and fisheries data managed by the organisation. Most of these contain terms used to populate control fields in the ICES database schema or are specific to chemical contaminants (e.g. CAS numbers of pollutants) or fishing (e.g. types of commercial fishing gear). However, a small number of these are of potential use to NETMAR:

- IC_Divisions: a sea area classification covering NW Europe and the NE Atlantic biased towards fishing grounds.
- IC_Sub-areas: a coarser grained sea area classification that forms a hierarchical layer above IC_Divisions.
- IC_Sub-divisions: a fine grained sea area classification that forms a hierarchical layer below IC_Divisions.
- JMP Area Codes: a list of terms for European coastal regions.
- PDMET: a list of plankton sampling methodologies.
- SFLAG: a list of qualifying terms that may be used to express the uncertainty in identification of biological taxa.
- STTYPE: a list of coastal station types.

7.2 ICES Governance

Governance of the RECO lists is internal to ICES with data centre staff making decisions in consultation with groups of experts in oceanography, fisheries, chemical monitoring and data management.

7.3 ICES Access and Methods

At the time of writing, the ICES RECO code lists were not accessible from the main ICES Data Centre page [ICES10a]. The code lists can still be accessed through the ICES Reference Code Utility [ICES10b]. This application allows the individual lists to be returned and searched, but only returns an HTML table of contents or results. Without the main website being available, no further analysis of access to these lists can be undertaken.

ICES report that currently the RECO Database includes internal codes which they do not want to expose to outside world. ICES are planning to rework the RECO database late in 2010, so in 2011 there will be access to all public RECO codes through a web service.

7.4 ICES Conclusions

There may be vocabularies of interest to NETMAR. For example, it may be useful for an oil spill management team to be able to communicate geographically with the fisheries community using geographic terms that the latter understands.

8 United States Geological Survey (USGS) Thesaurus

8.1 USGS Thesaurus Contents

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

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

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

8.2 USGS Thesaurus Governance

There is no explicit statement concerning governance, but there is a feedback e-mail on the web site indicating internal governance driven by user feedback.

8.3 USGS Thesaurus Access and Methods

The USGS thesauri are accessible through a series of web pages, with details of related terms displayed. This provides a useful easy to use human interface. However for the uses of NETMAR, programmatic access is preferred. This is so that services and processing tools can access and consume the thesauri. As the USGS thesauri are hosted using third-party software, access methods are provided external to the USGS website at vocabularyserver.com. This allows access to several useful methods.

All calls to the service use the following URL:

<http://www.vocabularyserver.com/usgs/services.php?task=fetchRelated&arg=41>

Where `task` is the method's name and the `arg` is the input value. An example of the XML payload for this URL is shown in Appendix D.

Table 8-1 USGS Thesaurus API methods

Method name	Inputs	Outputs
fetchRecord	Integer term ID	Returns an XML document with related terms and their ID's. a relationship value is also returned (couldn't find what they mean though)
fetchAlt	Integer term ID	Returns an XML document with terms that are alternative or non-preferred name for the input. A relationship value is also returned (should be 1 I think)
fetchDown	Integer term ID	Returns an XML document containing a tree of increasingly specific terms
fetchUp	Integer term ID	Returns an XML document containing a hierarchical structure for one ID
fetchTerm	Integer term ID	Returns an XML document containing the preferred label for the term
fetchNotes	Integer term ID	Returns an XML document with the preferred label and a description, i.e. a more complete term record
fetchDirectTerms	Integer term ID	Returns an XML document containing alternative, related and direct hierarchical terms for one term id
fetchTopTerms	NONE	Returns an XML document containing the top level terms
fetchTerms	Comma separated integer term id's (i.e. arg=41,63,96)	Returns an XML document with simple term data for the requested terms
fetchRelatedTerms	Comma separated integer term id's (i.e. arg=41,63,96)	Returns an XML document with related term data for the requested terms
Letter	A single letter	Returns an XML document with all terms starting with the specified letter
Search	String searchterm	Returns an XML document with matching terms including their location in a hierarchy where available
fetchSimilar	String searchterm	Returns an XML document with semantically similar terms(i.e. may actually be no relation just have a similar name)
fetchVocabularyData	None	Returns metadata about the vocabulary

8.4 USGS Thesaurus Conclusions

The thesaurus has the potential to be a valuable source of content for NETMAR. The mixture between US and UK English needs to be watched in a multi-lingual environment.

9 InterRisk

9.1 *InterRisk Contents*

In the 200812 branch there are two data ontologies (OWL files), an ontology designed for oil-spills and an ontology designed for harmful algal blooms, are served off the Coastal and Marine Resources Centre (CMRC), University College Cork, web site. The oil-spill ontology contains classifications of crude oil, condensate and refined petroleum product types together with rules specifying the relationships between these oil types.

The phytoplankton ontology primarily contains a formalised description of the semantic rules for a biological taxonomic classification, together with classification trees for selected harmful marine taxonomic groups down to Phylum or Class level that contain a mixture of scientific and common names. Accompanying this information is a classification of phytotoxins, human illnesses resulting from these and a small group of sampling methodologies.

In the same branch there are four keyword ontologies implemented as SKOS files, providing taxonomies for four of the five ISO19115 keyword types: disciplines, themes, temporal and strata. The disciplines taxonomy is a snapshot of the SeaDataNet top level parameter hierarchy³ linked to the concept 'discipline'. The themes taxonomy is a deeply hierarchical thesaurus of terms relevant to oils spills and harmful algal blooms, including term definitions and in many cases synonyms implemented as alternative names. The stratum taxonomy links two layers (sea surface and sea bed) to the concept of stratum. The temporal taxonomy links the concept 'month of year' to the names of the twelve months.

A later branch (200905) was developed in the final year of InterRisk based on the European Space Agency Open Access Ontology. This contains an a products ontology documenting datasets and services, a mappings ontology linking these to themes and an ontology defining the annotation terms used in the other ontologies.

9.2 *InterRisk Governance*

The governance of these ontologies has passed from the InterRisk project to the ontology governance group set up for the NETMAR project. This will oversee their development until the end of NETMAR (February 2013) after which they will be transferred to a persistent body to assure their long-term stewardship.

9.3 *InterRisk Access and Methods*

The InterRisk vocabularies are available as plain OWL files from the CMRC website⁴. They are also available through a drilldown interface on the same website. There is no documented API to this drilldown interface, but it is possible to analyse the calls being made from the webpage and to identify the URLs of the individual concepts.

The URLs take the form:

http://cmrc.ucc.ie/ontologies/interrisk/doc/namespace_concept.html

In which the "namespace" may be one of:

- amo Arc Marine Ontology
- habs Harmful Algal Blooms Ontology
- oilspills Oil Spills Ontology
- owl Web Ontology Language
- protege Protégé ontology editor
- rdf Resource Description Framework
- rdfs Resource Description Framework Schema

³ <http://vocab.ndg.nerc.ac.uk/list/P081/3>

⁴ <http://cmrc.ucc.ie/ontologies/interrisk/doc/index.html>

- xsd XML Schema

And the “concept” may be the name of a resource, a class, an object property, a datatype property, an annotation property or an individual. The resources, classes, object properties, datatype properties, annotation properties and individuals are available through indices at the following URLs

- Resources <http://cmrc.ucc.ie/ontologies/interrisk/doc/IndexAllResources.html>
- Classes <http://cmrc.ucc.ie/ontologies/interrisk/doc/IndexClasses.html>
- Object Properties <http://cmrc.ucc.ie/ontologies/interrisk/doc/IndexObjectProperties.html>
- Datatype Properties <http://cmrc.ucc.ie/ontologies/interrisk/doc/IndexDatatypeProperties.html>
- Annotation Properties <http://cmrc.ucc.ie/ontologies/interrisk/doc/IndexAnnotationProperties.html>
- Individuals <http://cmrc.ucc.ie/ontologies/interrisk/doc/IndexIndividuals.html>

Once both the namespace and the “concept” have been identified, a URL can be constructed to access that “concept”. For example, the individual “FuelOil” from the namespace “oilspills” is accessed thus:

http://cmrc.ucc.ie/ontologies/interrisk/doc/oilspills_fueloil.html

and the payload is

```
Individual
(
  oilspills:FuelOil
  annotation
  (
    rdfs:comment "[GPA]: A liquid petroleum product
having a flash point above 37,8 degrees C used, e.g.,
in industrial furnaces, domestic heaters, and ships"
    http://www.w3.org/2001/XMLSchema#string
  )
  type
  (
    oilspills:RefinedPetroleumProduct
  )
)
```

9.4 InterRisk Conclusions

These ontologies were developed for a NETMAR forerunner project covering two of the NETMAR example scenarios. Their relevance to NETMAR is therefore beyond question, both as a source of content and as technical guidance to the requirements of the NETMAR semantic framework. Consequently, they will be further developed during NETMAR. Extension of the phytoplankton ontology to utilise the knowledge encoded in the World Register of Marine Species is a particularly attractive idea.

10 Geoscience Markup Language (GeoSciML) vocabularies

10.1 GeoSciML Vocabularies Content

The GeoSciML vocabulary server presents three collections of vocabularies and thesauri designated as versions 'CGI200811', 'CGI201001' and '1GE201001'. CGI200811 contains 27 objects that are a mixture of flat lists and hierarchical thesauri. The content as might be expected is tightly focussed on the description of geological units, including rock type classifications, descriptions of the geological object geometry and geological processes. In addition there are two general lists describing value qualifiers and semantic relationships between vocabulary qualifiers. Version CGI2010001 is a development of these resources with a noticeable development of thesauri.

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

10.2 GeoSciML Vocabularies Governance

At present the procedure for adding to the GeoSciML vocabularies is to propose any additions via the chair of the GeoSciML concept definitions group. Depending on the nature of the request it will either be immediately acted upon or put out to the wider concept definitions group for comment. Some vocabularies are more complete than others and the GeoSciML governance body are aware of the need more concepts to be added. Any proposed additional concept must carry a definition.

10.3 GeoSciML Vocabularies Access and Methods

The vocabularies for GeoSciML are available through two clients: [GSML10a] and [GSML10b]. [GSML10a] provides a useful graphical user interface to the available methods on the server, while [GSML10b] provides web form access to a range of access methods. Although programmatic access to these methods is not documented, it is available and the following gives example calls to the access methods.

- AvailableVocabularies
 - This method provides access to a list of the vocabularies available from the server.
 - <http://srvgeosciml.brgm.fr/eXist2010/brgm/xql/concept.xql?collection=/db/SKOS/CGI201001&request=AvailableVocabularies&ts=1279723510115>
- GetVocabulary
 - This method returns a complete listing of an individual vocabulary as a SKOS document.
 - The following URL returns the simple lithology vocabulary.
 - <http://srvgeosciml.brgm.fr/eXist2010/brgm/xql/concept.xql?request=GetVocabulary&urlVocabulary=SimpleLithology200811.XML>
- GetConceptByTerm
 - This method searches a specified vocabulary for the presence of a keyword in the skos:prefLabel.
 - The following URL returns the SKOS concepts from the simple lithology vocabulary where the skos:prefLabel (that is the human readable concept name) contains the term "rock".

- <http://srvgeosciml.brgm.fr/eXist2010/brgm/xql/concept.xql?request=GetConceptByTerm&term=rock&urnVocabulary=SimpleLithology200811.XML>
- GetConceptByRegex
 - This method returns concept from the entire semantic resource in which the definition matches a given regular expression in an XML document.
 - The following URL returns a list of the concepts for which the definition contains the phrase “sediment”.
 - <http://srvgeosciml.brgm.fr/eXist2010/brgm/xql/concept.xql?request=GetConceptByRegex&term=sediment>
- GetConceptByProperties
 - This method searches the concepts in a given list for property and value pair matches and returns an XML document of the results.
 - The URL below returns the concepts from the “Simple Lithology” list with the property “ConsolidationDegree” equal to “consolidated”.
 - <http://srvgeosciml.brgm.fr/eXist2010/brgm/xql/concept.xql?request=GetConceptByProperties&urnVocabulary=SimpleLithology200811.xml&prop=geosciml%3AConsolidationDegree&val=consolidated>
- GetAllConceptRelations
 - http://srvgeosciml.brgm.fr/eXist2010/brgm/xql/concept.xql?collection=/db/SKOS/CGI201001&request=GetAllConceptRelations&urn=urn:cgi:classifer:CGI:EventEnvironment:201001:earth_interior_setting&ts=1279724132601
- GetProperties
 - This method returns the properties a concept may have within a given vocabulary in an XML document.
 - The following URL returns the properties a concept from the simple lithology vocabulary may have.
 - <http://srvgeosciml.brgm.fr/eXist2010/brgm/xql/concept.xql?request=GetProperties&urnVocabulary=SimpleLithology200811.XML>
- GetValues
 - This method returns the values of a given property from a given vocabulary as an XML document.
 - For example, the following URL returns the values for grain size from the simple lithology vocabulary
 - <http://srvgeosciml.brgm.fr/eXist2010/brgm/xql/concept.xql?request=GetValues&prop=geosciml%3AgrainSize&urnVocabulary=SimpleLithology200811.XML>
- GetTopmostConcepts

- This method provides top-level access to a vocabulary by returning only those concepts from the vocabulary which have no terms related to them by a skos:broader predicate.
- The following URL returns the top level concepts from the event environment vocabulary
- <http://srvgeosciml.brgm.fr/eXist2010/brgm/xql/concept.xql?collection=/db/SKOS/CGI201001&request=GetTopmostConcepts&urnVocabulary=EventEnvironment201001.XML&ts=1279723810536>
- GetConceptByURN
 - This method returns the SKOS document for an individual concept, defined by a given Uniform Resource Name
 - For the concept with the URN `urn:cgi:classifer:CGI:EventEnvironment:201001:earth_interior_setting` the following URL returns the SKOS document
 - http://srvgeosciml.brgm.fr/eXist2010/brgm/xql/concept.xql?collection=/db/SKOS/CGI201001&request=GetConceptByURN&urn=urn%3Acgi%3Aclassifer%3ACGI%3AEventEnvironment%3A201001%3Aearth_interior_setting&ts=1279724132628
 - The SKOS document returned from this URL is:

```
<repository>
  <owl:Thing
    rdf:about="urn:cgi:classifer:CGI:EventEnvironment:201001:earth_interior_s
    etting">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <dc:creator xml:lang="en">
      Concept Definition Task Group of IUGS CGI Interoperability
      Working Group (http://www.cgi-
      iugs.org/tech\_collaboration/interoperability\_working\_group.html)
    </dc:creator>
    <skos:prefLabel xml:lang="en">
      Earth interior setting
    </skos:prefLabel>
    <skos:definition xml:lang="en">
      Geologic environments within the solid Earth.
    </skos:definition>
    <dc:source xml:lang="en">
      this vocabulary
    </dc:source>
    <skos:narrower
      rdf:resource="urn:cgi:classifer:CGI:EventEnvironment:201001:contact
      _metamorphic_origin"/>
    <skos:narrower
      rdf:resource="urn:cgi:classifer:CGI:EventEnvironment:201001:crust"/
      >
    <skos:narrower
      rdf:resource="urn:cgi:classifer:CGI:EventEnvironment:201001:high_pr
      essure_low_temperature_earth_interior_setting"/>
    <skos:narrower
      rdf:resource="urn:cgi:classifer:CGI:EventEnvironment:201001:hypabys
      sal_setting"/>
```

```
<skos:narrower
  rdf:resource="urn:cgi:classifier:CGI:EventEnvironment:201001:low_pre
  ssure_high_temperature_setting"/>
<skos:narrower
  rdf:resource="urn:cgi:classifier:CGI:EventEnvironment:201001:mantle"
  />
<skos:narrower
  rdf:resource="urn:cgi:classifier:CGI:EventEnvironment:201001:regiona
  l_metamorphic_origin"/>
<skos:narrower
  rdf:resource="urn:cgi:classifier:CGI:EventEnvironment:201001:ultra_h
  igh_pressure_crustal_setting"/>
<skos:inScheme rdf:resource=""/>
<skos:topConceptOf rdf:resource=""/>
</owl:Thing>
</repository>
```

10.4 GeoSciML Vocabularies Conclusions

There is little subject area overlap between NETMAR and the GeoSciML vocabularies unless the International Coastal Atlas Network scenario incorporates coastal geology layers. However, the programmatic methods used to access the semantic resource are amongst the most comprehensive available and should be considered during the design phase of the access methods to the NETMAR semantic resource.

11 World Register of Marine Species (WoRMS) Aphia Records

11.1 WoRMS Content

WoRMS is a comprehensive on-line taxonomic resource for the marine domain. Their aim is 'to provide the most authoritative list of names of all marine species globally, ever published'. They are well on their way.

The catalogue contents at the time of writing consist of:

- 188,561 valid species, of which 147,609 checked (78%)
- 296,834 species names including synonyms
- 365,010 taxa (ranging hierarchically from infraspecies to Kingdoms)
- 12,417 images

Their final goal is to capture all of the estimated 230,000 marine species by the end of 2010, in synchrony with the completion of the Census of Marine Life programme.

The species names form the basis of a comprehensive knowledge base. The names are linked together by a full taxonomic hierarchy up to Kingdom level. Linkages are maintained between scientific names and common names (in several languages including French, German and Dutch as well as English), invalid and valid scientific names (termed synonyms) and the WoRMS entry and other taxonomic resources such as the US Integrated Taxonomic Information System.

Each entry carries extensive non-taxonomic information including the environment in which the species is found, its geographic distribution, links to references on the internet, notes on the biological characteristics of the species and images. The entry for the "edible mussel"⁵ exemplifies the richness of the information resource. A formal citation to tag any information accessed from the database (some of which has been harvested from other resources) and subsequently reused is provided.

11.2 WoRMS Governance

WoRMS has both a Steering Committee and a Board of Editors. The former includes some of the most respected names in marine biology. The latter is composed of experts specialising in particular groups of taxa who undertake the detailed content governance work reviewing both new and existing entries.

11.3 WoRMS Aphia Records access and methods

The aim of WoRMS is to provide an authoritative and comprehensive list of names of marine organisms, including information on synonymy. While highest priority goes to valid names, other names in use are included so that this register can serve as a guide to interpret taxonomic literature.

Aphia contains valid species names, synonyms and vernacular names, and extra information such as literature and biogeographic data. Besides species names, Aphia also contains the higher classification in which each scientific name is linked to its parent taxon. The classification used is a 'compromise' between established systems and recent changes. Its aim is to aid data management, rather than suggest any taxonomic or phylogenetic opinion on species relationships.

All of the WoRMS functions are accessed using the platform independent SOAP protocol. The WSDL⁶ can provide client access to all of the functions. Examples of WoRMS Aphia records are provided in Appendix E.

⁵ <http://www.marinespecies.org/aphia.php?p=taxdetails&id=140480>

⁶ <http://marinespecies.org/aphia.php?p=soap&wsdl=1>

Table 11-1 WoRMS API access methods. Inputs marked with * are optional

Method name	Inputs	Outputs
getAphiaID	String Scientific name boolean Marine_only	Returns the WoRMS ID as an integer
getAphiaRecords	String Scientific name boolean Marine_only* boolean Fuzzymatch* boolean Like*	Returns one or more matching AphiaRecord (max 50). The like Boolean, if true, allows the use of a % symbol as a wildcard similar to SQL
getAphiaNameByID	Integer AphiaID	Returns a string with the correct name for a given ID
getAphiaRecordByID	Integer AphiaID	Returns an aphia record object for the given id (see below for example return)
getAphiaRecordByTSN	Integer ITIS_TSN	Returns an aphia record object for the given id (see below for example return)
getAphiaRecordsByNames	scientificNames(custom type) Scientificnames boolean Marine_only* boolean Fuzzymatch* boolean Like*	For each given scientific name, returns one or more AphiaRecords. This allows you to match multiple names in one call. Limited to 500 names per call
getAphiaRecordsByVernacular	String vernacular	Returns one or more Aphia records for the given vernacular (species common name)
getAphiaClassificationById	Integer AphiaID	Returns the complete classification for a single taxon
getSourcesByAphiaID	Integer AphiaID	Returns one or more sources/references for a single AphiaID
getAphiaSynonymsById	Integer AphiaID	Returns all synonyms for a given Aphia ID
getAphiaVernacularsById	Integer AphiaID	Returns all vernaculars for a given AphiaID

11.4 WoRMS Conclusions

WoRMS is a comprehensive marine biological information resource that totally covers the NETMAR information requirements in this domain.

12 Quantities, Units, Dimensions and Types (QUDT)

12.1 QUDT Contents

The QUDT ontologies cover the following concepts

- Quantity: an observable property of an object, event or system that can be measured and quantified numerically. Quantities have a magnitude and a 'kind' that specifies the observable property quantified, e.g. length, force, frequency.
- Quantity Kind: the observable property attribute of a Quantity that can be measured and quantified numerically.
- Unit of Measure (UoM): a particular quantity of a given Quantity Kind that has been chosen as a scale for measuring other quantities of the same kind. The magnitude (Quantity Value) of any Quantity may be expressed as the product of a real number and a UoM.

The ontologies are distributed in a modular form comprising components that are distributed as individual OWL files:

QUDT	The QUDT Schema Ontology: 200 Classes and 61 Properties
Quantity	QUDT Measurable Quantities Individuals: 239 Individuals
Units	QUDT Units of Measure Individuals: 801 Individuals
Dimension	QUDT Physical Dimensions Individuals: 460 Individuals
NIST Constants	National Institute of Science and Technology Constants, Units & Uncertainty Individuals: 2 Classes, 1 Property, 648 Individuals
QUDT-DBPedia	Mappings of QUDT Classes and Individuals onto DBpedia Vocabulary: 262 skos:closeMatch mappings, 561 skos:exactMatch mappings
QUDT-SPIN	Functions for Units (e.g. conversion): 28 Functions

The domain scope of the ontologies is comprehensive from the perspective of a physicist, covering the dimensionality and UoM of multiple systems including SI, CGS and MKS. However, some oceanographic quantities in common use, such as salinity, are not included.

12.2 QUDT Governance

The ontologies have been built from Quantity and UoM references from standards organisations, such as the National Institute of Science and Technology Guide for the International System of Units. Content governance is therefore maintained by these organisations.

12.3 QUDT Access

There is no available API to the QUDT ontology; as such the only access to the semantic resource is through the published OWL files [QUDT10].

12.4 QUDT Conclusions

NETMAR has a requirement to be able to document service chain interfaces. Some of the required semantic information is contained within QUDT. In particular, concepts such as dimensionality and its relationship to measured phenomena and units of measure are covered, which provides the basis for a mechanism to ensure that a pressure is not delivered to a service when a temperature is expected. However, it is only part of the answer to this issue. Both nitrate concentration and chlorophyll concentration can be expressed in terms of the same dimensionality, but are obviously significantly different.

13 USGS Information Bank

The USGS Information Bank is a structured information storage scheme of databases and software that provide organized access to USGS Coastal and Marine data and metadata [USGS10a]. Two aspects of the Information Bank are of potential interest to the NETMAR project: the Field Activities Catalog and the Marine Realms Information Bank. Both of these are discussed below.

13.1 Field Activities Catalog Contents

The Field Activities Catalog contains details of field investigations undertaken by the USGS. The metadata and data can be accessed through a series of drilldowns with entry points of the platform from which the measurements were made; the year in which the field campaign took place; the geographical region(s) of interest; the scientists involved; the organizations responsible for the field measurements; and the project and scientific theme the field study belongs to.

As a result of browsing the Field Activities Catalog, it is obvious that several of these fields (namely: platform⁷; project / theme⁸; activity type; geographical area⁹; and organization¹⁰) are populated from controlled vocabularies which may be of use to the NETMAR project. Closer examination reveals that the content of most of these vocabularies leaves a lot to be desired. In particular the project / theme; activity type; and platform vocabularies do not have a well defined granularity for their concepts. As such they mix concepts which would be better split into separate lists, for example the platform vocabulary contains the concept of “airplane” and “fishing vessel” but also large numbers of individually named ships; the activity type vocabulary mixes concepts such as “cruise” with individual cruise names; and the project / theme vocabulary mixes both project titles with scientific concepts such as “coastal evolution”.

13.2 Marine Realms Information Bank Contents

The Marine Realms Information Bank (MRIB) provides access to free online scientific information about oceans, coasts, and coastal watersheds [USGS10b] through both a category search and geographical search. The categories available in the former are: author; agency; project; location; feature type; hot topics; biota; geologic time; discipline; research method; content type; and file type. Each of these twelve facets are populated from entries in controlled vocabularies available from [USGS10c]. The USGS defines these facets thus:

- Author:** The person(s) responsible for the intellectual content of the information resource (as distinct from the *custodian* of the web site providing the information resource).
- Agency:** Any organization that contributed to the information resource, by funding or conducting research, interpreting or compiling data, publishing a document, hosting a web site, and so forth. Subcategories include governments, academic institutions, museums and aquariums, other non-profit organizations, professional associations, and businesses.
- Project:** If applicable, the project name as specified in the information resource. This name may be different than the formal project title employed by the parent agency for administrative purposes.
- Location:** The named location(s) associated with the information resource. Matches are based on latitude and longitude values, allowing MRIB to assign multiple names to a given location. The **Location** facet can be used to view a map of all indexed information resources within a named geographic region.
- Feature Type:** The generic features, geographic and otherwise, with which the information resource is concerned. Included are landforms (e.g.,

⁷ <http://walrus.wr.usgs.gov/infobank/programs/html/ships2idshtml/ships.html>

⁸ <http://walrus.wr.usgs.gov/infobank/programs/html/projects2idshtml/projects.html>

⁹ <http://walrus.wr.usgs.gov/infobank/programs/gaz/arealist2html/arealist.html>

¹⁰ <http://walrus.wr.usgs.gov/infobank/programs/html/organizations2idshtml/organizations.html>

mountains and oceans), geological features (e.g., landslides and earthquake zones), biological features (e.g., coral reefs and kelp forests), administrative areas (e.g., marine sanctuaries and exclusive economic zones), and human constructions (e.g., dams and canals).

Hot Topics: Issues of concern to scientists, policy makers, and the general public. The Marine Realms Information Bank (MRIB) and the Monterey Bay Science Digital Library (a regional offshoot of MRIB) employ the generalized **Hot Topics** facet.

Biota: The common names of organisms, arranged in five kingdoms: animals, plants, fungi, protists, and bacteria. A subcategory for viruses is also included.

Geologic Time: The geologic units of time (eon, era, period, and epoch) addressed by the information resource. MRIB uses a simplification of [Grant's 2003 time scale](#).

Discipline: The traditional academic area within which an information resource falls.

Research Method: The means by which the scientific investigation was accomplished, including field observation, laboratory analysis, data processing, and modelling.

Content Type: The *intellectual form* of the information resource: text, data sets, images, and so on.

File Type: The *transmission form* of the information resource. MRIB employs the [IANA MIME](#) classification of electronic media types: application, audio, image, model, text, and video.

The MRIB gazetteer, feature type, topics, biota, geologic time, discipline, research method, content type and file type facet vocabularies are all of potential interest to the NETMAR project. These vocabularies are not simple lists, but contain a hierarchy of terms. In contrast with the vocabularies of the Field Activities Catalog, the MRIB vocabularies do not appear to suffer from mixed granularity.

13.3 USGS Information Bank Governance

There are no details provided on the governance of the controlled vocabularies for either the Field Activities Catalog or the MRIB.

13.4 USGS Information Bank Access

The controlled vocabularies of the Field Activities Catalog do not appear to be published for access in any way. The MRIB vocabularies are published via the MRIB website, but they are not published in an XML format – simply as HTML. There is no API to the MRIB facet vocabularies.

13.5 USGS Information Bank Conclusions

There are two potential semantic resources for the NETMAR project from within the USGS Information Bank: the Field Activities Catalog and the Marine Realms Information Bank (MRIB). The Field Activities Catalog suffers because the granularity of the vocabularies is not well defined.

The MRIB vocabularies present a much better possibility for integration into the NETMAR semantic resource, but they would require integration into an ontology server in order to be served in a semantic web standard format and be easily into the project's resource.

14 Conclusions

A wide range of available semantic resources relevant to the scientific domain of the NETMAR project have been reviewed in this report. A summary of the outcomes of this review process is presented below.

Table 14-1 A summary of access to the reviewed semantic resources

	Is XML returned?	Are individual lists addressed?	Are individual terms addressed?	Is search enabled?	Multilingual?
GEMET	Yes, but not from the API		Yes	Yes, keywords and regular expressions	Yes
NERC Vocabulary Server	Yes, SKOS	Yes	Yes	Yes, keywords	
MMI ORR	Yes, but not for all methods	Yes	Yes	Yes, through keywords and a SPARQL endpoint	
NASA GCMD		Yes, but there is no API			
NASA SWEET	Yes, but there is no API	Yes, but there is no API			
ICES RECO				Yes	
USGS Thesaurus	Yes	Yes	Yes	Yes, keywords	
InterRisk		Yes	Yes		
GeoSciML Vocabularies	Yes, SKOS	Yes	Yes	Yes, keywords, regular expressions and property values	
WoRMS			Yes	Yes, SQL-like search	Yes
QUDT	Yes, OWL	Yes			
USGS Information Bank		Yes			

Table 14-2 A summary of programmatic methods available for the reviewed semantic resources.

	What Lists?	Get List	Get Term	Get Related Terms	Regular Expressions search	SPARQL search	Keyword search	Verify Term
GEMET	Yes		Yes	Yes	Yes		Yes	
NERC Vocabulary Server	Yes	Yes	Yes	Yes			Yes	Yes
MMI ORR		Yes	Yes			Yes		
NASA GCMD								
NASA SWEET								
ICES RECO								
USGS Thesaurus			Yes	Yes			Yes	
InterRisk								
GeoSciML Vocabularies	Yes	Yes	Yes	Yes	Yes		Yes	
WoRMS			Yes				Yes	
QUDT								
USGS Information Bank								

The GEMET concept thesaurus is a rich multilingual resource with a suite of simple to use access methods that should be included in part if not in its entirety in the NETMAR semantic infrastructure. The only relevance of the INSPIRE themes, also served through the GEMET infrastructure, to NETMAR is as a source of theme URIs, if semantic linkages to INSPIRE are required within NETMAR services.

The NVS holds a very large amount of English language content and resource bridges that could be subsumed developed and extended through ontology extension to form the core of the NETMAR semantic resource.

The MMI ORR is a resource that is of significant interest to NETMAR as it is the preferred repository for the US ICAN community. It also has provides an ideal test bed for ontology extension experiments and as a mechanism for exposure of semantic resources to NETMAR that are out of scope for ingestion into the NERC Vocabulary Service.

The GCMD keyword lists, particularly the Science Keywords, have significant potential for NETMAR. However, investment in the currently public content, particularly mapping work is considered unwise. Should the promised keyword server become available during NETMAR then this assessment may need to be revised.

The NASA SWEET ontologies represent a rich resource of knowledge encoded in machine readable form of relevance to the NETMAR semantic framework. Every effort should be made to incorporate them.

There may be ICES vocabularies of interest to NETMAR. For example, it may be useful for an oil spill management team to be able to communicate geographically with the fisheries community using geographic terms that the latter understands.

The USGS thesaurus has the potential to be a valuable source of content for NETMAR. The mixture between US and UK English needs to be watched in a multi-lingual environment.

The InterRisk ontologies were developed for a NETMAR forerunner project covering two of the NETMAR example scenarios. Their relevance to NETMAR is therefore beyond question, both as a source of content and as technical guidance to the requirements of the NETMAR semantic framework. Consequently, they will be further developed during NETMAR. Extension of the phytoplankton ontology to utilise the knowledge encoded in the World Register of Marine Species is a particularly attractive idea.

There is little subject area overlap between NETMAR and the GeoSciML vocabularies unless the International Coastal Atlas Network scenario incorporates coastal geology layers. However, the programmatic methods used to access the semantic resource are amongst the most comprehensive available and should be considered during the design phase of the access methods to the NETMAR semantic resource.

WoRMS is a comprehensive marine biological information resource that totally covers the NETMAR information requirements in this domain.

NETMAR has a requirement to be able to document service chain interfaces. Some of the required semantic information is contained within QUDT. In particular, concepts such as dimensionality and its relationship to measured phenomena and units of measure are covered, which provides the basis for a mechanism to ensure that a pressure is not delivered to a service when a temperature is expected. However, it is only part of the answer to this issue. Both nitrate concentration and chlorophyll concentration can be expressed in terms of the same dimensionality, but are obviously significantly different.

There are two potential semantic resources for the NETMAR project from within the USGS Information Bank: the Field Activities Catalog and the Marine Realms Information Bank (MRIB). The Field Activities Catalog suffers because the granularity of the vocabularies is not well defined. The MRIB vocabularies present a much better possibility for integration into the NETMAR semantic resource, but they would require integration into an ontology server in order to be served in a semantic web standard format and be easily into the project's resource.

15 References

- [GE10a] GEMET RDF Files <http://www.eionet.europa.eu/gemet/rdf> Accessed 20th July 2010
- [GE10b] GEMET Web Service API
<https://svn.eionet.europa.eu/projects/Zope/wiki/GEMETWebServiceAPI> Accessed 19th July 2010
- [GSML10a] CGI / GeoSciML – Vocabularies <http://srvgeosciml.brgm.fr/eXist2010/brgm/client.html>
Accessed 21st July 2010
- [GSML10b] GeoSciML : search vocabularies <http://srvgeosciml.brgm.fr/eXist2010/brgm/> Accessed
21st July 2010
- [ICES10a] ICES Data Centre – Data Submissions
<http://www.ices.dk/datacentre/Submissions/index.aspx> Accessed 21st July 2010
- [ICES10b] ICES Reference and Field List Index <http://www.ices.dk/datacentre/reco/> Accessed 21st
July 2010
- [MMI09] Programmatic Access | Marine Metadata Interoperability. Published 19th November 2009
<http://marinemetadata.org/mmiorrusrman/mmiorrprg> Accessed 21st July 2010
- [NASA07] GCMD's Science Keywords and Associated Directory Keywords.
http://gcmd.gsfc.nasa.gov/Resources/valids/archives/keyword_list.html Accessed 21st
July 2010
- [NASA10a] <http://sweet.jpl.nasa.gov/2.0/>
- [QUDT10] QUDT Version 1.0.0 Ontologies & Documentation.
<http://www.qudt.org/qudt/owl/1.0.0/index.html> Accessed 31st August 2010
- [USGS10a] USGS CMG InfoBank <http://walrus.wr.usgs.gov/infobank/> Accessed 31st August 2010
- [USGS10b] Marine Realms Information Bank <http://mrib.usgs.gov/> Accessed 31st August 2010
- [USGS10c] MRIB Classification Scheme – Facets <http://mrib.usgs.gov/doc/facets.html> Accessed 31st
August 2010
- [W3C04] Resource Description Framework (RDF): Concepts and Abstract Syntax. Published 10th
February 2004 <http://www.w3.org/TR/rdf-concepts/> Accessed 21st July 2010

Appendices

Appendix A. List of abbreviations

API	Application Programming Interface
BGS	British Geological Survey
BODC	British Oceanographic Data Centre
CF	Climate and Forecast
EIONET	European Environment Information and Observation Network
EUMIS	European Marine Information System
GCMD	Global Change Master Directory
GEMET	GEneral Multilingual Environmental Thesaurus
GeoSciML	GeoScience Markup Language
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
HTTP-POX	HyperText Transfer Protocol – Plain Old XML
ICAN	International Coastal Atlas Network
ICES	International Council for the Exploration of the Seas
IHB	International Hydrographic Bureau
INSPIRE	Infrastructure for Spatial Information in Europe
IOC	Intergovernmental Oceanographic Commission
ISO	International Organization for Standardization
JCOMM	Joint WMO- IOC Technical Commission on Oceanography and Marine Meteorology
JSON	JavaScript Object Notation
MEDIN	Marine Environmental Data & Information Network, United Kingdom
MMI	Marine Metadata Interoperability project
MRIB	Marine Realms Information Bank
NASA	National Aeronautics and Space Administration
NERC	Natural Environment Research Council, United Kingdom
NETMAR	Open Service Network for Marine Environmental Data
NVS	NERC Vocabulary Server
ORR	Ontology Registry and Repository

OWL	Web Ontology Language
PDF	Portable Document Format
RDF	Resource Description Framework
RECO	ICES REference Code
SeaVoX	SeaDataNet and MarineXML Vocabulary Content Governance Group
SKOS	Simple Knowledge Organization System
SOAP	Simple Object Access Protocol
SPARQL	SPARQL Protocol and RDF Query Language
SQL	Structured Query Language
SWEET	Semantic Web for Earth and Environmental Terminology
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name
USGS	United States Geological Survey
WMO	World Meteorological Organization
WoRMS	World Register of Marine Species
WSDL	Web Service Description Language
XML	eXtensible Markup Language

Appendix B. Complete Catalogue of Vocabulary Lists Served from BODC through the NERC Vocabulary Server

IDENTIFIER	NAME	DEFINITION	GOVERNANCE
C101	BODC series feature types	Terms that describe groups of BODC series with common independent variable characteristics.	BODC
C161	International Hydrographic Bureau (1953) sea areas	Terms used for sea areas from International Hydrographic Bureau, Limits of Oceans and Seas (Special Publication No. 23), 3rd edition 1953.	IHO
C162	SeaDataNet sea area extensions	Sea area terms added to the IHB(1953) list under SeaDataNet governance until the proposed hierarchical SeaVox sea area gazetteer is available	SeaDataNet
C171	United Kingdom ship names	Names of ships registered in the UK or operated by UK organisations that are known to be engaged in the collection of oceanographic data.	ICES
C172	Non-United Kingdom ship names	Names of ships operated by organisations outside the UK that are known to be engaged in the collection of oceanographic data.	ICES
C173	Partnership for Observation of the Global Ocean ships of interest	Research vessels deemed to be of interest to POGO. 'Of interest' is defined as active ocean-going research vessels greater than 60m in length	POGO
C174	SeaDataNet Cruise Summary Report ship metadata	Ship instances (a hull operating under a given name and governance type) used in SeaDataNet CSR forms including metadata in the definition to allow reliable mapping to ship hull databases.	SeaDataNet
C180	Intergovernmental Oceanographic Commission countries	Country names published by ICES on behalf of IOC as part of their RNODC (formats) responsibility. Maintained until 2003.	ICES
C191	SeaVoX salt and fresh water body gazetteer	Terms specified by the SeaVoX vocabulary governance to describe coherent regions of the hydrosphere. Includes land masses enclosing freshwater bodies.	SeaVoX
C301	Active vocabulary content governance authorities	Bodies responsible for the intellectual control of vocabularies served by the NDG/SeaDataNet vocabulary server	BODC
C310	World Meteorological Organisation voluntary observing ship category	WMO ship type codes used to classify voluntary observing ships from http://www.wmo.int/pages/prog/www/ois/pub47/pub47-home.htm .	WMO
C320	International Standards Organisation countries	ISO country codes from ISO3166-1 list taken from www.iso.org on 22/08/2007.	ISO
C321	International Standards Organisation deprecated country codes	Deprecated ISO country codes from the ISO3166-3 list.	ISO

IDENTIFIER	NAME	DEFINITION	GOVERNANCE
C322	International Standards Organisation country code nulls	Terms used to populate fields normally populated by ISO country codes in cases where the information is unknown or populating the field is considered inappropriate	ISO
C330	NERC DataGrid data production tool category	Abbreviated terms used to specify categories of data production tools in the NERC DataGrid MOLES metadata schema.	
C342	Monitoring activity rationale	Terms describing the reasons why a monitoring activity was undertaken.	SeaDataNet
C351	European Nature Information System Level 3 Habitats	Medium-granularity (hierarchical level 3 of 5) terms describing the environment associated with an activity from a biological perspective.	EEA
C361	Monitoring activity legislative drivers	Legislative acts, agreements and treaties that have provided the impetus for monitoring activities to be undertaken.	SeaDataNet
C371	Ten-degree Marsden Squares	Labels applied to areas of ten degrees latitude by ten degrees longitude in the Marsden Square system.	SeaDataNet
C372	Five-degree Marsden Squares	Labels applied to areas of five degrees latitude by five degrees longitude in the Marsden Square system.	SeaDataNet
C373	One-degree Marsden Squares	Labels applied to areas of one degree latitude by one degree longitude in the Marsden Square system.	SeaDataNet
C381	Ports Gazetteer	Geographic locations from which a cruise may begin or end	SeaDataNet
C382	Ports Gazetteer Deprecated Entries	Entries that have been defined in the C381 cruise start and end point gazetteer, but have been deprecated.	SeaDataNet
C390	World Meteorological Organisation sea states	Codes and terms developed from the Douglas wind sea states to describe the general condition of the surface of a large water body.	WMO
C400	Bonn Agreement pollution report accuracy	Terms used to classify the accuracy and reliability of pollution reports filed under the Bonn Agreement.	BA
C410	BODC marine pollution sources	Terms developed by BODC to provide a standard classification of pollution sources to be used in UK pollution reports filed under the Bonn Agreement.	BODC
C430	BODC oilspill quantity	Terms developed by BODC to classify the magnitude of a discharge of oil into the marine environment	BODC
C601	NERC DataGrid vocabulary term relationships	Terms used in the NDG project to describe the relationship between subjects and objects in RDF triples, including vocabulary term mappings	
C610	Post town nil-value terms	Terms to be used in 'city' or 'post town' metadata fields to indicate that no value has been supplied to the field.	BODC
C611	United Kingdom post towns	Terms to be used in 'city' or 'post town' metadata fields for addresses in the United Kingdom.	BODC

IDENTIFIER	NAME	DEFINITION	GOVERNANCE
C620	County or administrative region nil-value terms	Terms to be used in 'county' metadata fields to indicate that no value has been supplied for that field.	BODC
C621	United Kingdom counties	Terms to be used in 'county' metadata fields for addresses in the UK. Covers mainland Britain, Northern Ireland, Channel Isles, Isle of Man, Isle of Wight, Orkney, Shetland Islands and Hebrides.	BODC
C624	German states	Terms to be used in 'county' metadata fields for addresses in post-reunification Germany.	BODC
C627	Belgian Provinces	Terms to be used in 'county' metadata fields for addresses in Belgium.	BODC
C628	Canadian Provinces and Territories	Terms to be used in 'county' metadata fields for addresses in Canada.	BODC
C629	Danish regions	Terms to be used in 'county' metadata fields for addresses in Denmark.	BODC
C6210	French Provincial Regions	Terms to be used in 'county' metadata fields for addresses in France.	BODC
C6211	Dutch provinces	Terms to be used in 'county' metadata fields for addresses in the Netherlands.	BODC
C6212	Irish Republic counties	Terms to be used in 'county' metadata fields for addresses in the Republic of Ireland.	BODC
C6219	United States of America states	Terms to be used in 'county' metadata fields for addresses in the United States of America	BODC
C6220	Swedish counties	Terms to be used in 'county' metadata fields for addresses in Sweden.	BODC
C751	Natural Environment Research Council laboratories and collaborative centres	Research centres and project offices supplying data to BODC that are either part of the Natural Environment Research Council or funded by them to a significant degree.	BODC
C752	United Kingdom Government funded laboratories	Research centres supplying data to BODC that are either part of the UK Government or directly funded by it to a significant degree.	BODC
C753	United Kingdom higher education institutions	Research centres supplying data to BODC that are part of the UK higher education sector.	BODC
C754	Commercial organisations	Research centres in the commercial sector supplying data to BODC. Includes companies based both in the UK and abroad.	BODC
C755	Foreign academic and government organisations	Research centres from outside the UK supplying data to BODC that are in either the academic or government sectors.	BODC
C756	Charitable organisations	A legally incorporated non-profit organisation that operates for the public benefit and has federally registered charitable status. Includes charities registered in any country.	BODC

IDENTIFIER	NAME	DEFINITION	GOVERNANCE
C771	ROSCOP physical oceanography category	Terms used to describe shipboard activities (sometimes inaccurately described as parameters) in the physical oceanography domain.	SeaDataNet
C772	ROSCOP chemical oceanography category	Terms used to describe shipboard activities (sometimes inaccurately described as parameters) in the chemical oceanography domain.	SeaDataNet
C773	ROSCOP contaminants category	Terms used to describe shipboard activities (sometimes inaccurately described as parameters) in the contaminants domain.	SeaDataNet
C774	ROSCOP biology and fisheries category	Terms used to describe shipboard activities (sometimes inaccurately described as parameters) in the biology and fisheries domain.	SeaDataNet
C775	ROSCOP meteorology category	Terms used to describe shipboard activities (sometimes inaccurately described as parameters) in the meteorology domain.	SeaDataNet
C776	ROSCOP geology and geophysics category	Terms used to describe shipboard activities (sometimes inaccurately described as parameters) in the geology and geophysics domain.	SeaDataNet
C865	SeaDataNet Contact Roles	Terms used to define the responsibilities for a contact (person or organisation) either within the SeaDataNet project or for the datasets described by SeaDataNet metadata.	SeaDataNet
C866	SeaDataNet security access control roles	Terms used by the SeaDataNet project to describe roles associated with access rights to data, metadata and project information.	SeaDataNet
C970	Vocabulary Server Mappings Index	A catalogue of the mappings between vocabularies held in the BODC Vocabulary Server.	BODC
C980	Vocabulary Server list subject categories	Terms used to classify the lists held in the Vocabulary Server (formerly known as code tables) by the subject of their contents.	BODC
E011	ESEAS active tide gauge sites	Terms used to describe stations providing sea level data into the ESEAS network	BODC
EVO1	Environment of the Organism (EnvO) ontology base terms	Terms assembled by the EnvO community to describe the environment of an organism or biological sample.	
GGS1	GEBCO 30-second grid data sources	Provenance attributes of data source files used in the production of the GEBCO 30-second grid	
L021	SeaDataNet Geospatial Feature Types	SeaDataNet profile of ISO MD_GeometricObjectTypeCodeTerms code list. Known in SEA-SEARCH as 'Library 2' or cdi_measurement_codes.	SeaDataNet
L031	SeaDataNet Measurement Periodicity Classes	Terms used to express a classification of measurement periodicity in the SeaDataNet project.	SeaDataNet

IDENTIFIER	NAME	DEFINITION	GOVERNANCE
L051	SeaDataNet sample collector categories	Terms used to describe groups of devices that collect water, suspended matter, sediment, rock, air or biota samples.	SeaDataNet
L052	SeaDataNet sample processor categories	Terms used to describe groups of devices that are used to modify water, suspended matter, sediment, rock, air or biota samples prior to analysis.	SeaDataNet
L053	SeaDataNet sample analyser categories	Terms used to describe groups of laboratory instruments that are used to analyse water, suspended matter, sediment, rock, air or biota samples.	SeaDataNet
L054	SeaDataNet sensor and instrument package categories	Terms used to describe groups of sensors or packages that make physical, chemical, biological or geological measurements in place in the field.	SeaDataNet
L055	SeaDataNet remote sensor categories	Terms used to describe groups of devices that take images of the environment or make measurements in the field without making contact with the measured entity.	SeaDataNet
L056	SeaDataNet positioning system categories	Terms used to describe groups of devices that determine the spatial location of a platform	SeaDataNet
L061	SeaVoX Platform Classes	Grouping terms used for vehicles, objects, structures or organisms capable of bearing instruments or tools for the collection of physical, chemical, geological or biological samples and data.	SEAVOX
L062	SeaVoX Platform Class Categories	Classification terms for the grouping terms contained in the SeaVoX Platform Classes vocabulary	SEAVOX
L071	SeaDataNet data access mechanisms	Terms for mechanisms by which data objects described by SeaDataNet Central Data Index (CDI) records may be obtained before SeaDataNet is fully implemented.	SeaDataNet
L081	SeaDataNet Data Access Restriction Policies	Terms used to represent and classify data access policies in operation in the SeaDataNet project	SeaDataNet
L101	SeaDataNet geographic co-ordinate reference frames	Co-ordinate reference systems used for positions (latitude/longitude or grid references) in SeaDataNet metadata. May include 3D co-ordinate systems, but not vertical elevation datums like OD Newlyn.	SeaDataNet
L111	Vertical Co-ordinate Reference System Origins	Terms used to describe the zero point for vertical co-ordinate reference systems (datums).	SeaDataNet
L121	SeaDataNet Parameter Validation Terms	Terms that are used to describe quality control procedures applied to data parameters.	SeaDataNet
L131	SeaVoX Vertical Co-ordinate Coverages	Terms used to describe data coverage over the vertical (z) co-ordinate.	SEAVOX
L141	SeaDataNet Activity Operational Stati	Terms used to describe the operational status of data collection activities within the SeaDataNet project.	SeaDataNet

IDENTIFIER	NAME	DEFINITION	GOVERNANCE
L151	SeaDataNet quality management system accreditations	Terms used to describe formally recognised standards for data acquisition and handling procedures.	SeaDataNet
L181	ROSCOP sample quantification units	Terms used as units in the quantification of what was collected or measured in a Cruise Summary Report (ROSCOP) record.	SeaDataNet
L191	SeaDataNet keyword types	Terms used in SeaDataNet metadata to describe the purpose of a keyword. An extension of the ISO19115 KeywordTypeCode codelist.	SeaDataNet
L201	SeaDataNet measurand qualifier flags	Flags used to provide additional information, usually referring to data quality, about data values in SeaDataNet data files.	SeaDataNet
L211	SeaDataNet device category types	High-level classification of devices based on function and mode of deployment.	SeaDataNet
L221	SeaVoX Device Catalogue	Terms for distinct sampling or measuring devices that may be identified in the real world in terms of manufacturer and model number.	SEAVOX
L231	SeaDataNet metadata entities	Types of metadata record in the SeaDataNet metadata catalogue repositories	SeaDataNet
L241	SeaDataNet data transport formats	Formats used for data delivery by SeaDataNet	SeaDataNet
L300	MEDATLAS Data Centres	Organisations supplying or holding data concerned with the EU MEDATLAS project.	
M010	MEDIN data format categories	Terms describing what types of data formats exist in a data set so a human can identify if the data is likely to be of use and select a tool to interrogate that data.	MSG
N010	NERC DataGrid Data Provider Groups	Groups of organisations who provide discovery metadata collections to NERC DataGrid that are held such that the collections may be searched separately.	
N021	NERC DataGrid Recognised Formats	Formats recognised by NDG, where this may allow conditional treatment. See also http://proj.badc.rl.ac.uk/ndg/browser/TI07-MOLES/trunk/Documents/NDG_MOLES_Instructions_for_Creation.doc	
N031	Climate Science Modelling Language supported feature types	Terms used to specify the feature types available for coverages in the Climate Science Modelling Language (CSML).	
N041	NERC DataGrid on-line reference types	Terms describing types of on-line references in NERC DataGrid service bindings	

IDENTIFIER	NAME	DEFINITION	GOVERNANCE
P011	BODC Parameter Usage Vocabulary	Terms built using the BODC parameter semantic model designed to describe individual measured phenomena. May be used to mark up sets of data such as a NetCDF array or spreadsheet column.	BODC
P021	SeaDataNet Parameter Discovery Vocabulary	Terms describing fine-grained related groups of measurement phenomena designed to be used in dataset discovery interfaces.	SeaDataNet
P022	SeaDataNet Parameter Discovery Vocabulary deprecates	Deprecated terms describing fine-grained related groups of measured phenomena designed to be used in dataset discovery interfaces.	SeaDataNet
P031	SeaDataNet Agreed Parameter Groups	Terms agreed within the EU SeaDataNet community to describe coarse-grained groupings of related measurement phenomena.	SeaDataNet
P032	SeaDataNet Agreed Parameter Group deprecates	Deprecated terms agreed within the EU SeaDataNet community to describe coarse-grained groupings of related measurement phenomena.	SeaDataNet
P041	Global Change Master Directory Science Keywords	Concatenated category, topic, term and variable separated by '>'. From Olsen L.M. et. al (2006) NASA/GCMD Earth Science Keywords Version 5.3.3.	GCMD
P042	Global Change Master Directory deprecated Science Keywords	Concatenated category, topic, term and variable separated by chevrons that were formerly in a published version of GCMD Earth Science Keywords.	GCMD
P051	International Standards Organisation ISO19115 Topic Categories	Terms defined by ISO describing data themes. Used to populate the mandatory ISO19115 topic category field.	ISO
P061	BODC data storage units	Terms used by BODC to describe the measurement units for data held in its repositories.	BODC
P071	Climate and Forecast Standard Names	Terms used for definitive but not necessarily comprehensive descriptions of measured phenomena in the CF conventions (a content standard for data stored in NetCDF).	CF Standard Names Committee
P072	Climate and Forecast deprecated Standard Names	Terms used for definitive descriptions of measured phenomena in the CF conventions (NetCDF content standard) that have been replaced by other terms.	CF Standard Names Committee
P081	SeaDataNet Parameter Disciplines	Terms used to classify SeaDataNet Agreed Parameter Groups to provide topic/theme level terms in a hierarchical parameter discovery interface	SeaDataNet
P091	MEDATLAS Parameter Usage Vocabulary	Terms under the content governance of SISMER used to describe measured phenomena within the MEDATLAS project.	
P092	MEDATLAS Parameter Usage Vocabulary deprecated terms	Terms formerly used to describe parameters in the MEDATLAS data format that have subsequently been withdrawn.	

IDENTIFIER	NAME	DEFINITION	GOVERNANCE
P101	Global Change Master Directory Instrument Keywords	Terms used to describe sensors, instruments and other pieces of scientific equipment in the NASA Global Change Master Directory metadatabase.	GCMD
P102	Global Change Master Directory deprecated Instrument Keywords	Terms formerly used to describe sensors, instruments and other pieces of scientific equipment in the NASA Global Change Master Directory metadatabase that have been deleted from the list by GCMD.	GCMD
P111	Global Change Master Directory Science Keyword categories	Category level terms from the Science Keywords from Olsen L.M. et. al (2006) NASA/GCMD Earth Science Keywords Version 5.3.3.	GCMD
P121	Global Change Master Directory Science Keyword topics	Topic level terms from the Science Keywords from Olsen L.M. et. al (2006) NASA/GCMD Earth Science Keywords Version 5.3.3.	GCMD
P131	Global Change Master Directory Science Keyword terms	Term level terms from the Science Keywords from Olsen L.M. et. al (2006) NASA/GCMD Earth Science Keywords Version 5.3.3.	GCMD
P132	Global Change Master Directory deprecated Science Keyword terms	Term level terms formerly used in the Science Keywords from Olsen et. al (2006) GCMD Earth Science Keywords that have been deleted by GCMD.	GCMD
P141	Global Change Master Directory Science Keyword variables	Variable level terms from the Science Keywords from Olsen L.M. et. al (2006) NASA/GCMD Earth Science Keywords Version 5.3.3.	GCMD
P142	Global Change Master Directory Science Keyword deprecated variables	Variable level terms that have previously existed in published versions of the GCMD Science Keywords.	GCMD
P151	Climate and Forecast Cell Methods	Terms used to describe measured phenomena derivation algorithms in the CF conventions (a content standard for data stored in NetCDF), primarily targeted at statistical parameters.	CF Standard Names Committee
P181	Global Change Master Directory Chronostratigraphic Units	Chronostratigraphic unit terms from Olsen L.M. et. al (2006) NASA/GCMD Earth Science Keywords Version 5.3.3.	GCMD
P191	Global Change Master Directory platforms	Terms used to describe sensor-bearing platforms from Olsen L.M. et. al (2006) NASA/GCMD Earth Science Keywords Version 5.3.3.	GCMD
P201	Global Change Master Directory URL content types	Terms used in NASA's GCMD to describe the function of related URLs embedded in DIF metadata records. Also used for NERC DataGrid service bindings.	GCMD
P211	Marisaurus thesaurus terms	Base terms (all hierarchical levels) interlinked by a relationship network to build the Marisaurus thesaurus.	BODC
P220	INSPIRE themes	Groupings of spatial data according to Annex I, II and III of the INSPIRE Directive [DS-D2.5]	European Commission

IDENTIFIER	NAME	DEFINITION	GOVERNANCE
P231	MEDIN Parameter Discipline Keywords	Coarse grained grouping terms designed by MEDIN to provide the top level for a parameter discovery thesaurus	MEDIN Standards Group
S010	BODC parameter semantic model component names	Terms used to describe the information entities that are combined in the BODC semantic model to produce a phenomenon description.	BODC
S020	BODC parameter semantic model relationships between what theme and where theme	Terms used to describe the relationship between the 'where theme' and the 'what theme' in the BODC semantic model for measured phenomena descriptions	BODC
S030	BODC parameter semantic model sample preparation entity descriptions	Controlled vocabulary defining the terms that may be used for a sample preparation entity (part of the how theme) in the BODC parameter semantic model.	BODC
S040	BODC parameter semantic model analytical method entity descriptions	Controlled vocabulary defining the terms that may be used for an analytical method entity (part of the how theme) in the BODC parameter semantic model.	BODC
S060	BODC parameter semantic model parameter entity names	Controlled vocabulary defining the terms that may be used for a parameter entity (part of the "what" theme) in the BODC parameter semantic model.	BODC
S070	BODC parameter semantic model parameter statistic	Controlled vocabulary defining the terms that may be used for a parameter statistic (part of the "what" theme) in the BODC parameter semantic model.	BODC
S080	BODC parameter semantic model parameter entity subgroup names	Controlled vocabulary defining the terms that may be used for a parameter entity subgroup (part of the what theme qualifying a generic parameter entity) in the BODC parameter semantic model.	BODC
S090	BODC parameter semantic model biological entity size terms	Controlled vocabulary defining the terms that may be used for biological entity size (part of the what theme) in the BODC parameter semantic model.	BODC
S100	BODC parameter semantic model biological entity gender terms	Controlled vocabulary defining the terms that may be used for biological entity gender (part of the what theme) in the BODC parameter semantic model.	BODC
S110	BODC parameter semantic model biological entity development stage terms	Controlled vocabulary defining the terms that may be used for biological entity development stage (part of the "what" theme) in the BODC parameter semantic model.	BODC
S120	BODC parameter semantic model biological entity subcomponent terms	Controlled vocabulary defining the terms that may be used for biological entity (component of the "what" theme) plant or animal parts in the BODC parameter semantic model.	BODC
S130	BODC parameter semantic model biological entity subgroup terms	Controlled vocabulary defining the terms that may be used for ad-hoc qualification of a biological entity (component of the what theme) in the BODC parameter semantic model.	BODC

IDENTIFIER	NAME	DEFINITION	GOVERNANCE
S140	BODC parameter semantic model biological entity morphology terms	Controlled vocabulary defining the terms that may be used for morphology of a biological entity (component of the what theme) in the BODC parameter semantic model.	BODC
S150	BODC parameter semantic model biological entity colour terms	Controlled vocabulary defining the terms that may be used for colour or optical appearance of a biological entity (component of the what theme) in the BODC parameter semantic model.	BODC
S210	BODC parameter semantic model sphere names	Controlled vocabulary defining the terms that may be used to describe the sphere (SWEET EarthRealm) in the BODC parameter semantic model. Part of the where theme.	BODC

Appendix C. Example MMI ORR list return

```

<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:omvmmi="http://mmisw.org/ont/mmi/20081020/ontologyMetadata/"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:omv="http://omv.ontoware.org/2005/05/ontology#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns="http://mmisw.org/ont/calcofi/bottleqc/"
  xml:base="http://mmisw.org/ont/calcofi/bottleqc">
  <owl:Ontology rdf:about="">
    <omv:description>The California Cooperative Oceanic Fisheries Investigations (CalCOFI) are a
    unique partnership of the California Department of Fish and Game, NOAA Fisheries Service, and
    Scripps Institution of Oceanography. They conduct quarterly cruises off southern & central
    California, collecting a suite of hydrographic and biological data on station and underway. These are
    Bottle Data Precision-Quality codes.</omv:description>
    <omvmmi:origVocKeywords>bottle, quality, code</omvmmi:origVocKeywords>
    <omv:acronym>bottleqc</omv:acronym>
    <omvmmi:hasContentCreator>CalCOFI</omvmmi:hasContentCreator>
    <omvmmi:origVocUri>http://cisweb4.ucsd.edu/references/data-formats/215-art-
    dbtables.html</omvmmi:origVocUri>
    <dc:contributor>CalCOFI: SIO, NOAA, California Department of Fish & Game, CCE-
    LTER</dc:contributor>
    <omv:creationDate>2010-09-07T08:28:30+0000</omv:creationDate>
    <omvmmi:origVocSyntaxFormat>Tab-delimited</omvmmi:origVocSyntaxFormat>
    <omvmmi:hasResourceType>qcCategory</omvmmi:hasResourceType>
    <omv:version>20100706T014033</omv:version>
    <omvmmi:origVocDescriptiveName>CalCOFI Database
    Tables</omvmmi:origVocDescriptiveName>
    <omv:name>CalCOFI Bottle Data Precision-Quality Codes</omv:name>
    <omvmmi:contact>http://www.calcofi.org/component/contact/12.html</omvmmi:contact>
    <dc:source>http://cisweb4.ucsd.edu/references/data-formats/215-art-dbtables.html</dc:source>
    <dc:date>2010-07-06T01:40:33+0000</dc:date>

    <omvmmi:shortNameUri>http://mmisw.org/ont/mmi/resourcetype/qcCategory</omvmmi:shortNameUri>
    <omvmmi:origMaintainerCode>calcofi</omvmmi:origMaintainerCode>
    <dc:description>The California Cooperative Oceanic Fisheries Investigations (CalCOFI) are a
    unique partnership of the California Department of Fish and Game, NOAA Fisheries Service, and
    Scripps Institution of Oceanography. They conduct quarterly cruises off southern & central
    California, collecting a suite of hydrographic and biological data on station and underway. These are
    Bottle Data Precision-Quality codes.</dc:description>
    <omv:documentation>http://www.calcofi.org/references.html</omv:documentation>
    <omv:hasContributor>CalCOFI: SIO, NOAA, California Department of Fish & Game, CCE-
    LTER</omv:hasContributor>
    <omvmmi:creditRequired>yes</omvmmi:creditRequired>
    <omv:reference>http://www.calcofi.org/publications.html</omv:reference>
    <omvmmi:contactRole>content manager</omvmmi:contactRole>
    <omvmmi:creditCitation>Copyright (c) 2010 - CalCOFI</omvmmi:creditCitation>
    <omv:usedOntologyEngineeringTool

  rdf:resource="http://mmisw.org/ont/mmi/20081020/ontologyMetadata/voc2rdf"/>
  <omvmmi:origVocManager>CalCOFI</omvmmi:origVocManager>
  <dc:creator>CalCOFI</dc:creator>
  <omv:keywords>bottle, data, quality</omv:keywords>
  <omv:uri>http://mmisw.org/ont/calcofi/bottleqc</omv:uri>
  <omvmmi:origVocDocumentationUri>http://cisweb4.ucsd.edu/references/data-formats/215-art-
  dbtables.html</omvmmi:origVocDocumentationUri>
  <omv:hasCreator>CalCOFI</omv:hasCreator>

```

```

    <omvmmi:temporaryMmiRole>ontology producer</omvmmi:temporaryMmiRole>
  </owl:Ontology>
  <owl:Class rdf:about="http://mmisw.org/ont/calcofi/bottleqc/Bottle_Qual">
    <rdfs:label>Bottle_Qual</rdfs:label>
  </owl:Class>
  <owl:DatatypeProperty rdf:about="http://mmisw.org/ont/calcofi/bottleqc/Name">
    <rdfs:domain rdf:resource="http://mmisw.org/ont/calcofi/bottleqc/Bottle_Qual"/>
    <rdfs:label>Name</rdfs:label>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:about="http://mmisw.org/ont/calcofi/bottleqc/Description">
    <rdfs:domain rdf:resource="http://mmisw.org/ont/calcofi/bottleqc/Bottle_Qual"/>
    <rdfs:label>Description</rdfs:label>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:about="http://mmisw.org/ont/calcofi/bottleqc/Type">
    <rdfs:domain rdf:resource="http://mmisw.org/ont/calcofi/bottleqc/Bottle_Qual"/>
    <rdfs:label>Type</rdfs:label>
  </owl:DatatypeProperty>
  <Bottle_Qual rdf:about="http://mmisw.org/ont/calcofi/bottleqc/Btl_Cnt">
    <Description>Autonumbered Bottle Count</Description>
    <Type>Long Integer</Type>
    <Name>Btl_Cnt</Name>
    <rdfs:label>Btl_Cnt</rdfs:label>
  </Bottle_Qual>
  <Bottle_Qual rdf:about="http://mmisw.org/ont/calcofi/bottleqc/S_prec">
    <Description>Salinity precision</Description>
    <Type>Integer</Type>
    <Name>S_prec</Name>
    <rdfs:label>S_prec</rdfs:label>
  </Bottle_Qual>
  <Bottle_Qual rdf:about="http://mmisw.org/ont/calcofi/bottleqc/T_prec">
    <Description>Temperature precision</Description>
    <Type>Integer</Type>
    <Name>T_prec</Name>
    <rdfs:label>T_prec</rdfs:label>
  </Bottle_Qual>
  <Bottle_Qual rdf:about="http://mmisw.org/ont/calcofi/bottleqc/O_qual">
    <Description>Oxygen precision</Description>
    <Type>Integer</Type>
    <Name>O_qual</Name>
    <rdfs:label>O_qual</rdfs:label>
  </Bottle_Qual>
  <Bottle_Qual rdf:about="http://mmisw.org/ont/calcofi/bottleqc/S_qual">
    <Description>Salinity quality</Description>
    <Type>Integer</Type>
    <Name>S_qual</Name>
    <rdfs:label>S_qual</rdfs:label>
  </Bottle_Qual>
  <Bottle_Qual rdf:about="http://mmisw.org/ont/calcofi/bottleqc/O2Satq">
    <Description>Oxygen saturation quality</Description>
    <Type>Integer</Type>
    <Name>O2Satq</Name>
    <rdfs:label>O2Satq</rdfs:label>
  </Bottle_Qual>
  <Bottle_Qual rdf:about="http://mmisw.org/ont/calcofi/bottleqc/T_qual">
    <Description>Temperature quality</Description>
    <Type>Integer</Type>
    <Name>T_qual</Name>
    <rdfs:label>T_qual</rdfs:label>
  </Bottle_Qual>
  <Bottle_Qual rdf:about="http://mmisw.org/ont/calcofi/bottleqc/SThtaq">
    <Description>Sigma-theta quality</Description>

```

```
<Type>Integer</Type>
<Name>SThtaq</Name>
<rdfs:label>SThtaq</rdfs:label>
</Bottle_Qual>
</rdf:RDF>
```

Appendix D. Example USGS Thesaurus XML Payload

```
<vocabularyservices>
  <result>
    <term>
      <term_id>
        42
      </term_id>
      <string>
        algal blooms
      </string>
      <relation_type_id>
        2
      </relation_type_id>
    </term>
    <term>
      <term_id>
        122
      </term_id>
      <string>
        bacteria
      </string>
      <relation_type_id>
        2
      </relation_type_id>
    </term>
    <term>
      <term_id>
        1242
      </term_id>
      <string>
        periphyton
      </string>
      <relation_type_id>
        2
      </relation_type_id>
    </term>
    <term>
      <term_id>
        1263
      </term_id>
      <string>
        phycology
      </string>
      <relation_type_id>
        2
      </relation_type_id>
    </term>
    <term>
      <term_id>
        1336
      </term_id>
      <string>
        protests
      </string>
      <relation_type_id>
        2
      </relation_type_id>
    </term>
  </result>
</vocabularyservices>
```

```
<resume>
  <status>
    available
  </status>
  <param>
    <task>
      fetchRelated
    </task>
    <arg>
      41
    </arg>
  </param>
  <web_service_version>
    0.4
  </web_service_version>
  <version>
    TemaTres 1.1
  </version>
  <cant_result>
    5
  </cant_result>
</resume>
</vocabularyservices>
```

Appendix E. Example WoRMS Aphia Records

Example Aphia Record :

```
(AphiaRecord){
  AphiaID = 137111
  url = "http://www.marinespecies.org/aphia.php?p=taxdetails&id=137111"
  scientificname = "Tursiops truncatus"
  authority = "(Montagu, 1821)"
  rank = "Species"
  status = "accepted"
  valid_AphiaID = 137111
  valid_name = "Tursiops truncatus"
  valid_authority = "(Montagu, 1821)"
  kingdom = "Animalia"
  phylum = "Chordata"
  cls = "Mammalia"
  order = "Cetartiodactyla"
  family = "Delphinidae"
  genus = "Tursiops"
  citation = "Perrin, W. (2009). Tursiops truncatus (Montagu, 1821). In: Perrin, W.F. World Cetacea Database.
  Accessed through: World Register of Marine Species at
  http://www.marinespecies.org/aphia.php?p=taxdetails&id=137111 on 2010-08-18"
}
```

Example aphia vernacular return demonstrating multi-lingual searching:

```
[(Vernacular){
  vernacular = "afale"
  language_code = "tr"
  language = "Turkish"
}, (Vernacular){
  vernacular = "afalina"
  language_code = "ru"
  language = "Russian"
}, (Vernacular){
  vernacular = "afalina"
  language_code = "tr"
  language = "Turkish"
}, (Vernacular){
  vernacular = "bolshoi delfin"
  language_code = "ru"
  language = "Russian"
}, (Vernacular){
  vernacular = "bottlenose dolphin"
  language_code = "en"
  language = "English"
}, (Vernacular){
  vernacular = "bottlenosed dolphin"
  language_code = "en"
  language = "English"
}, (Vernacular){
  vernacular = "chernyi delfin"
  language_code = "ru"
  language = "Russian"
}, (Vernacular){
  vernacular = "common bottlenose dolphin"
  language_code = "en"
  language = "English"
}
```

```
}, (Vernacular){  
  vernacular = "dauphin à gros nez"  
  language_code = "fr"  
  language = "French"  
}, (Vernacular){  
  vernacular = "dauphin nesarnak"  
  language_code = "fr"  
  language = "French"  
}, (Vernacular){  
  vernacular = "delfin geddumu qasir"  
  language_code = "mt"  
  language = "Maltese"  
}, (Vernacular){  
  vernacular = "delfin kabir"  
  language_code = "ar"  
  language = "Arabic"  
}, (Vernacular){  
  vernacular = "delfín mular"  
  language_code = "es"  
  language = "Spanish"  
}, (Vernacular){  
  vernacular = "délfin-mular"  
  language_code = "pt"  
  language = "Portuguese"  
}, (Vernacular){  
  vernacular = "delfino maggiore"  
  language_code = "it"  
  language = "Italian"  
}, (Vernacular){  
  vernacular = "delfino soffiatore"  
  language_code = "it"  
  language = "Italian"  
}, (Vernacular){  
  vernacular = "dobri dupin"  
  language_code = "cr"  
  language = "Croatian"  
}, (Vernacular){  
  vernacular = "dolphinan yam hatichon"  
  language_code = "he"  
  language = "Hebrew"  
}, (Vernacular){  
  vernacular = "golfinho-flíper"  
  language_code = "pt"  
  language = "Portuguese"  
}, (Vernacular){  
  vernacular = "golfinho-nariz-de-garrafa"  
  language_code = "pt"  
  language = "Portuguese"  
}, (Vernacular){  
  vernacular = "grand dauphin"  
  language_code = "fr"  
  language = "French"  
}, (Vernacular){  
  vernacular = "grand souffleur"  
  language_code = "fr"  
  language = "French"  
}, (Vernacular){  
  vernacular = "grosser Tümmler"  
  language_code = "de"  
  language = "German"  
}, (Vernacular){
```

```
vernacular = "handa iruka"  
language_code = "ja"  
language = "Japanese"  
}, (Vernacular){  
vernacular = "hochcöpliger Tümmeler"  
language_code = "de"  
language = "German"  
}, (Vernacular){  
vernacular = "lielaa-iuras-tsuuka"  
language_code = "lv"  
language = "Latvian"  
}, (Vernacular){  
vernacular = "nezarnak"  
language_code = "ge"  
language = "Greenlandic"  
}, (Vernacular){  
vernacular = "oresvin"  
language_code = "sv"  
language = "Swedish"  
}, (Vernacular){  
vernacular = "oudre"  
language_code = "fr"  
language = "French"  
}, (Vernacular){  
vernacular = "rinodélfino"  
language_code = "gr"  
language = "Greek"  
}, (Vernacular){  
vernacular = "roaz"  
language_code = "pt"  
language = "Portuguese"  
}, (Vernacular){  
vernacular = "souffleur"  
language_code = "fr"  
language = "French"  
}, (Vernacular){  
vernacular = "tandthöije"  
language_code = "no"  
language = "Norwegian"  
}, (Vernacular){  
vernacular = "toninha"  
language_code = "pt"  
language = "Portuguese"  
}, (Vernacular){  
vernacular = "tuimelaar"  
language_code = "nl"  
language = "Dutch"  
}, (Vernacular){  
vernacular = "tümeler"  
language_code = "no"  
language = "Norwegian"  
}, (Vernacular){  
vernacular = "Tümmeler"  
language_code = "de"  
language = "German"  
}, (Vernacular){  
vernacular = "tursio"  
language_code = "it"  
language = "Italian"  
}, (Vernacular){  
vernacular = "tursion"
```

```
language_code = "fr"  
language = "French"  
, (Vernacular){  
  vernacular = "tursione"  
  language_code = "it"  
  language = "Italian"  
, (Vernacular){  
  vernacular = "tursiope troncato"  
  language_code = "it"  
  language = "Italian"  
, (Vernacular){  
  vernacular = "tursiops"  
  language_code = "fr"  
  language = "French"  
, (Vernacular){  
  vernacular = "Velika pliskovka"  
  language_code = "sn"  
  language = "Slovenian"  
}]
```