Project No. 249024

NETMAR

Open service network for marine environmental data

Instrument: *Please tick*	CA	✓ **STREP**	IP	NOE

ICT - Information and Communication Technologies Theme

Deliverable 3.8 Prototype interlinked ontology resource fully populated to NETMAR requirements

Reference: d3_8_revision_2012_July_30.docx

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

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

British Oceanographic Data Centre, Natural Environment Research Council

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)	

 	**NETMAR** Open service network for marine environmental data Project Reference: 249024 Contract Type: Collaborative Project Start/End Date: 01/03/2010 - 31/01/2013 Duration: 36 months
	Coordinator: Prof. Stein Sandven Nansen Environmental and Remote Sensing Center Thormøhlensgate 47, Bergen, Norway Tel.: +47-55205800 Fax. +47 55 20 58 01 E-mail: stein.sandven@nersc.no

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)
  Contact: Mr Mickael Treguer (mickael.treguer@ifremer.fr)
- Norwegian Meteorological Institute (METNO), Norway.
  Contact: Dr Jan Ivar Pladsen (janip@met.no)

Author(s)

- Dr. Adam Leadbetter, British Oceanographic Data Centre (NERC). alead@bodc.ac.uk
- Dr. Roy Lowry, British Oceanographic Data Centre (NERC). rkl@bodc.ac.uk

Document approval

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

Revision History

Issue	Date	Change records	Author(s)
Draft	2012-06-25	First draft of the report.	A. M. Leadbetter R. K. Lowry
Draft 2	2012-07-10	Second draft incorporating comments from Torill Hamre; and Pilot 2 semantic requirements from Vincent Gouriou	A. M. Leadbetter
Draft 3	2012-07-25	Third draft responding to external review by Sean Bechhoffer. Clarification as to the purpose of Chapter 3. Justification for use of OWL sameAs.	A. M. Leadbetter
Draft 4	2012-07-30	Fourth draft incorporating comments from Roy Lowry. Added extra definitions to chapter 2's glossary. Updated scheme concept counts and ICAN scheme diagram.	A. M. Leadbetter
1	2012-07-30	Final version of the report approved by the coordinator.	T. Hamre

Executive Summary

The NETMAR project is designed around the aim of producing a pilot system for searching, downloading and integrating satellite, in-situ and model data within a web based context. This system utilises semantic resources to identify and accessing distributed data and services exposed to the system.

The content of the semantic resources developed within the NETMAR project was designed to support a set of scenarios for the four defined pilots which were designed to test the pilot software system.

1. Arctic Sea Ice and Met-ocean Observing System
2. Near real time monitoring and forecasting of oil spill
3. Ocean Colour - Marine Ecosystem, Research and Monitoring
4. International Coastal Atlas Network (ICAN) for coastal zone management

As such the users identified by these scenarios were consulted through a series of workshops and meetings about their requirements. The semantic requirements for three of the pilots have been amalgamated into one oceanographic domain resource which is structured around the broad themes, known as facets, of

- Measured or modelled parameter (sub-divided into theme and discipline)
- Scientific instrument used to generate the parameter
- Platform from which the instrument was deployed
- Platform class
- Research or monitoring project from which the data were generated
- Spatial coverage

A second semantic resource has been developed for the final pilot, focussed on supporting the needs of the International Coastal Atlas Network, which contains the facets of

- Agents of coastal change
- Effects of coastal change
- Human responses to coastal change

and maps to resources developed by existing nodes of the Network.

In order to be of use in marking up data and metadata for both data and service discovery and use, the semantic resources must be accessible through the internet. To this end, the NERC Vocabulary Server was brought to version 2.0 (NVS2.0) and the application programming interface used to access NVS2.0 is described herein.

The resources available through NVS2.0 provide a significant backbone to the outputs of the NETMAR project, being used most visibly to users in the search and discovery client and in the service chaining editor for validation of links between separate processing services. It is also adopted by the wider international community, for instance providing the semantic resources to the FP-7 project SeaDataNet 2.

Contents

1 Introduction

1.1 Background

The Open service network for marine environmental data (NETMAR) project[1] 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 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.

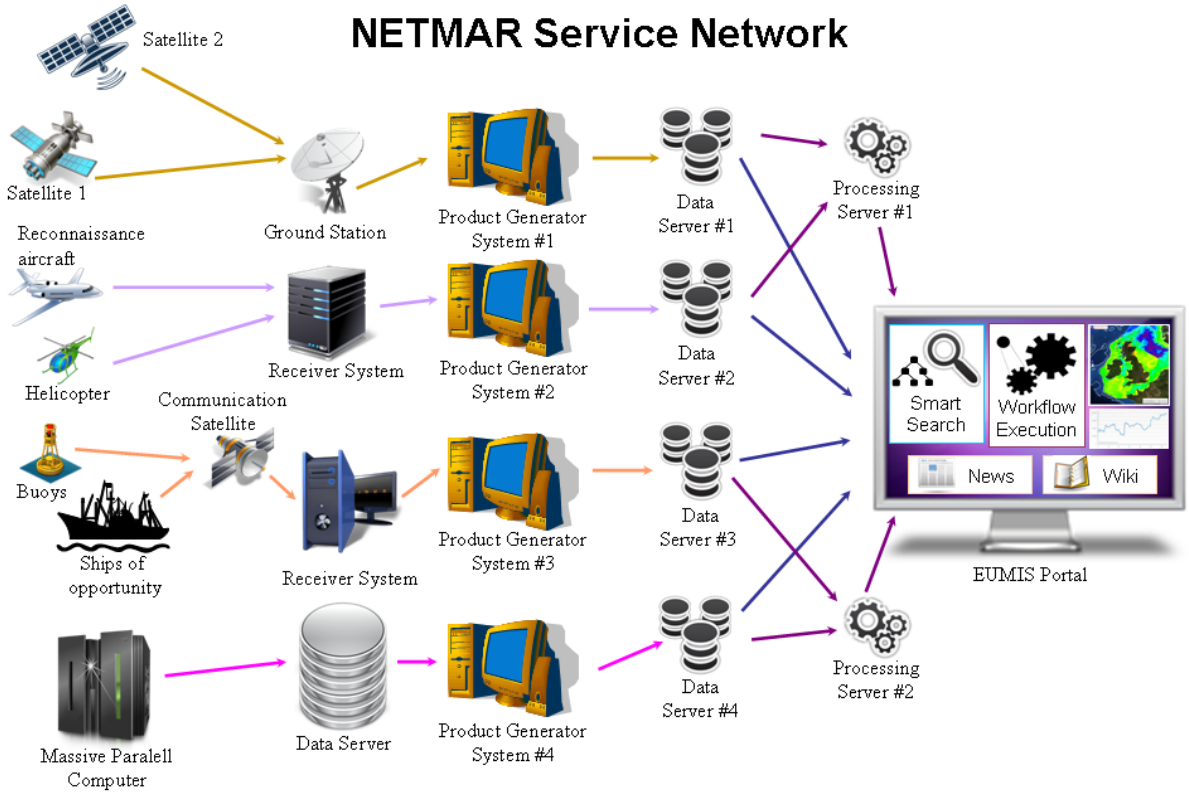

Figure 1.1 The NETMAR Service Network, where WP3 contributes to the Smart Search and to appropriate markup of data and processes from the servers of the network.

1.2 Objective of this report

The objective of developing semantic resources within the context of the NETMAR project was to develop an interconnected network of ontologies for marine environmental product and services that support dynamic search, access and chaining of services by a semantic framework.

To this end, users have been consulted about their requirements during the course of the project through a series of meetings and workshops, and the identified requirements captured in scenario and

[1] http://netmar.nersc.no/

use case descriptions, requirements lists and dedicated questionnaires [PTS+11]. The semantic requirements have been documented in [LLC11], which stated that the following facets should be supported for “smart search”:

- Theme
- Discipline
- Instrument
- Platform
- Platform_class
- Project
- Vertical_coverage

To support the “smart search” it was decided to build on existing ontologies from SeaDataNet and other projects, and extend these with new terms that we needed to support the search needed for the four pilots. Also, the extended ontologies need to support the need for semantic validation of workflows constructed by the service chaining editor [LL11].

This report aims to show the developments which have created the final NETMAR semantic resource through:

- Adopt/adapt existing tools for ontology construction and population to enable the bridging and extension of ontologies.

  The NERC Vocabulary Server (NVS) has been upgraded to Version 2. This firstly involved a redesign and reengineering of the Relational Database Management System which contains the knowledge which is served through the NVS. The software layer which converts this into an internationally accepted standard format for delivery was then completely rewritten.

  This software engineering task has made the documents returned by the NVS fully compatible with latest World Wide Web Consortium recommendations for the representation of controlled vocabularies, namely the 2009 version of the Simple Knowledge Organization System[2]. We have also incorporated the design aspects required by the NETMAR - namely the provision for: multilingual titles and definitions; and the ability to bridge into semantic resources external to the NERC Vocabulary Server. This latter functionality has been shown through mapping biological entities to the World Register of Marine Species and the Life Sciences Identifiers.

  The resulting system is now fully operational with two application programming interfaces and has been accessed from locations with a global coverage. Discussions are ongoing with the INSPIRE specification drafting teams for Atmosphere and for Oceans to host their code list on the NVS. It will also provide the semantic backbone for the EU FP7 SeaDataNet-II project.

- Populating the developed ontologies with relevant concepts and instances from the domains of the NETMAR use cases.

  The NETMAR Oceanography Thesaurus and ICAN Coastal Erosion Thesaurus have been made available.

  The NETMAR Oceanography Thesaurus has been designed in consultation with the partners involved in NETMAR pilots 1, 2 and 3. It incorporates 461 concepts from seventeen collections (including two external resources) and defines six facets:

  - Parameter
  - Instrument
  - Platform

[2] http://www.w3.org/TR/2009/REC-skos-reference-20090818/

- Platform_Class
- Project
- Vertical_Coverage

The ICAN Coastal Erosion Thesaurus was designed in consultation with the technical team of the ICAN community. It contains 340 concepts from four concept collections and defines three facets:

- Agents of Coastal Change
- Effects of Coastal Change
- Human Responses to Coastal Change

1.3 Terminology

1.3.1 *Semantic resource, thesaurus and ontology*

The phrase “semantic resource” is used to describe a group of concepts published from a given domain which may be accessed through the World Wide Web. In this context “semantic resource” is distinct from the term “resource” which may mean a concept, a list of concepts or a concept hierarchy.

In the context of the World Wide Web and semantics, a thesaurus is a related group of concepts which can be arranged in a hierarchy. Thus, some concepts will describe a wide range of ideas (have a large granularity) and can be used as an entry point to the thesaurus. These concepts will be declared as being related to other concepts in the thesaurus which have a tighter definition (have a smaller granularity) and so a concept hierarchy can be built.

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.

1.3.2 *Simple Knowledge Organisation System*

The Simple Knowledge Organization System (SKOS) is a W3C recommendation, and provides a scheme for the serialization of controlled vocabularies and thesauri. SKOS is based upon concepts that it defines as a “unit of thought”, i.e. an idea or notion such as “shoreline emergency access” or “oil spill”. Concepts may also carry other information, such as their relationships to other concepts and information about their provenance and version history. SKOS provides the means for grouping those concepts together as either collections or schemes.

A SKOS collection is a grouping of concepts which share something in common and can be conveniently grouped under a common label, for example “SeaDataNet agreed parameter groups” or “ISO19115 topic categories”. Similarly, SKOS concept schemes are also groupings of concepts but the relationships between the concepts are also a part of the concept scheme, so it is a useful model for the publication of thesauri, for example the “ICAN coastal erosion thesaurus.”

SKOS also defines three forms of relationship between concepts. A concept may be broader or narrower than another concept, or related to another concept. The related attribute allows the loose mapping of one concept to another. The broader and narrower attributes allow the construction of a hierarchy. If a concept belongs to a hierarchical scheme and is an entry point to that hierarchy (that is, at the top of the tree) it can be declared as a SKOS topConcept. For concepts in the same scheme, the broader and narrower relations may be said to be transitive; that is a concept two levels below a given concept can be inferred to be narrower than the concept in question without explicitly stating a relationship. For example (and illustrated below), eBay has “Sporting Goods” as a top level auction category, or a topConcept. Narrower than this is “Sailing”, and still narrower is “Rope”. If these relationships were declared as transitive “Rope” could be inferred to be narrower than “Sporting Goods”. The NETMAR semantic framework has additionally extended the SKOS model to allow synonyms to be identified using the Web Ontology Language’s sameAs attribute. The differences between SKOS concept collections and concept schemes are very limited in the W3C’s specification.

The NETMAR project has chosen to use schemes as a discovery tool for concepts, and collections to store and publish concepts and for referencing their identifiers.

NVS2.0 uses OWL's sameAs only in its strictest sense: where the terms identified are identical and their properties should be inherited (Halpin et al., 2010). However, this has its drawbacks in situations where the two concepts being mapped are not necessarily identical in both definition and properties and may be addressed by the addition of extra near-synonymous relationships in future releases of the NERC Vocabulary Server.

1.3.3 *Representational State Transfer and Simple Object Access Protocol*

There are two main standards for building web services, both of which are implemented in the NERC Vocabulary server.

The first is Representational State Transfer (ReST or ReSTful) the architecture of which services relies on clients and servers. In ReST, clients issue requests to the servers, which then process the requests and return appropriate responses. Requests and responses are built around the transfer of representations, or serializations, of resources. A resource is any meaningful concept that may be addressed through a network address, such as a Uniform Resource Identifier (URI) or Uniform Resource Locator (URL).

The Simple Object Access Protocol (SOAP) is a protocol specification for exchanging information in a structured manner across computer networks. SOAP relies on the XML for its messaging format. The XML based protocol incorporates

- an envelope defining the contents of the XML message and how to process it
- encoding rules for expressing application specific data types
- a representation of method calls and responses

1.3.4 *NERC Vocabulary Server*

The Natural Environment Research Council (NERC) Vocabulary Server (NVS2.0) provides access through the World Wide Web to semantic resources defining standardised terms that cover a broad spectrum of disciplines of relevance to the oceanographic and wider community.

1.4 Organisation of this report

This report is organised to present the improvements made to the initial prototype NETMAR semantic resource presented in NETMAR Deliverable 3.6, it then shows the full details of the API which can be used to access the content of NVS2.0 and goes on to illustrate the uses of the semantic resource within both the NETMAR project and the wider scientific community.

2 Glossary

For conciseness, the NERC Vocabulary Server concept collections are referred to in this section by their collection identifiers, which may be translated using the following table:

Table 2-1 NERC Vocabulary Server concept collections referred to in the remainder of this deliverable

Identifier	Label	Description
A01	International Coastal Atlas Network Coastal Erosion Global Thesaurus	Terms used at all hierarchical levels in the ICAN global thesaurus for coastal erosion. Terms used to populate drop-down lists in discovery portals.
A02	Oregon Coastal Atlas Coastal Erosion Thesaurus markup terms	Terms used at all hierarchical levels in the Oregon Coastal Atlas for coastal erosion layers, datasets and discovery.
A03	Oregon Coastal Atlas Coastal Erosion Thesaurus discovery terms	Terms used at all hierarchical levels in the Oregon Coastal Atlas for categorisation and discovery, but not for layer or dataset markup.
A04	MIDA Coastal Erosion Thesaurus	A collection of terms used by the Irish Marine Data Atlas in association with the topic of coastal erosion
C17	ICES Platform Codes	Identifiers and metadata for platform instances (combinations of names and physical entities such as hulls or airframes).
C40	Bonn Agreement pollution report accuracy	Terms used to classify the accuracy and reliability of pollution reports filed under the Bonn Agreement.
C41	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.
L05	SeaDataNet device categories	Terms used to classify groups of sensors, instruments or samplers (devices that collect water, suspended matter, sediment, rock, air or biota samples).
L06	SeaVoX Platform Categories	2-level grouping term hierarchy used for vehicles, structures or organisms capable of bearing instruments or tools for the collection of physical, chemical, geological or biological samples or data.
L13	SeaVoX Vertical Co-ordinate Coverage's	Terms used to describe data coverage over the vertical (z) co-ordinate.
L19	SeaDataNet keyword types	Terms used in SeaDataNet metadata to describe the purpose of a keyword. An extension of the ISO19115 KeywordTypeCode codelist.
P01	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.
P02	SeaDataNet Parameter Discovery Vocabulary	Terms describing fine-grained related groups of measurement phenomena designed to be used in dataset discovery interfaces.
P03	SeaDataNet Agreed Parameter Groups	Terms agreed within the EU SeaDataNet community to describe coarse-grained groupings of related measurement phenomena.
P06	BODC data storage	Terms used by BODC to describe the measurement units for

	units	data held in its repositories.
P07	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).
P08	SeaDataNet Parameter Disciplines	Terms used to classify SeaDataNet Agreed Parameter Groups to provide topic/theme level terms in a hierarchical parameter discovery interface.
P19	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.
P24	Units of measure dimensions	Classification concepts for units of measure based upon the fundamental SI quantities upon which they are based.
P25	NETMAR aggregation parameters	Terms used to describe phenomena within the NETMAR project at a granularity level suitable for labelling data aggregations or assessing whether the parameter is appropriate as an input to a given WPS.
P27	Oil spill spillet location	Terms used to describe the environmental spatial location of an aliquot of a simulated oil spill by the METEO France MOTHY model.
P28	Cedre pollution site cleanup operational stati	Terms used to describe the operational stati of pollution cleanup sites in the Cedre cleanup site service.
S25	BODC parameter semantic model biological entity names	Terms used to describe biological entities (organisms or parts thereof) in the BODC Parameter Usage Vocabulary

Concept schemes are also referred to by their identifiers, which may be translated using the following table.

Table 2-2 NERC Vocabulary Server concept schemes referred to in the remainder of this deliverable

Identifier	Label	Description
ICANCOERO	International Coastal Atlas Network Coastal Erosion Thesaurus	Thesaurus containing coastal erosion dataset (including GIS layer) terms compiled by ICAN and mapped to a global thesaurus. Includes both markup and discovery terms from the mapped components.
ICANDIS	International Coastal Atlas Network Coastal Erosion Discovery Thesaurus	Thesaurus of discovery terms that may be mapped to terms used to mark up atlas component data or metadata.
NETMAR_OCEAN	NETMAR Oceanography Thesaurus	Thesaurus centred on parameter descriptions from the BODC PUV and CF Standard Names mapping to parameter discovery, units, instrument, platform, project and coverage concepts. Underpins discovery and semantic service chain validation for the NETMAR sea ice and oceanography pilots.
NETOC_INSTRUMENT	NETMAR Oceanography Thesaurus Instrument Facet	Instrument facet of a thesaurus centred on parameter descriptions from the BODC PUV and CF Standard Names mapping to parameter discovery, units, instrument, platform, project and coverage concepts. Underpins discovery and semantic service chain validation for the NETMAR sea ice and oceanography pilots.
NETOC_PARAM	NETMAR Oceanography	Parameter facet of a thesaurus centred on

	Thesaurus Parameter Facet	parameter descriptions from the BODC PUV and CF Standard Names mapping to parameter discovery, units, instrument, platform, project and coverage concepts. Underpins discovery and semantic service chain validation for the NETMAR sea ice and oceanography pilots.
NETOC_PLAT	NETMAR Oceanography Thesaurus Platform Facet	Platform facet of a thesaurus centred on parameter descriptions from the BODC PUV and CF Standard Names mapping to parameter discovery, units, instrument, platform, project and coverage concepts. Underpins discovery and semantic service chain validation for the NETMAR sea ice and oceanography pilots.
NETOC_PLATCLASS	NETMAR Oceanography Thesaurus Platform Class Facet	Platform class facet of a thesaurus centred on parameter descriptions from the BODC PUV and CF Standard Names mapping to parameter discovery, units, instrument, platform, project and coverage concepts. Underpins discovery and semantic service chain validation for the NETMAR sea ice and oceanography pilots.
NETOC_PROJ	NETMAR Oceanography Thesaurus Platform Project Facet	Project facet of a thesaurus centred on parameter descriptions from the BODC PUV and CF Standard Names mapping to parameter discovery, units, instrument, platform, project and coverage concepts. Underpins discovery and semantic service chain validation for the NETMAR sea ice and oceanography pilots.
NETOC_VCOV	NETMAR Oceanography Thesaurus Platform Vertical Coverage Facet	Vertical coverage class facet of a thesaurus centred on parameter descriptions from the BODC PUV and CF Standard Names mapping to parameter discovery, units, instrument, platform, project and coverage concepts. Underpins discovery and semantic service chain validation for the NETMAR sea ice and oceanography pilots.
OCACETHES	Oregon Coastal Atlas Coastal Erosion Thesaurus	Thesaurus of terms associated with coastal erosion in the OCA. Includes both markup terms and their discovery categories.

3 Actions required in response to Deliverable 3.7

3.1 Action for Pilot 1 (Arctic Sea Ice and Met-ocean Observing System)

Two issues were raised by this pilot:

1. The prototype NETMAR semantic resource did not include any parameters for atmospheric winds.

2. The instrument facet didn't cover data generated by numerical models.

The solution to issue (1) was to expand the NETMAR semantic resource to include wind velocity derived from SAR images and calculated (at a standard height of 10m) by numerical models. This was achieved by the extensions to the vocabulary server detailed below.

The second issue was described as 'resolved' in D3.7 by incorporation of the 'MODELS' term from the GCMD Platforms vocabulary into the semantic resource. However, this was considered unsatisfactory. The definition used for the platform concept is ' vehicles, structures or organisms capable of bearing instruments or tools for the collection of physical, chemical, geological or biological samples or data'. A numerical model doesn't conform to this definition. However, it does conform to the definition of a 'data production tool' (devices that create data).

The ideal solution would be to incorporate a concept from the vocabulary used for the instrument facet (SeaDataNet L05). Unfortunately one doesn't exist. Furthermore, numerical models do not conform to the SeaDataNet definition of a 'device' (sensors, instruments or mechanisms that collect water, suspended matter, sediment, rock, air or biota samples). However, it is believed that there is a case that could be made to the SeaDataNet governance for broadening this definition and including terms in L05 to cover models. Should this not be accepted then it will be necessary to replace L05 in the semantic resource by a vocabulary under NETMAR governance. The details of the required process are given below.

3.1.1 *Vocabulary Server Back Office Changes*

The following changes were required:

- Addition of P01 codes to the NETOC_PARAM and NETMAR_OCEAN concept schemes
    - WINDSSAR - Wind speed in the atmosphere by synthetic aperture radar
    - WINDXMOD - Wind velocity (10m along x-axis) in the atmosphere by model prediction
    - WINDYMOD - Wind velocity (10m along y-axis) in the atmosphere by model prediction
- Extension of P01 to include 10m wind speed and wind direction parameters. New codes to be added to the NETOC_PARAM and NETMAR_OCEAN concept schemes.
- Ensure mapping of P01 to P06 includes all the above parameters, and there are P24 mappings to all the P06 entries mapped. This may require an extension of P24. Any new codes to be added to the NETOC_PARAM and NETMAR_OCEAN concept schemes.
- Extension of P25 to cover the new P01 codes. New codes to be added to the NETOC_PARAM and NETMAR_OCEAN concept schemes.

- Map the new P25 terms to L13, EDMERP, L05 and L06. If the terms mapped are not already included in the NETMAR concept schemes add them to NETMAR_OCEAN plus the appropriate subset scheme.

- Addition of P07 entries ‘wind_from_direction’, ‘wind_to_direction’, ‘wind_speed’, ‘wind_speed_of_gust’, ‘eastward_wind’, ’geostrophic_eastward_wind’, ’x_wind’, ‘northward_wind’, ‘geostrophic_northward_wind’ and ’y_wind’ to the NETOC_PARAM and NETMAR_OCEAN concept schemes. Creation of mappings for these concepts to P25.

- Ensure the additional P07 entries have mappings to P02 in place. If not, create them.

- Add P02 concept ‘Wind speed and direction’ (EWSB) to the NETOC_PARAM and NETMAR_OCEAN concept schemes.

- Consider whether any additional concepts from P19 or C174 are required (e.g. SAR satellites). If they are add them to the NETOC_PARAM and NETMAR_OCEAN concept schemes and create mappings to L06.

3.1.2 *L05 Vocabulary Extension or Replacement*

The following work plan was proposed.

3.1.2.1 Primary strategy

- Consult with Bob Arko (lead systems analyst at Lamont Doherty Earth Observatory, Columbia University, United Statues who trialled the approach of considering numerical models as sensors and reported problems through personal communication) to determine whether the issues he encountered apply at the L05 granularity level. If not, activate the SeaDataNet governance to change the L05 entity definition and create L05 concept or concepts for numerical models.

- Add new L05 term to the NETOC_PARAM, NETOC_INSTRUMENT and NETMAR_OCEAN concept schemes and create mappings toP25.

3.1.2.2 Backup strategy

- Create a new data production tool category vocabulary under NETMAR governance

- Replace the L05 entries in the NETOC_PARAM, NETOC_INSTRUMENT and NETMAR_OCEAN concept schemes by concepts from the new vocabulary.

- Replace mappings between L05 and P25 by mappings between P25 and the new vocabulary.

3.2 Action for Pilot 2 (Oil Spill Drift Forecasting and Shoreline Cleanup)

The issues raised by pilot 2 were as follows:

1. There are no markup resources in the NETMAR ontology to cover the keyword types:
    a. pollution probability
    b. pollution source
    c. pollution type
    d. cleanup techniques
    e. Pollution nature

2. There is a need for these keyword types to be discovered by personnel who do not speak English

3. The instrument facet in the NETMAR semantic resource didn't cover data generated by numerical models or human observers (although the latter is present in the SeaDataNet L05 vocabulary).

The solution to issue (1) required a little analysis. Both P01 and P25 are collections of 'parameter' concepts. P01 is primarily designed for markup of data streams, but may also be used for detailed metadata markup. P25 is specifically designed to hold parameter descriptions in support of semantic verification of WPS connections. The term 'parameter' may be considered as a description of what has been quantified by a measurement. Based upon this, 'pollution probability' can be regarded as the parameter and the concepts in the C40 collection may be regarded as measurements of the 'pollution probability' parameter expressed as descriptive text. There are clear precedents for this, such as describing wind speed as 'gale force'.

As these 'parameters' relate to metadata markup and not to the markup of WPS connections the vocabulary in which they belong is P01. Therefore it was proposed that P01 be extended to cover the markup requirements of pilot 2.

The solution to issue (2) was to provide multilingual entries (English, French, Norwegian, Portuguese, Spanish) for the additional terms added to P01.

The first part of issue (3) is exactly the same issue raised by Pilot 1 and was resolved through implementation of the strategy described above for that pilot. The second part simply required an additional L05 term (91 – observers) adding to the NETMAR_OCEAN and NETOC_INSTRUMENT concept schemes.

3.2.1 *Vocabulary Server Back Office Changes*

- Extension of P01 to include 'pollution probability', 'pollution source',' pollution type', 'cleanup techniques', and 'pollution nature' parameters. Add the new codes to the NETOC_PARAM and NETMAR_OCEAN concept schemes and map them to the P06 'dimensionless' concept.

- Add the P02 'Pollution Events' concept (GP0001), the P03 'Anthropogenic contamination' concept (H001) and the P08 'Environment' concept (DS10) to the NETOC_PARAM and NETMAR_OCEAN concept schemes. The mappings between P01 and P02 are automatically created as part of the P01 concept creation process. Mappings between P02/P03 and P03/P08 already exist.

- Add the L05 'observers' concept to the NETOC_INSTRUMENT and NETMAR_OCEAN concept schemes and map it to the new P01 entries. Note this is an additional mapping in the NETMAR semantic resource, but its introduction makes more sense than the introduction of non-numeric parameters to P25.

3.3 Action for Pilot 3 (Ocean Colour – Marine Ecosystem, Research and Monitoring)

Three issues were raised by this pilot:

1. The prototype NETMAR semantic resource did not include any parameters for normalised water-leaving radiance, dissolved oxygen, fluorescence or turbidity.

2. A discovery strategy is required to distinguish between data from the MODIS and MERIS instruments.

3. The instrument facet didn't cover data generated by numerical models.

The solution to issue (1) was to expand the NETMAR semantic resource to include the parameters specified. A detailed description of how this was achieved is given in the 'Vocabulary Server Back Office Changes' section.

The second issue was addressed by the addition of a small number of mappings to the NETMAR concept schemes. MODIS and MERIS are described in the instrument facet by a single term - 'ocean colour radiometers'. However, their data are marked up by sensor-specific concepts from P01 such as 'Concentration of chlorophyll-a {chl-a} per unit volume of the water body by programmable medium-spectral resolution imaging spectrometer (MERIS) and OC5 algorithm'. If these concepts are mapped as 'related to' the satellites carrying the appropriate sensor then the platform facet will automatically be populated thereby providing a differentiation mechanism. The detailed work to achieve this is described below.

The third issue was exactly the same issue raised by Pilot 1 and was resolved through implementation of the strategy described above for that pilot.

3.3.1 *Vocabulary Server Back Office Changes*

The following changes were required:

- Addition of P01 codes to the NETOC_PARAM and NETMAR_OCEAN concept schemes
    - RXXX412N - Normalised water-leaving radiance (412nm wavelength) from the water body
    - RXXX443N - Normalised water-leaving radiance (443nm wavelength) from the water body
    - RXXX490N - Normalised water-leaving radiance (490nm wavelength) from the water body
    - RXXX510N - Normalised water-leaving radiance (510nm wavelength) from the water body
    - RXXX555N - Normalised water-leaving radiance (555nm wavelength) from the water body
    - RXXX670N - Normalised water-leaving radiance (670nm wavelength) from the water body
    - DOXMZZXX- Concentration of oxygen {O2} per unit mass of the water body [dissolved phase]
    - DOXYZZXX - Concentration of oxygen {O2} per unit volume of the water body [dissolved phase]
    - OXYSZZ01 - Saturation of oxygen {O2} in the water body [dissolved phase]
    - FLUOZZZZ - Fluorescence of the water body
    - TURBXXXX - Turbidity of the water body
    - TURBMOXX - Turbidity of the water body by Moderate Resolution Imaging Spectroradiometer (MODIS)

  - TSEDBVOL - Concentration (volume per volume) of suspended particulate material per unit volume of the water body

  - TSEDZZZZ - Concentration of suspended particulate material {SPM} per unit volume of the water body

- Ensure mapping of P01 to P06 includes all the above parameters, and there are P24 mappings to all the P06 entries mapped. This may require an extension of P24. Any new codes to be added to the NETOC_PARAM and NETMAR_OCEAN concept schemes.

- Extension of P25 to cover the new P01 codes, including P01/P25 mappings. New codes to be added to the NETOC_PARAM and NETMAR_OCEAN concept schemes.
- Map the new P25 terms to L13, EDMERP, L05 and L06.

- Addition of P07 entries 'surface_upwelling_spectral_radiance_in_air_emerging_from_sea_water', 'volume_fraction_of_oxygen_in_sea_water', 'mole_concentration_of_dissolved_molecular_oxygen_in_sea_water' , 'moles_of_oxygen_per_unit_mass_in_sea_water','mass_concentration_of_oxygen_in_sea_water', 'fractional_saturation_of_oxygen_in_sea_water' and 'mass_concentration_of_suspended_matter_in_sea_water' and to the NETOC_PARAM and NETMAR_OCEAN concept schemes.

- Map the additional P07 entries to P25 and ensure they have mappings to P02 in place. If not, create them.

- Add P02 concepts 'Ocean colour and earth-leaving visible waveband spectral radiation' (R410),'Dissolved oxygen parameters in the water column' (DOXY), 'Raw fluorometer output' (FVLT) and 'Concentration of suspended particulate material in the water column' (TSED) to the NETOC_PARAM and NETMAR_OCEAN concept schemes.

- Add P03 concepts 'Optical properties ' (D015), 'Dissolved gases' (C015), 'Suspended particulate material' (G015) to the NETOC_PARAM and NETMAR_OCEAN concept schemes.

- Add P08 concept 'Marine geology' (DS04) to the NETOC_PARAM and NETMAR_OCEAN concept schemes.

- Add P19 concept 'Environmental Satellite' (PLSR0093) into the NETOC_PLAT and NETMAR_OCEAN concept schemes together with a mapping to L06.

- Add mappings between MERIS P01 codes and 'Environmental Satellite' (PLSR0093) in P19. Add mappings between MODIS P01 codes and 'Earth Observing System, AQUA' (PLSR0032) and 'Earth Observing System, TERRA (AM-1)' (PLSR0415).

3.4 Action for Pilot 4 (ICAN)

The following issues were identified with the draft semantic resources produced for the ICAN pilot:

1. The mapping between the A01 and the Oregon Coastal Atlas ontology was created by mapping to the markup terms in A02 rather than the discovery terms in A03. Whilst this approach was more reliable due to increased semantic clarity of the finer-grained concepts, it

provides a less able vehicle for the capabilities of semantic mediation. An approach based on mappings to broader local discovery concepts in the ontology would be preferred.

2. Less abstract sub-themes mapped to the ICAN coastal erosion themes were requested.

3. The MIDA Coastal Erosion Ontology should be converted from OWL to SKOS, mapped to A01 and included in the ICANCOERO concept scheme.

Issue (1) required the creation of a mapping between A01 and A03 place of the existing mapping between A01 and A02. The A01 to A02 mapping was retained in a readily available backup should its restoration be required in future.

Issue (2) required the creation or acquisition of additional concepts that describe 'coastal erosion' in more concrete terms. A primary aim of NETMAR is to replace large-scale creation of semantic content by utilisation of pre-existing semantic resources. The approach to this issue was therefore to trawl through existing resources, such as GEMET, for suitable concepts and map them to A01.

Issue (3) was addressed by ingestion of the MIDA Coastal Erosion Ontology into the NERC Vocabulary Server as a concept collection, including any existing mappings to A01. The collection will then be added to the ICANCOERO concept scheme.

3.5 Other Issues Requiring Work

Subsequent to the publication of D3.7, the NETMAR concept schemes were imported into the Jena triple store underpinning the NETMAR Semantic Web Service and exposed to the Jena reasoning engine. This exposed the following concepts that according to the mappings were broader than each other (i.e. mapped in an unexpected, reflexive way):

http://vocab.nerc.ac.uk/collection/P02/current/TEMP <==> http://vocab.nerc.ac.uk/collection/P02/current/TEMP
http://vocab.nerc.ac.uk/collection/P01/current/NTRIZZXX <==> http://vocab.nerc.ac.uk/collection/P01/current/NTRIZZXX
http://vocab.nerc.ac.uk/collection/P01/current/NTRIZZXX <==> http://vocab.nerc.ac.uk/collection/P02/current/NTRI
http://vocab.nerc.ac.uk/collection/P25/current/SLCA <==> http://vocab.nerc.ac.uk/collection/P25/current/SLCA
http://vocab.nerc.ac.uk/collection/P25/current/SLCA <==> http://vocab.nerc.ac.uk/collection/P02/current/SLCA
http://vocab.nerc.ac.uk/collection/P25/current/SLCA <==> http://vocab.nerc.ac.uk/collection/P01/current/SLCAZZXX
http://vocab.nerc.ac.uk/collection/P02/current/SLCA <==> http://vocab.nerc.ac.uk/collection/P25/current/SLCA
http://vocab.nerc.ac.uk/collection/P02/current/SLCA <==> http://vocab.nerc.ac.uk/collection/P02/current/SLCA
http://vocab.nerc.ac.uk/collection/P02/current/SLCA <==> http://vocab.nerc.ac.uk/collection/P01/current/SLCAZZXX
http://vocab.nerc.ac.uk/collection/P02/current/NTRA <==> http://vocab.nerc.ac.uk/collection/P02/current/NTRA
http://vocab.nerc.ac.uk/collection/P02/current/NTRA <==> http://vocab.nerc.ac.uk/collection/P01/current/NTRAZZXX
http://vocab.nerc.ac.uk/collection/P02/current/NTRZ <==> http://vocab.nerc.ac.uk/collection/P02/current/NTRZ
http://vocab.nerc.ac.uk/collection/P02/current/PSAL <==> http://vocab.nerc.ac.uk/collection/P02/current/PSAL
http://vocab.nerc.ac.uk/collection/P02/current/PHOS <==> http://vocab.nerc.ac.uk/collection/P02/current/PHOS
http://vocab.nerc.ac.uk/collection/P02/current/PHOS <==> http://vocab.nerc.ac.uk/collection/P01/current/PHOSZZXX
http://vocab.nerc.ac.uk/collection/P01/current/NTRAZZXX <==> http://vocab.nerc.ac.uk/collection/P02/current/NTRA
http://vocab.nerc.ac.uk/collection/P01/current/NTRAZZXX <==> http://vocab.nerc.ac.uk/collection/P01/current/NTRAZZXX
http://vocab.nerc.ac.uk/collection/P02/current/NTRI <==> http://vocab.nerc.ac.uk/collection/P01/current/NTRIZZXX
http://vocab.nerc.ac.uk/collection/P02/current/NTRI <==> http://vocab.nerc.ac.uk/collection/P02/current/NTRI
http://vocab.nerc.ac.uk/collection/P01/current/SLCAZZXX <==> http://vocab.nerc.ac.uk/collection/P25/current/SLCA
http://vocab.nerc.ac.uk/collection/P01/current/SLCAZZXX <==> http://vocab.nerc.ac.uk/collection/P02/current/SLCA
http://vocab.nerc.ac.uk/collection/P01/current/SLCAZZXX <==> http://vocab.nerc.ac.uk/collection/P01/current/SLCAZZXX
http://vocab.nerc.ac.uk/collection/P01/current/PHOSZZXX <==> http://vocab.nerc.ac.uk/collection/P02/current/PHOS
http://vocab.nerc.ac.uk/collection/P01/current/PHOSZZXX <==> http://vocab.nerc.ac.uk/collection/P01/current/PHOSZZXX

The mappings stored for each of these concepts were checked and any inconsistencies in the semantic logic that are identified were resolved.

3.6 Future updates

As the NETMAR Ontology Governance Group [L10] was set up with a plan for continuity, further *ad hoc* additions to either the NETMAR Oceanographic Thesaurus or the ICAN Coastal Erosion Thesaurus may be made through under the governance of this body. This will allow future users of the system to add their own semantic requirements to the existing resource.

4 Access to the NETMAR Semantic Resource

Consumers may access the vocabulary server either using the ReSTful URLs described below or via SOAP.

SOAP consumers should generate their client implementation from the WSDL available at http://vocab.nerc.ac.uk/vocab2.wsdl.

4.1 Collection, concept and scheme URLs

Collections, concepts and schemes are presented to the server as Uniform Resource Identifiers (URIs) (in this case actually URLs) having the syntax

Collections: http://vocab.nerc.ac.uk/collection/
http://vocab.nerc.ac.uk/collection/{*colRef*}/{*colVer*}/
http://vocab.nerc.ac.uk/collection/{*colRef*}/{*colVer*}/{*status*}/

Concepts: http://vocab.nerc.ac.uk/collection/{*colRef*}/{*colVer*}/{*conRef*}/

Schemes: http://vocab.nerc.ac.uk/scheme/
http://vocab.nerc.ac.uk/scheme/{*schemeRef*}/

where

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

respectively provide catalogues of the available concept collections and concept schemes.

colRef is an internal opaque identifier for the concept collection, e.g. P02 for the SeaDataNet Parameter Discovery Vocabulary.

colVer may be a valid concept collection version number or 'current' to specify the latest version of the collection.

status may be ‘all’, ‘accepted’ or ‘deprecated’ to indicate whether all concepts related to a collection should be returned, or only the accepted or deprecated concepts.

conRef is an internal opaque identifier for the concept within the concept collection, e.g. TEMP for ‘Temperature of the water column’ in the SeaDataNet Parameter Discovery Vocabulary.

schemeRef is an internal opaque identifier for the concept scheme, e.g. ICANCOERO for the International Coastal Atlas Network Coastal Erosion Thesaurus.

4.2 ReSTful and SOAP API Method Details

4.2.1 *GetCollections*

The `GetCollections` method allows the client to retrieve a list of the available SKOS concept collections from NVS2.0. This allows a client to discover the content of NVS2.0 which is available through the concept collection paradigm.

API	*Method Call Details*
ReST	**Base URL:** http://vocab.nerc.ac.uk/ **URL suffix:** collection/ **Example fully encoded URL:** http://vocab.nerc.ac.uk/collection/ **Returns:** A SKOS concept collection RDF XML document
SOAP	**Method:** getCollections **Input Parameters**: No Parameters needed **Returns:** ConceptCollection complex data type

4.3 GetConceptCollection

The `GetConceptCollection` method allows the client to retrieve all of the available metadata and all of the concepts and associated information for a given SKOS concept collection identified by its URL.

API	*Method Call Details*
ReST	**Base URL:** http://vocab.nerc.ac.uk/collection/ **URL suffix:** collectionID/versionID/status versionID is optional. If it is omitted, the versionID defaults to "current", which may also be used as a valid versionID, and returns the most up to date version of the concept collection. status is also optional. If omitted the status defaults to "all" which returns all concepts registered to the specified concept collection. Other values for status which are valid are "accepted" and "deprecated". **Example fully encoded URLs:** http://vocab.nerc.ac.uk/collection/A01/ http://vocab.nerc.ac.uk/collection/C19/2/ http://vocab.nerc.ac.uk/collection/A01/current/ http://vocab.nerc.ac.uk/collection/A01/current/all/ http://vocab.nerc.ac.uk/collection/A01/current/accepted/ http://vocab.nerc.ac.uk/collection/A01/current/deprecated/ **Returns:** A SKOS concept collection RDF XML document
SOAP	**Method:** getConceptCollection(*collectionURL,status*) **Input Parameters:** *collectionURL*: String - concept collection URL e.g. http://vocab.nerc.ac.uk/collection/A01/ *status*: String of value "all", "accepted" or "deprecated" **Returns:** ConceptCollection complex data type

4.4 GetConcept

The `GetConcept` method allows the client to retrieve all available information about a given concept, identified by its URL.

API	*Method Call Details*
ReST	**Base URL:** http://vocab.nerc.ac.uk/collection/ **URL suffix:** collectionID/versionID/conceptID versionID may either be the string "current" to return the most up to date version of the concept, or an integer number to return the version of the concept from a given version of the concept collection. **Example fully encoded URLs:** http://vocab.nerc.ac.uk/collection/C18/current/72/ http://vocab.nerc.ac.uk/collection/A01/current/Human_Responses_to_Coastal_Change **Returns:** A SKOS concept RDF XML document
SOAP	**Method:** GetConcept(*conceptURL*) **Input Parameters:** *conceptURL*: String - concept URL e.g. http://vocab.nerc.ac.uk/collection/C18/current/72/ **Returns:** concept complex data type

4.5 GetSchemes

The `GetSchemes` method allows the client to retrieve a list of and the descriptions of the concept schemes available through NVS2.0.

API	*Method Call Details*
ReST	**Base URL:** http://vocab.nerc.ac.uk/ **URL suffix:** scheme/ **Example fully encoded URL:** http://vocab. nerc.ac.uk/scheme/ **Returns:** A SKOS concept scheme RDF XML document
SOAP	**Method:** GetSchemes **Input Parameters**: No Parameters needed **Returns:** ConceptScheme complex data type

4.6 GetConceptScheme

The `GetConceptScheme` method allows the client to retrieve all of the available metadata and all of the concepts and associated information for a given SKOS concept scheme, as identified by its URL.

API	*Method Call Details*
ReST	**Base URL:** http://vocab.nerc.ac.uk/scheme/ **URL suffix:** schemeID/ **Example fully encoded URL:** http://vocab.nerc.ac.uk/scheme/ICANCOERO/ **Returns:** A SKOS concept scheme RDF XML document
SOAP	**Method:** GetConceptScheme(*schemeURL*) **Input Parameters:** *schemeURL*: String - concept Scheme URL e.g. http://vocab.nerc.ac.uk/scheme/ICANCOERO/ **Returns:** ConceptScheme complex data type

4.7 GetRelatedConcepts

The `getRelatedConcepts` method allows the client to access all of the concepts which are related to a given concept, identified by that concept's URL.

A relationship type flag is provided to the method call to determine which types of relationship are returned by the method call. This will allow, for example, the selection of only narrower matches or only broader matches facilitating relationship tree building in client interfaces. The flag is a four digit number where each integer value may be 1 or 0 to determine if the relationship should be returned or not. e.g.:

broader	narrower	sameAs	related
0	0	1	0

searches only for related concepts which are synonyms to the specified concept.

The method returns a representation of the input concept along with individual concept records of the related concepts.

This method is not available through the ReST API, only through the SOAP API.

API	*Method Call Details*
ReST	This method is unavailable through the ReST API.
SOAP	**Method:** getRelatedConcepts(*conceptURL*, *relationshipType*, *status*) **Input Parameters:** *conceptURL*: String - concept URL e.g. http://vocab.nerc.ac.uk/collection/P01/current/PSALCU01/ *relationshipType*: Integer indicating level of relationship to return as defined above *status*: String of value "all", "accepted" or "deprecated" **Returns:** RelatedConcepts complex data type

4.8 GetTopConcepts

The `getTopConcepts` method allows the client to access the concepts which are explicitly stated to be the entry points of a given SKOS concept scheme, identified by its URL.

This method is not available formally through the ReST API, only through the SOAP API.

API	Method Call Details
ReST	This method is unavailable through the ReST API.
SOAP	**Method:** getTopConcepts(*schemeURL*) **Input Parameters:** *schemeURL*: String - concept Scheme URL e.g. http://vocab.nerc.ac.uk/scheme/ICANCOERO/ **Returns:** A list of Concept complex data type objects enclosed by <getTopConcepts><topConcepts>…</topConcepts></getTopConcepts> tags

4.9 SearchVocab

The `searchVocab` method allows the client to search the knowledge encoded within NVS2.0.

This method is not available through the ReST API, only through the SOAP API. Note, there is no guaranteed consistency to the order in which the concepts are returned.

API	Method Call Details
ReST	This method is unavailable through the ReST API.
SOAP	**Method:** searchVocab(*query*, *case_sensitivity*, *term_type*, *max_results*, *multilang*, *uri_list*, *status*) **Input Parameters:** *query*: The search term to be acted on. Valid wildcard characters are: * = 1 or more characters e.g. Searches for “Salinity*” and “*alinity*” on http://vocab.nerc.ac.uk/collection/P01/current/ will yield different result sets. *case_sensitivity*: Optional Boolean value: *true* or *false*. Default action is *false*. *term_type*: String of value “uri”, “preflabel” or “altlabel” *max_results*: An optional integer to limit the number of resturned results *multilang*: Optional Boolean value to search on non-English labels: *true* or *false*. Default action is *false*. This option is included to significantly reduce the response time for searches in which multilingual functionality is not required. *uri_list*: A list of the concept collection URLs to search e.g. http://vocab.nerc.ac.uk/collection/P01/current/,http://vocab.nerc.ac.uk/collection/P02/current *status*: String of value “all”, “accepted” or “deprecated” **Returns:** SearchResults complex data type

4.10 VerifyConcept

The `verifyConcept` method is used to check the existence of a given concept within NVS2.0, as identified by its URL, its preferred label or its alternative label. This is of particular use to a client that is validating the markup of its metadata or data. The return of this method is a Boolean value, equal to `true` if the concept in question exists in NVS2.0 and `false` if it does not.

This method is not formally available through the ReST API, only through the SOAP API. However, an HTTP 200 “OK” header is returned when a ReSTful call is made to a valid concept, and an HTTP 400 header is returned when the concept is invalid.

API	*Method Call Details*
ReST	This method is formally unavailable through the ReST API, however it is informally avaialbe through HTPP header analysis.
SOAP	**Method:** verifyConcept(*concept*, *collectionURI*, *conceptType*, *status*) **Input Parameters:** String - concept Collection URL, String – Concept label or URL *concept*: String of one of the following: The full URL to the concept to be verified – use with *conceptType* = *"*uri" e.g. http://vocab.nerc.ac.uk/collection/P01/current/PSALCU01 or the concept preferred label or alternative label to be verified – use with *conceptType="*preflabel*" or conceptType="*altlabel" e.g. "Practical salinity of the water body by CTD and computation using UNESCO 1983 algorithm and NO calibration against independent measurements" collectionURI: String - The URL to the concept collection against which the concept should be verified e.g. http://vocab.nerc.ac.uk/collection/P01/ *conceptType*: String of value "uri", "preflabel" or "altlabel" *status*: String of value "all", "accepted" or "deprecated" **Returns:** Boolean value, i.e.: <verifyConcept><verified>true</verified></verifyConcept> <verifyConcept><verified>false</verified></verifyConcept>

5 ReSTful Interface XML Payload Details

The ReSTful URL access methods return XML documents conforming to the W3C's Simple Knowledge Organization System model. The use of XML tags outside the scope of that specification within these payload documents is explained in detail below.

5.1 Namespaces used

DC	http://purl.org/dc/elements/1.1/
GRG	http://www.isotc211.org/schemas/grg
OWL	http://www.w3.org/2002/07/owl#
RDF	http://www.w3.org/1999/02/22-rdf-syntax-ns#
RDFS	http://www.w3.org/2000/01/rdf-schema#
SKOS	http://www.w3.org/2004/02/skos/core#
XSD	http://www.w3.org/2001/XMLSchema#

5.2 skos:Collection

A SKOS concept collection is a group of related concepts. Each controlled vocabulary served from NVS2.0 are formalised in their representation as a `skos:Collection`. Each concept which is a member of a `skos:Collection` is enclosed by a `skos:member` tag.

5.2.1 *dc:title and skos:prefLabel*

- Mandatory
- Number
    - One per concept collection document

The Dublin Core metadata element set provides the `dc:title` tag to present the formal name given to a resource. In this instance, the `dc:title` and `skos:prefLabel` tags will carry the title of the concept collection.

5.2.2 *dc:alternative and skos:altLabel*

- Optional
- Number
    - One per concept collection document

Where a resource has more than one title by which it is known, the `dc:alternative` and `skos:altLabel` tags provides a method of encoding the alternative titles.

5.2.3 *dc:description*

- Optional
- Number
    - One per concept collection document

Often, the formal name (or names) of a resource cannot carry enough information to make the resource both discoverable and usable. In this case, a plain text description of the resource can aid in the usage of the resource. In this case, the account of the content of the resource shall be contained within `dc:description` tags.

5.2.4 *dc:date*

- Mandatory
- Number
  - One per concept collection document

The Dublin Core metadata element `dc:date` allows the inclusion of an important point in the lifecycle of the resource. In this case we use the time and date of creation of the version of the concept collection requested.

```
<dc:date>
        2011-05-31T08:00:20.136+0000
</dc:date>
```

5.2.5 *owl:versionInfo*

- Mandatory
- Number
  - One per concept collection document

The `owl:versionInfo` tag gives the published version number of the concept collection.

```
<owl:versionInfo>
        2
</owl:versionInfo>
```

5.2.6 *dc:creator*

- Mandatory
- Number
  - One or more per concept collection document

The Dublin Core metadata elements provide the creator element defined as “the entity primarily responsible for making the resource”. The Dublin Core guidelines give examples of a creator as “include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.” Therefore a `dc:creator` tag will be used to store the content governance body for a given concept scheme.

```
<dc:creator>
        SeaVox: SeaDataNet and MarineXML Vocabulary Content Governance Group
</dc:creator>
```


5.2.7 *grg:RE_RegisterOwner*

- Mandatory
- Number
    - One or many per concept collection document

The `grg:RE_RegisterOwner` tag allows the inclusion of ISO19135 compatible information regarding the person or body who owns a concept collection.

```
< grg:RE_RegisterOwner>
      SeaVox: SeaDataNet and MarineXML Vocabulary Content Governance Group
</ grg:RE_RegisterOwner>
```

5.2.8 *rdfs:comment*

- Optional for inclusion in concept collection documents
- Number
    - Zero, one or many per concept collection document

An RDF Schema (RDFS) comment is added to the NVS2.0 payload in order to provide further information about the body in charge of the content governance for the concept collection or concept scheme.

```
<rdfs:comment>
      Group set up under the joint auspices of the SeaDataNet project and
      the Intergovernmental Oceanographic Commission MarineXML Steering
      Group for controlled vocabulary governance in the marine domain
</rdfs:comment>
```

5.2.9 *grg:RE_RegisterManager*

- Mandatory
- Number
    - One per concept collection document

The `grg:RE_RegisterManager` tag allows the inclusion of ISO19135 compatible information regarding the person or body appointed by a register owner to manage a register.

```
< grg:RE_RegisterManager>
      British Oceanographic Data Centre
</ grg:RE_RegisterManager>
```

5.2.10 *dc:publisher*

- Mandatory
- Number
    - One per concept collection document

The `dc:publisher` tag allows the inclusion of the publisher of the resource

```
<dc:publisher>
      Natural Environment Research Council
</dc:publisher>
```


5.3 skos:scheme

SKOS concept schemes represent an aggregation of concepts with interconnecting semantic relationships. A concept scheme is likely to contain a hierarchy, so the SKOS collections (or parts thereof) in NVS2.0 which form thesauri may be grouped together and formalised as concept schemes. The definition of a SKOS concept scheme gives the entry points to the broadest concept definitions within the hierarchy, which are referred to as the top concepts. Each top concept is also declared to be the top concept of a concept scheme, and each concept member of a concept scheme is declared to be a member of each scheme to which it belongs.

5.3.1 *dc:title and skos:prefLabel*

- Mandatory
- Number
    - One per concept scheme document

The Dublin Core metadata element set provides the `dc:title` tag to present the formal name given to a resource. In this instance, the `dc:title` and `skos:prefLabel` tags will carry the title of the concept scheme.

5.3.2 *dc:alternative and skos:altLabel*

- Optional
- Number
    - One per concept scheme document

Where a resource has more than one title by which it is known, the `dc:alternative` and `skos:altLabel` tags provides a method of encoding the alternative titles.

5.3.3 *dc:description*

- Optional
- Number
    - One per concept scheme document

Often, the formal name (or names) of a resource cannot carry enough information to make the resource both discoverable and usable. In this case, a plain text description of the resource can aid in the usage of the resource. In this case, the account of the content of the resource shall be contained within `dc:description` tags.

5.3.4 *dc:date*

- Mandatory
- Number
    - One per concept scheme document

The Dublin Core metadata element `dc:date` allows the inclusion of an important point in the lifecycle of the resource. In this case we use the time and date of creation of the current version of the concept scheme.

```
<dc:date>
        2011-05-31T08:00:20.136+0000
</dc:date>
```


5.3.5 *owl:versionInfo*

- Mandatory
- Number
  - One per concept scheme document

The `owl:versionInfo` tag gives the published version number of the concept scheme.

```
<owl:versionInfo>
      2
</owl:versionInfo>
```

5.3.6 *dc:creator*

- Mandatory
- Number
  - One per concept scheme document

The Dublin Core metadata elements provide the creator element defined as “the entity primarily responsible for making the resource”. The Dublin Core guidelines give examples of a creator as “include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.” Therefore a `dc:creator` tag will be used to store the content governance body for a given concept scheme.

```
<dc:creator>
      SeaVox: SeaDataNet and MarineXML Vocabulary Content Governance Group
</dc:creator>
```

5.3.7 *rdfs:comment*

- Optional for inclusion in concept scheme documents
- Number
  - Zero or one per concept scheme document

An RDFS comment is added to the NVS2.0 payload in order to provide further information about the body in charge of the content governance for the concept collection or concept scheme.

```
<rdfs:comment>
      Group set up under the joint auspices of the SeaDataNet project and
      the Intergovernmental Oceanographic Commission MarineXML Steering
      Group for controlled vocabulary governance in the marine domain
</rdfs:comment>
```

5.3.8 *dc:publisher*

- Mandatory
- Number
  - One per concept scheme document

The `dc:publisher` tag allows the inclusion of the publisher of the resource

```
<dc:publisher>
      Natural Environment Research Council
</dc:publisher>
```


5.4 skos:concept

- Mandatory
- Number
    - One if the payload is in response to a request for a concept by identifier (URI)
    - Many if the payload is returned in response to a request for a list or thesaurus or a request for a concept using a query parameterized by anything other than a concept's URI

A `skos:Concept` is the base unit of currency within the NERC Vocabulary Server, on which other units such as lists and thesauri are built. Therefore each NVS response to a request shall return at least one concept. Each `skos:Concept` opening tag shall also contain the URL to the concept as an `rdf:about` subtag, e.g.:

```
<skos:concept rdf:about="http://vocab.nerc.ac.uk/collection/collid/ver/concept/">
</skos:concept>
```

Each `skos:Concept` may have associated annotations (including human-language translations), mappings, concept collections, concept schemes and provenance information.

5.4.1 skos:prefLabel

- Mandatory
- Number
    - A maximum of one per human readable language into which the concept has been translated

A `skos:Concept` returned from NVS2.0 shall have a preferred label in at least one human readable language. The `skos:prefLabel` is to contain the preferred human readable representation of the concept.

```
<skos:prefLabel xml:lang="en">
        Adriatic Sea
</skos:preLabel>
```

5.4.2 skos:altLabel

- Optional
- Number
    - Zero or one per concept

The `skos:altLabel` element can be used to provide alternative spellings or synonyms to a given concept, or to provide a lexical label for use in alternative circumstances, e.g. axes labels in plotting software.

```
<skos:altLabel xml:lang="en">
        Haloc_WC
</skos:altLabel>
```

5.4.3 skos:definition

- Mandatory

- Number
    - One per human readable language into which the concept has been translated

The `skos:definition` tag is used to carry supporting information which describes a concept in greater detail than is carried in the human readable title of the concept enclosed in `skos:prefLabel`. For concepts which require structured information to be carried with them, the contents of the `skos:definition` tag may be encoded as a JavaScript Object Notation (JSON) string.

```
<skos:definition xml:lang="en">
      {"country": "Italy",
      "platformclass": "self-propelled boat",
      "callsign": "IMNQ",
      "length": "19.05",
      "built": "1977",
      "notes": "Leased tugboat"}
</skos:definition>
```

5.4.4 *dc:identifier and skos:notation*

- Mandatory
- Number
    - One per concept

Version 1.X of the NVS uses the `skos:externalID` property to define the SeaDataNet Uniform Resource Name (URN) of a given concept. However, this property was deprecated in 2004, and the recommended replacement is the `dc:identifier` property from the Dublin Core metadata element set. `dc:identifier` is defined as "an unambiguous reference to the resource within a given context" which fits the usage of the property to declare the SeaDataNet URN or any other external identifiers given to a concept. The `skos:notation` tag is defined as a character string, not normally recognizable as a word or sequence of words in any human readable language, used to uniquely identify a concept within the scope of a concept scheme. This formal scope restriction is the reason that both `dc:identifier` and `skos:notation` tags are used for the same content.

```
<dc:identifier>
      SDN:C191::3_1_2_4
</dc:identifier>

<skos:notation>
      SDN:C191::3_1_2_4
</skos:notation>
```

5.4.5 *dc:date*

- Mandatory
- Number
    - One per concept

The Dublin Core metadata element `dc:date` allows the inclusion of an important point in the lifecycle of the concept. In this case we use the time and date of creation of this version of the concept.

```
<dc:date>
        2011-05-31T08:00:20.136+0000
</dc:date>
```

5.4.6 *owl:versionInfo*

- Mandatory
- Number
    - One per concept

The `owl:versionInfo` tag gives the version number of the concept included in the document.

```
<owl:versionInfo>
        2
</owl:versionInfo>
```

5.4.7 *skos:note*

- Mandatory
- Number
    - One per concept

The `skos:note` tag is designed to allow ancillary information about a SKOS concept. In the context of the payload documents under discussion here it is used to define a concept's publication status. A concept may be "accepted", "proposed" or "deprecated". This value is set respectively according to whether it has been accepted by the content governance body, it is being considered by the content governance body or the concept has been deprecated by the governance body.

```
<skos:note xml:lang="en">
        accepted
</skos:note>
```

5.4.8 *owl:deprecated*

- Mandatory
- Number
    - One per concept

The `owl:deprecated` tag encloses a Boolean value indicating if the concept has been deprecated ("false") or not ("true").

```
<owl:deprecated>
        true
</owl:deprecated>

<owl:deprecated>
        false
</owl:deprecated>
```


5.5 Multi-lingual provisioning

The encoding of which human language a SKOS annotation tag is written in should follow the World Wide Web Consortium guidelines of language encoding in XML. These recommendations state that the language tags from the Internet Assigned Numbers Authority (IANA) repository should be used with a hierarchy of

Primary language – extended language – script – region – variant – extension – private use

For the purposes of the NERC Vocabulary Server, a primary language encoding is the deepest the hierarchy need go.

```
<skos:prefLabel xml:lang="en">
        colour
</skos:prefLabel>

<skos:prefLabel xml:lang="fr">
        couleur
</skos:prefLabel>
```

5.6 Mappings

- Optional
- Number
    - Zero or many per concept or concept collection

Mappings, or semantic relations, indicate the links that a given concept has to another concept. Mappings in the original version of the SKOS specification could be broader, narrower, exact or close. However, in the latest version of the specification the concepts of exact and close matches have been superseded by the related tag.

Broader relations indicate that the current concept has a narrower definition than the concept to which it is related, narrower relations imply the inverse, and close matches imply that the two concepts are more loosely coupled.

Broader and narrower matches may also have the transitive property associated with them, which allows the use of a semantic inference engine. When a thesaurus is delivered through concept scheme in NVS2.0, the mappings internal to that thesaurus are supplied as transitive and those external to the thesaurus are supplied as non-transitive.

```
<skos:narrower rdf:resource="http://a/Term/Url" />
<skos:narrowerTransitive rdf:resource="http://a/Term/Url" />
<skos:broader rdf:resource="http://a/Term/Url" />
<skos:broaderTransitive rdf:resource="http://a/Term/Url" />
```

Loosely coupled concepts are tagged using the `skos:related` tag thus:

```
<skos:related rdf:resource="http://a/RelatedTerm/Url" />
```

Finally, synonyms (which were identified using the `skos:exactMatch` tag in version 1 of the NVS) are now specified using the Web Ontology Language's `sameAs` tag, thus:

```
<owl:sameAs rdf:resource="http://a/synonym/Url" />
```

It should be noted that `skos:exactMatch` has not been deprecated from the latest version of the SKOS specification, but has been specified only within the scope of concept schemes. As the concepts in NVS2.0 are registered to concept collections, `skos:exactMatch` cannot be used to signify synonymous relationships between concepts in NVS2.0.

6 Conclusions

6.1 How does the semantic resource benefit NETMAR?

The NETMAR project identified pilots for the semantically aware data and service discovery portal it is developing. Following a survey of user requirements for these pilots, three were identified as having a common semantic resource while a second resource was deployed for the fourth, the International Coastal Atlas Network (ICAN) [DD11]. This latter resource is an alternative approach, at a lower level of complexity, to semantics within ICAN to that developed by [WL08].

The main NETMAR semantic resources is designed around the faceted approach outlined above and describes: observed parameter; measuring instrument; observing platform; observing platform class; project; vertical coverage; and horizontal coverage. A concept map of this faceted resource is shown Figure 6-1.

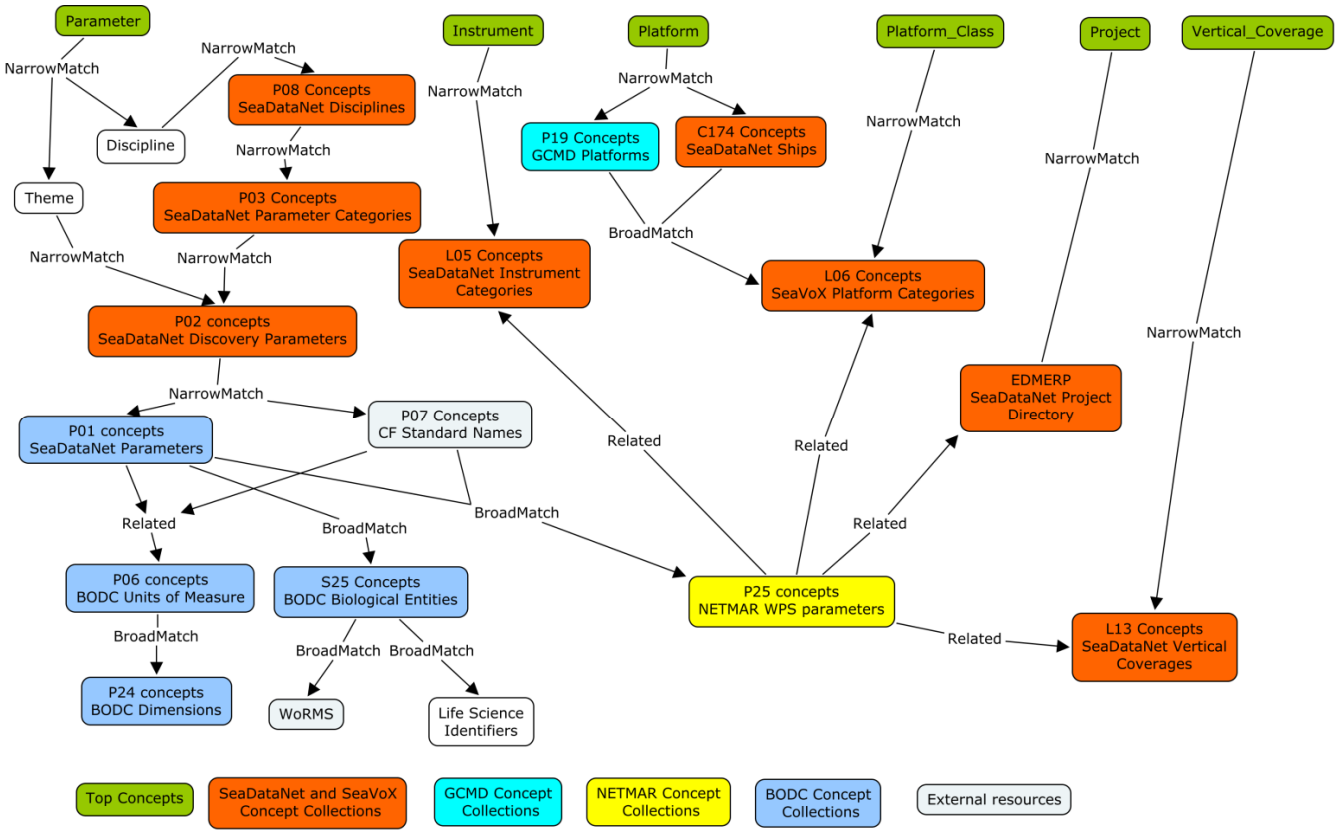

Figure 6-1 The NETMAR ocean semantic resource

The facets are interconnected at a narrow level of description, placing the NETMAR semantic resource into the taxonomy category of the semantic spectrum. Each of these facets is has an explicitly declared entry point at the broadest level of description, and the concepts below these entry points in the hierarchy can be inferred to inherit the facet to which they belong, therefore moving the NETMAR resource into the realm of informal ontology. Furthermore, the ocean biogeochemistry data ontology described by [CA11] defines the class “parameter” and “instrument” which are instantiated by concepts from the narrowest description level of the NETMAR semantic resource. This moves portions of the NETMAR resource into formal ontologies.

The ICAN community aims to make Coastal Web Atlases (CWAs) interoperable across jurisdictional boundaries and to create a smart search mediator for OGC Catalogue Services. The authors of local CWAs often use local vocabularies to describe their content and for interoperability a global resource mapped to these local resources must be defined. Earlier work in this area [GIR12] took the approach of using OWL ontologies, but there was little uptake following an initial concerted effort. The NETMAR

project has taken a SKOS based approach and, as a proof of concept, produced a mapping between the Oregon State CWA markup vocabularies; the Irish MIDA markup vocabularies and a global set of markup terms in the coastal erosion domain as defined by the ICAN community[3] (see Figure 6-2).

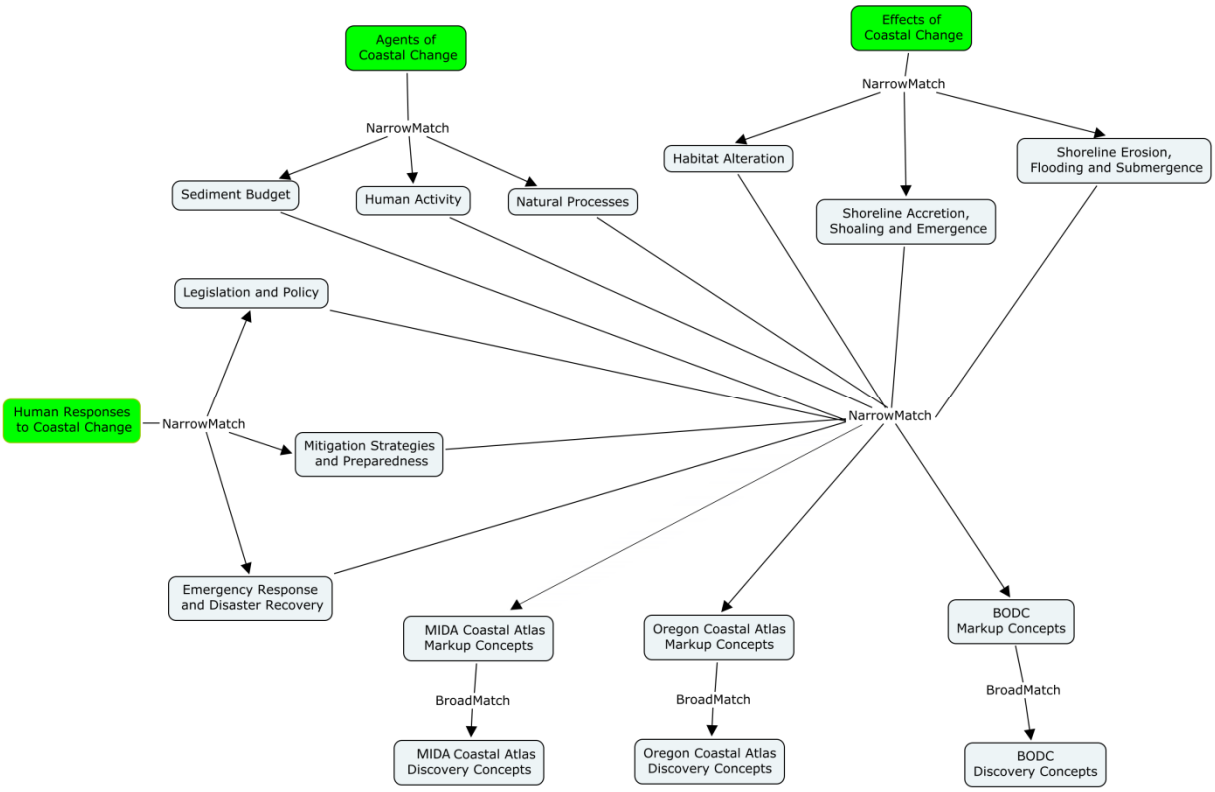

Figure 6-2 ICAN Coastal Erosion Thesaurus

6.1.1 *Client layer*

There are a number of client layer applications which have been built on top of the NETMAR semantic resource in order to allow human interaction with the resource, in addition to machine interaction. The simplest of these is a XSLT stylesheet over the SKOS concept schemes, which allows users to browse the SKOS concept hierarchy (Figure 6-3).

[3] http://vocab.nerc.ac.uk/scheme/ICANCOERO/

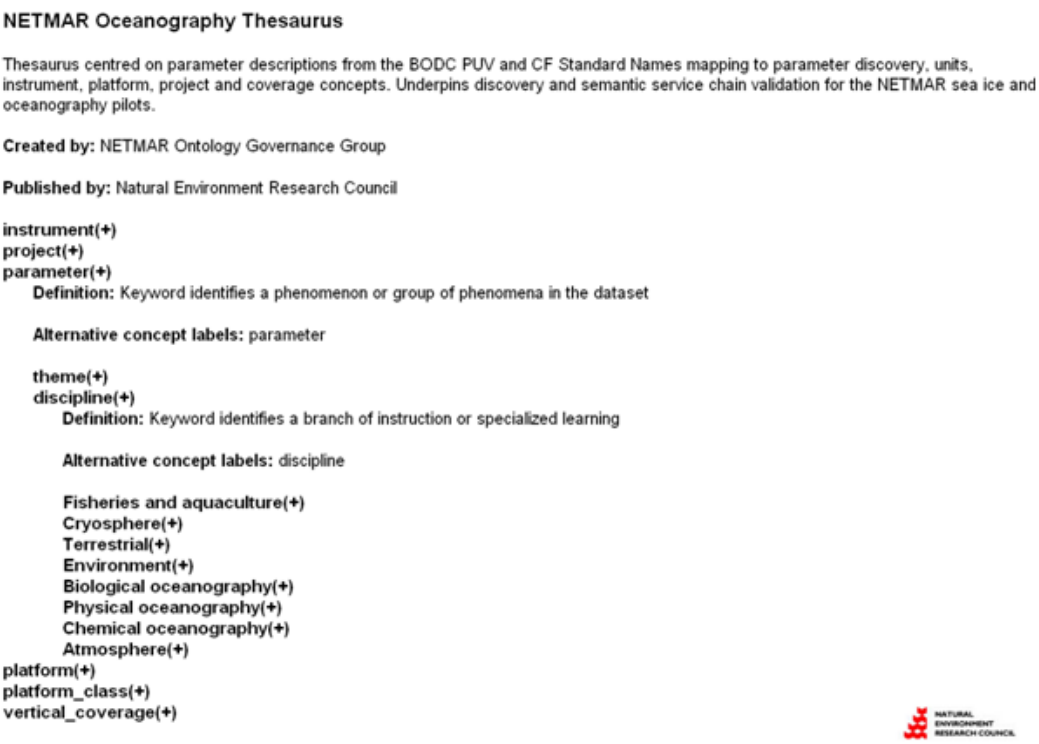

Figure 6-3 Browsing a SKOS concept scheme on NVS2.0

A more complex client is built on top of the NETMAR semantic resource and extends the OGC Catalogue Service[4] mediator presented by [LW08]. This client (Figure 6-4) allows the user to both browse and search the semantic resource. Once a term of interest has been located, the user may then discover and download datasets tagged with semantically related terms.

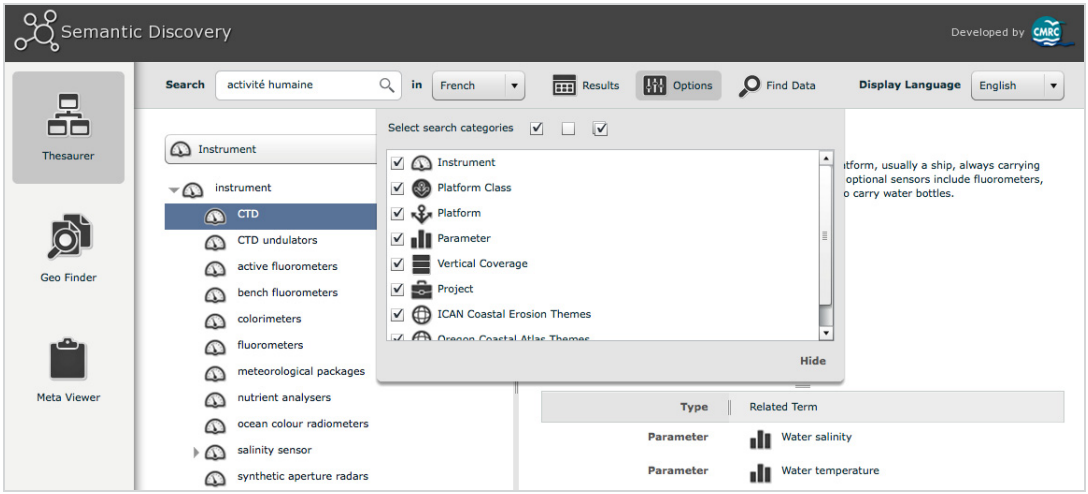

Figure 6-4 Screenshot of the Semantic Discovery Client showing concept search options in the Ontology Browser

6.1.2 *Service validation*

A main goal of the NETMAR project was to orchestrate the chaining of Open Geospatial Consortium Web Processing Services[5] (WPS) using a standard workflow management tool [JW11]. In order to

[4] http://www.opengeospatial.org/standards/cat

[5] http://www.opengeospatial.org/standards/wps

correctly chain services, the inputs must be of the basic physical dimension expected by the service. As the NETMAR semantic resource contains concepts related to units of measure[6] and aggregates of basic physical dimension[7], a service can be labelled with the expected input and also a dimensional analysis tool can be constructed. In its simplest form, this tooling takes the form a “traffic light” checker (Figure 6-5) which gives a green light for equivalent units, an amber light for units which require a conversion to be applied to achieve equivalence or a red light for incompatible units. This tool may be embedded inside a workflow editor or orchestration engine so that compatibility between WPSs may be established prior to executing a workflow.

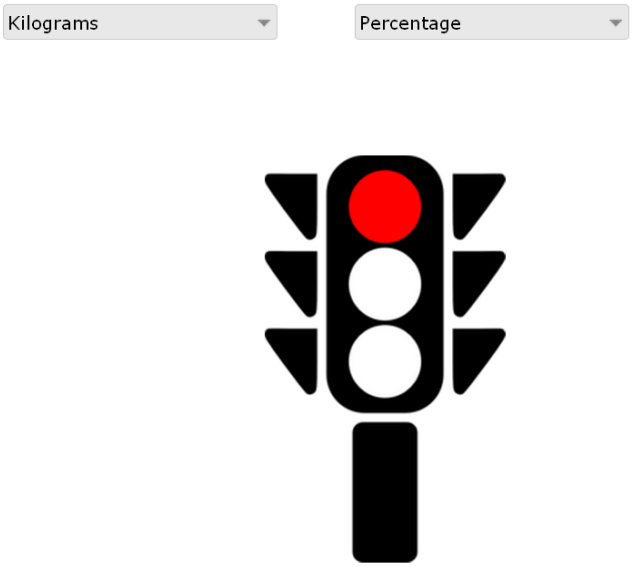

Figure 6-5 A traffic light system to illustrate semantic validation of input units to a web service

6.2 How does the semantic resource benefit the wider community?

The NVS2.0 is being used to support the EU FP-7 project SeaDataNet-II where the resources it serves are used in data and metadata markup and API calls have been embedded within the project’s NEMO software package[8], which is used to generate the correct data file format for ingestion to SeaDataNet, and will soon be incorporated within the Mikado metadata authoring tool[9].

Discussions are also ongoing with a number of other projects as to their use of NVS2.0: the INSPIRE programme are contemplating using for serving legally-binding codelists; the NERC Science Information Strategy will use NVS2.0 as the corporate node in a federated vocabulary architecture; the United States National Science Foundation Rolling Deck to Repository project and the Australian government’s Integrated Marine Observing System are making use of codelists served from NVS2.0; and avenues are being explored with instrument manufacturers to embed codes from NVS2.0 for semantic integration from the time of data collection.

The rapid uptake of NVS2.0 throughout the community is illustrated in the total monthly API calls since December 2011, as illustrated in Table 6-1.

[6] http://vocab.nerc.ac.uk/collection/P06/current/

[7] http://vocab.nerc.ac.uk/collection/P25/current/

[8] http://www.seadatanet.org/Standards-Software/Software/NEMO

[9] http://www.seadatanet.org/Standards-Software/Software/MIKADO

Table 6-1 NVS2.0 usage statistics, logged from 450 unique IP addresses

Month	Number of NVS2.0 API calls logged
December 2011	322
January 2012	739
February 2012	1602
March 2012	2352
April 2012	2380
May 2012	1863
June 2012	149021
July 2012 (to 30th)	181846

7 References

[CA11] Chandler, C., Allison, M., Groman, R., West, P., Zednick, S. and Meffei, A., 2011. Developing an ontology for ocean biogeochemistry data. Geophysical Research Abstracts 13, EGU2011-8578.

[DD11] O'Dea, E., Dwyer, E., Cummins, V. and Wright, D., 2011. Potentials and limitations of coastal web atlases. Journal of Coastal Conservation: Planning and Management 15(1), 1-21.

[GIR12] Graybeal, J., Isenor, A. and Rueda, C., 2012. Semantic mediation of vocabularies for ocean observing systems, Computers & Geosciences 40, 120-131.

[HHM10] Halpin, H., Hayes, P., McCusker, J. McGuinness, D. and Thompson, H., 2010. When owl:sameAs isn't the same: an analysis of identity in linked data, International Semantic Web Conference 2010 (1): 305-320

[JW11] de Jesus, J., Walker, P., Grant, M. and Groom, S., 2011. WPS orchestration using the Taverna workbench: The eScience approach. Computers & Geosciences, doi:10.1016/j.cageo.2011.11.011

[L10] Lowry, Roy, 2010. Establishment of NETMAR ontology governance body. NETMAR (Open service for marine environmental data) Deliverable 3.1. European Commission Information Society and Media Directorate-General Grant Agreement Number 24902.

[LL11] Lowry, Roy and Leadbetter, Adam, 2011. Prototype interlinked ontology resource. NETMAR (Open service for marine environmental data) Deliverable 3.6. European Commission Information Society and Media Directorate-General Grant Agreement Number 24902.

[LLC11] Lowry, Roy; Leadbetter, Adam and Clements, Olly, 2011. Strategy for ontology and tool development to fulfil NETMAR requirements. NETMAR (Open service for marine environmental data) Deliverable 3.4. European Commission Information Society and Media Directorate-General Grant Agreement Number 24902.

[LW08] Lassoued, Y., Wright D., Bermudez L., Boucelma, O., 2008. Ontology-based Mediation of OGC Catalogue Service for the Web: A Virtual Solution for Integrating Coastal Web Atlases. In: Proceedings of the 3rd International Conference on Data and Software Engineering ICSOFT 2008 , Berlin, Germany.

[PTS+11] Pladsen, Jan Ivar, et. al, 2011. D1.1 Use case specification and schedule for testing. NETMAR (Open service network for marine environmental data) Deliverable D1.1. European Commission Information Society and Media Directorate-General Grant Agreement Number 249024

[WL08] Wright, D., Lassoued, Y., Bermudez, L., Nyerges, T., Haddad, T. and Dwyer, N., 2008. Semantic mediation as a gateway to interoperability, with a case study of the International Coastal Atlas Network (ICAN). In: T. Cova, H. Miller, K. Beard, A. Frank, and M. Goodchild, eds. Geographic Information Science: 5th International Conference, GIScience 2008, Park City, Utah, September 23-26, Extended Abstracts, 201-218

