

Controlled Vocabularies in the Oceanographic domain

ALEXANDRA KOKKINAKI - NOC- BODC (UNITED KINGDOM)



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Overview

- Introduction
- Vocabularies
- NERC Vocabularies
- Features
- Application use and technologies of NERC vocabularies
- Vocabulary governance



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Metadata

“Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information.”

-- National Information Standards Organization

-- <http://www.niso.org/home/>



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

E.g. A digital image's metadata



- Name
 - Title
 - Type
 - Owner
 - Description
 - Tags
 - Colour
 - Time and date of last modification
 - Date of creation
- JPG
TIF
PNG
GIF



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

- Collect similar information
- Assist content authors in consistently tagging content
- Enable users to find the information they need by translating their language into the language of the information store



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Controlled vocabularies

- In information science controlled vocabulary is a carefully selected list of words and phrases, which are used to tag units of information (document or work) so that they may be more easily retrieved by a search
- They are important:
 - capture expertise in agreed, well-defined descriptions
 - enable population of a given field in a metadata model with standardised unambiguous terms
 - enable records to be interpreted by computers
 - promote consistency and interoperability

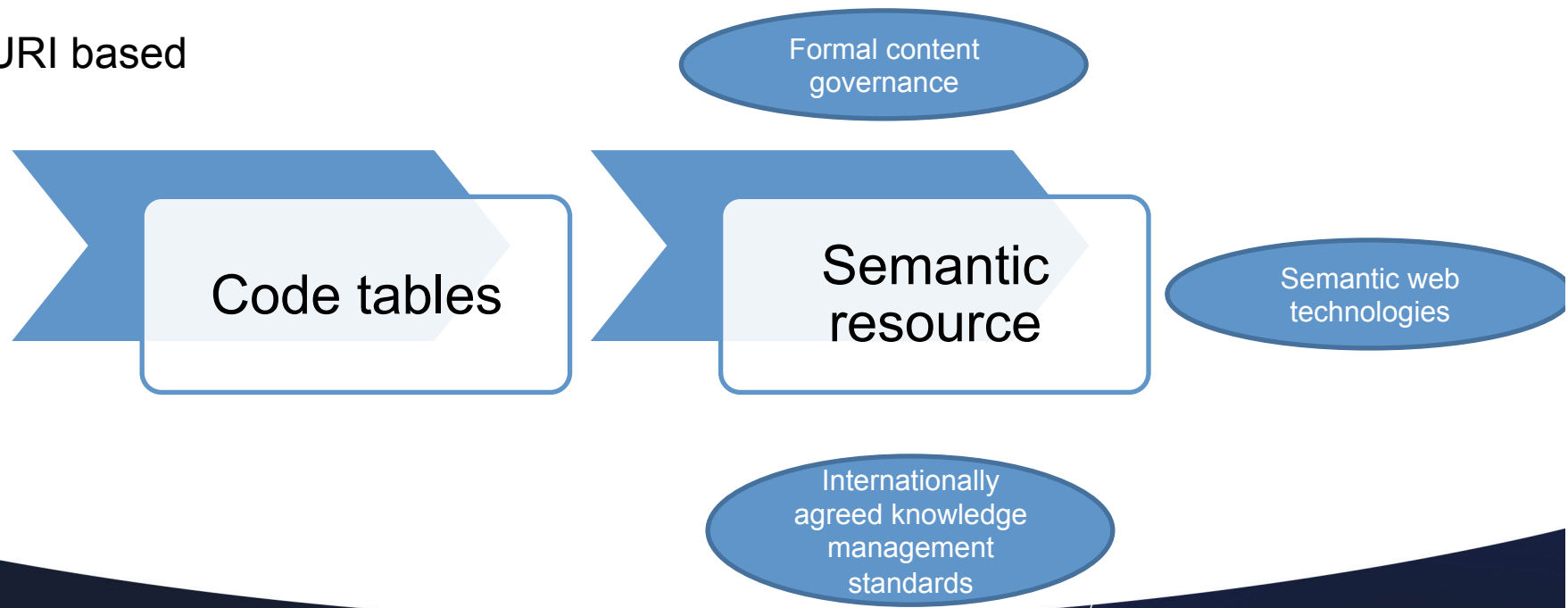


Controlled vocabularies in BODC

Resource Description Framework (RDF)

Simple Knowledge Organisation System (SKOS)

URI based



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Uniform Resource Identifier

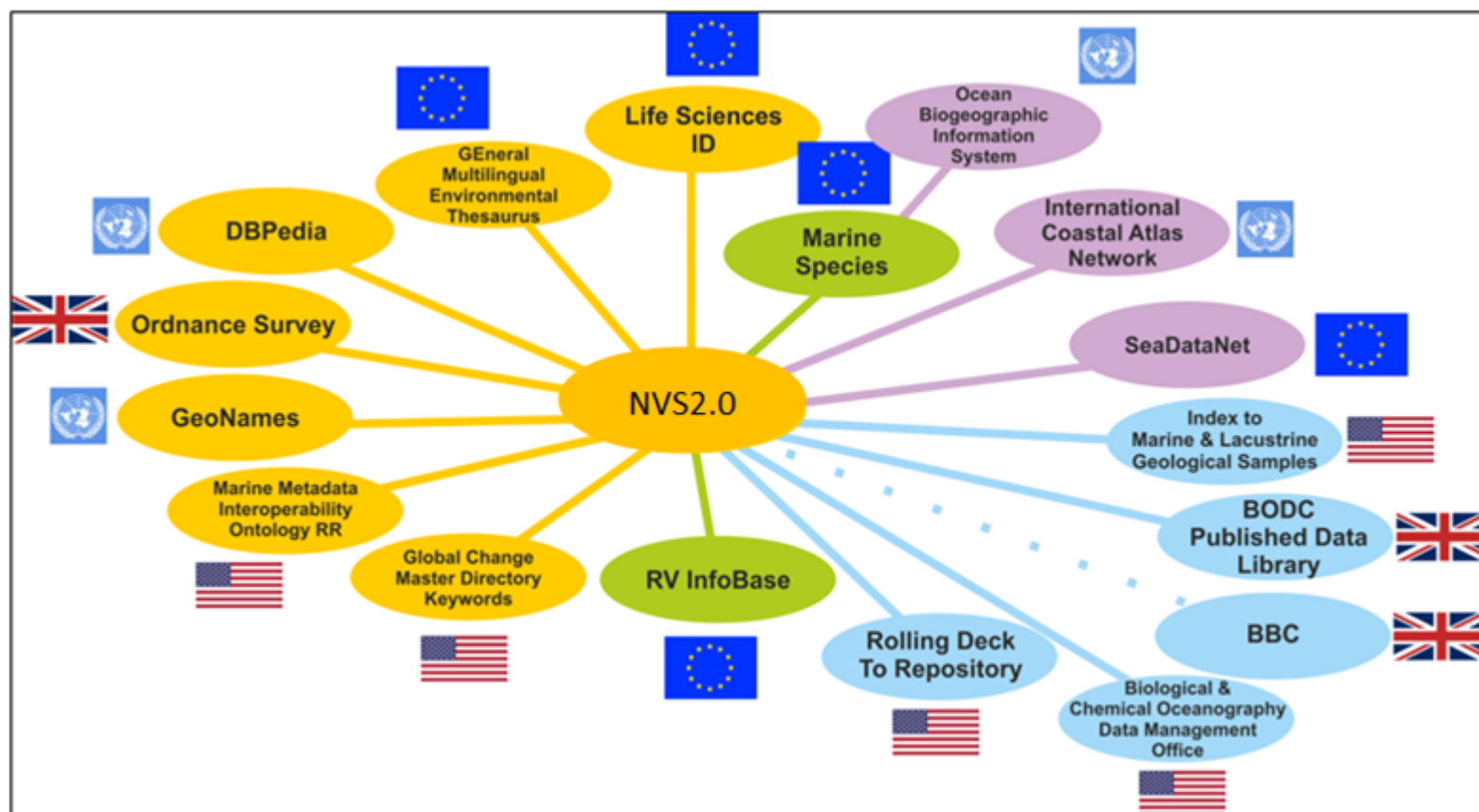
- Documents returned are fully standards compliant
- RDF/XML as specified by W3C
- Conform to Linked Data standards
- Has allowed worldwide adoption of the vocabularies







National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT



	RDF Vocabulary resources
	Non RDF Vocabulary resources
	Non RDF datasets
	RDF Datasets

Vocabulary checklist

It is best practice to use or extend an existing vocabulary before creating a new vocabulary

- ensure vocabularies you use are published by a trusted group or organization;
- ensure vocabularies have permanent URIs;
- and confirm the versioning policy.

Reference: Best Practices for Publishing Linked Data [<https://www.w3.org/TR/ld-bp/#VOCABULARIES>]



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Extended Checklist

- **Vocabularies MUST be documented**
- **Vocabularies SHOULD be self-descriptive**
- **Vocabularies SHOULD be described in more than one language**
- **Vocabularies SHOULD be used by other datasets**
- **Vocabularies SHOULD be accessible for a long period**
- **Vocabularies SHOULD be published by a trusted group or organization**
- **Vocabularies SHOULD have persistent URLs**
- **Vocabularies SHOULD provide a versioning policy**

Reference: [<https://www.w3.org/TR/ld-bp/#VOCABULARIES>]



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

NERC Vocabulary Server v2 (NVS2)

Concept: Fundamental entity in Simple Knowledge Organisation System (SKOS)

- an atomic piece of knowledge.
- and consists of the following attributes:
 - **An identifier:** alphanumeric string, ideally random and devoid of semantic content
 - **A preferred label:** Human readable string
 - **An alternative label:** Abbreviation mostly used for software, where space is a problem
 - **A definition :** Clear and unambiguous statement of the semantics associated with the concept
 - **A status:** accepted / deprecated
 - **A version :** indicates the number of times the concept attributes have been changed since the concept was first created. The number of changes is the version number minus 1.
 - **Publication date:** the timestamp of the last change to the concept attributes.



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Browser tabs: vocab.nerc.ac.uk/collection, URL vs URI vs URN: The Sc, How to Set Up Your Own, W3 URIs, URLs, and URNs: Cla, W3 URIs, URLs, and

Address bar: vocab.nerc.ac.uk/collection/S25/current/BE000006/

Bookmarks: Apps, ★ Bookmarks, Projects, BODCInternal, RDF study, PHD, Vocabularies, rdf_staff, medepad, Suggested Sites, Im

-- Abra alba (ITIS: 81306: WoRMS 141433) --

URI	http://vocab.nerc.ac.uk/collection/S25/current/BE000006/
Identifier ()	SDN:S25::BE000006
Preferred label (en)	Abra alba (ITIS: 81306: WoRMS 141433)
Alternative label ()	
Definition ()	
Version Info ()	1
Deprecated()	false
Same as	http://www.marinespecies.org/aphia.php?p=taxdetails&id=141433
Same as	http://lsid.tdwg.org/urn:lsid:marinespecies.org:taxname:141433
Narrower	http://vocab.nerc.ac.uk/collection/P01/current/ZBAF0006/
Narrower	http://vocab.nerc.ac.uk/collection/P01/current/LTZBM001/
Narrower	http://vocab.nerc.ac.uk/collection/P01/current/N601S10Z/
Narrower	http://vocab.nerc.ac.uk/collection/P01/current/ASHLM050/
Date ()	2010-09-16 12:58:56.0

Vocabularies SHOULD be self-descriptive

Concept aggregations

Two types of concept aggregations:

- aggregation of the concepts that may be used to populate a specified field in a data or metadata model and has been mapped to the **'concept collection'** in SKOS
- This may be thought of as a 'concept soup' interlinked by a simple set of semantic relationships (broader, narrower, same and related). A small number of these concepts, known as 'top concepts' are declared as entry points. These thesauri have been mapped to **'concept schemes'** in SKOS.



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

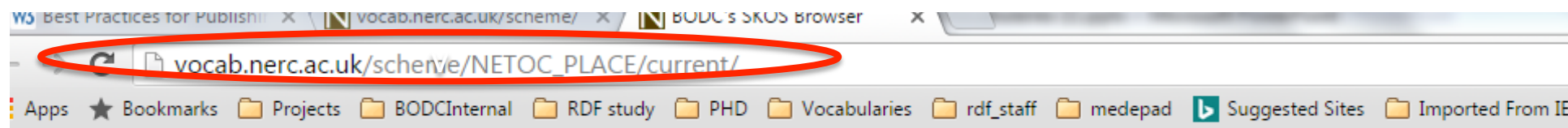
noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

title-:	SeaDataNet Parameter Disciplines
alternative-:	SeaDataNet Disciplines
description-:	Terms used to classify SeaDataNet Agreed Parameter Groups to provide topic/theme discovery interface.
date-:	2008-05-01 02:00:03.0
publisher-:	Natural Environment Research Council
creator-:	SeaDataNet
versionInfo-:	3
RE_RegisterManager:	British Oceanographic Data Centre
RE_RegisterOwner:	SeaDataNet
comment-:	Governance for vocabularies used in the EU SeaDataNet project implemented as cons members of the SeaDataNet Technical Task Team

↑ -- Administration and dimensions --

URI	http://vocab.nerc.ac.uk/collection/P08/current/DS07/
Identifier ()	SDN:P08::DS07
Preferred label (en)	Administration and dimensions
Alternative label ()	Administration and dimensions
Version Info ()	1
Definition (en)	Parameters related to spatial and temporal co-ordinates, entity referencing (eg record numbering
Deprecated ()	false
Same as ()	http://vocab.nerc.ac.uk/collection/P03/current/Z005/
Broader	http://vocab.nerc.ac.uk/collection/L19/current/001/
Narrower	http://vocab.nerc.ac.uk/collection/W01/current/035/
Date ()	2005-03-10 14:31:52.0



NETMAR Sea Area Gazetteer

Thesaurus that is the subset of the C19 sea area gazetteer covering the region of interest of the NETMAR project.

Created by: NETMAR Ontology Governance Group

Published by: Natural Environment Research Council

Arctic Ocean(+)

Northeast Atlantic Ocean (40W)(+)

Machine readable

Discoverable

Interoperable

NVS2 Vocabulary Server

- Concept URIs
- Description based in Resource Description Framework (RDF) and (Simple Knowledge Organization System) SKOS

<http://vocab.nerc.ac.uk/collection/{XXX}/current/>

definition

- Vocabularies are Collections

<http://vocab.nerc.ac.uk/collection/{XXX}/current/{XXX001}>

- Each Collection consists of many concepts
- A concept is a term that belongs to a collection

preferred label

alternative label

<http://vocab.nerc.ac.uk/collection/P06/current/MMPL/>

Mappings

Interlink data with
a rich and fast-
growing network of
other data sources

-206 Vocabularies
-161.000 terms



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Mappings

- Internal, External
 - Narrower
 - Broader
 - Same As
 - Related

Concepts from different controlled vocabularies may be connected using simple mapping relationships:

- Bacillariophyceae same-as diatoms IPTS68
- temperature related ITS90 temperature
- Nutrients in rivers related nitrate in water bodies
- Salinity is-narrower-than physical oceanography
- Physical oceanography is-broader-than salinity –



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Mail - Koki x Drake - Ho x Workshop x vocab.nerc x BODC's SK x vocab.nerc x Beethov x sdn2_wp8 x Nat

← → ↻ vocab.nerc.ac.uk/collection/P08/current/DS09/

Apps ★ Bookmarks Projects BODCInternal RDF study PHD Vocabularies rdf_staff medepad Suggested Sites Imported From

↑ -- Cryosphere --

URI	http://vocab.nerc.ac.uk/collection/P08/current/DS09/
Identifier ()	SDN:P08::DS09
Preferred label (en)	Cryosphere
Alternative label (en)	Cryosphere
Definition (en)	The cryosphere science domain, including ice on both land and sea
Version Info ()	1
Deprecated()	false
Same as	http://vocab.nerc.ac.uk/collection/P03/current/M015/
Broader	http://vocab.nerc.ac.uk/collection/L19/current/001/
Narrower	http://vocab.nerc.ac.uk/collection/C77/current/M04/
Narrower	http://vocab.nerc.ac.uk/collection/W01/current/035/
Related	http://vocab.nerc.ac.uk/collection/P03/current/H003/
Related	http://www.bbc.co.uk/programmes/b00818zy
Date ()	2006-08-09 09:59:42.0

NVS2 Versioning

- In NVS V2, each concept carries its own version number and publication date
- Each time any concept attribute is modified a mechanism archives the record as it was, increments the version number and sets the last modification/publication timestamp to the current time.
- Changes are only for broadening the meaning of the concept
- Any change triggers a new version of the concept
- If concept changes version then the Collection changes version too.



Deletion/Deprecation

- NVS concepts are not deleted
- Concepts are deprecated and replaced by other concepts

↑ -- Concentration of alkanols (alcohols), phenols and ethers in sediment --

URI	http://vocab.nerc.ac.uk/collection/P02/current/CA11/
Identifier ()	SDN:P02::CA11
Preferred label (en)	Concentration of alkanols (alcohols), phenols and ethers in sediment
Alternative label (en)	Sed_Alkanol+Phenol+Ethers
Definition (en)	Concentration of alkanols, phenols and ethers per unit mass or unit volume of sediment (material collected by a corer including fluff)
Version Info ()	2
Deprecated()	true
ReplacedBy	http://vocab.nerc.ac.uk/collection/P02/current/CASD/
Broader	http://vocab.nerc.ac.uk/collection/P02/current/CASD/
Broader	http://vocab.nerc.ac.uk/collection/L19/current/005/
Date ()	2011-02-24 16:52:36.0



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

↑ -- Natural Processes --

URI	http://vocab.nerc.ac.uk/collection/A01/current/Natural_Processes/
Identifier ()	SDN:A01::Natural_Processes
Preferred label (en)	Natural Processes
Preferred label (fr)	Processus naturels
Preferred label (no)	Naturlige prosesser
Preferred label (pt)	Processos naturais
Preferred label (es)	Procesos naturales
Alternative label ()	
Definition (en)	Actions and processes that result from the forces of nature such as weather and tides.
Definition (fr)	Non disponible
Definition (no)	Ikke tilgjengelig
Definition (pt)	não disponível
Definition (es)	no disponible
Version Info ()	1
Deprecated()	false
Broader	http://vocab.nerc.ac.uk/collection/A01/current/Agents_of_Coastal_Change/
Narrower	http://vocab.nerc.ac.uk/collection/A02/current/Rain/
Narrower	http://vocab.nerc.ac.uk/collection/A02/current/AverageMaximumTemperature/
Narrower	http://vocab.nerc.ac.uk/collection/A02/current/AverageMinimumTemperature/
Narrower	http://vocab.nerc.ac.uk/collection/A02/current/AveragePrecipitation/
Narrower	http://vocab.nerc.ac.uk/collection/A02/current/EarthSurfaceGeology/
Narrower	http://vocab.nerc.ac.uk/collection/A02/current/Erosion/
Narrower	http://vocab.nerc.ac.uk/collection/A02/current/Flood/
Narrower	http://vocab.nerc.ac.uk/collection/A02/current/Landslide/
Narrower	http://vocab.nerc.ac.uk/collection/A02/current/SeaLevelRise/
Narrower	http://vocab.nerc.ac.uk/collection/A02/current/Tsunami/

Semantic modeling of vocabularies

There are a few BODC vocabularies that are built from an underlying semantic model, which use other BODC vocabularies as the building blocks or components.

<http://vocab.nerc.ac.uk/collection/P01/current/WB000284/>

Concentration of lipids per unit wet weight of biota {Mytilus galloprovincialis (ITIS: 79456: WoRMS 140481) [Sex: male Subcomponent: gill]}

- ✓ **Measurement property** Concentration
- ✓ **Measurement statistical qualifier**
- ✓ **Chemical substance** lipids
- ✓ **Measurement-matrix relationship** per unit wet weight of
- ✓ **Matrix** biota
- ✓ **Biological entity** {Mytilus galloprovincialis (ITIS: 79456: WoRMS 140481) [Sex: male Subcomponent: gill]}



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

NVS2 application and use







**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Instruments

Instrument	code list value
<div>     </div>	

Category code

- 62 - aeroplane
- 54 - airship
- 96 - amphibious crawler
- 95 - amphibious vehicle
- 69 - autogyro
- 3B - autonomous surface water vehicle
- 25 - autonomous underwater vehicle**
- 13 - beach/intertidal zone structure
- 50 - buoyant aircraft
- 75 - cetacean
- 17 - coastal structure
- 90 - cryosphere
- 72 - diver
- 94 - drift ice
- 24 - drifting manned submersible
- 44 - drifting subsurface float
- 46 - drifting subsurface profiling float
- 42 - drifting surface float
- 3D - drillship
- 9A - DUKW
- 76 - fish

Name *

Code L06 *

Quantity Unit

HE
NT

Uses

- Searching existing metadata records.
 - Discovery interfaces
 - [BODC discovery interface](#)
 - [SeaDataNet discovery interface](#)
 - R2R (Rolling deck to repository)

NEMO: Conversion of any ASCII format to the SeaDataNet ODV4 ASCII format

- netCDF files, ODV
- Maps data column names to vocabulary concepts

Change_Vocab_V1toV2: Conversion of SeaDataNet ODV or MEDATLAS with V1 vocabs to NVS V2.0



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Access-Technologies-Publication-Tools



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

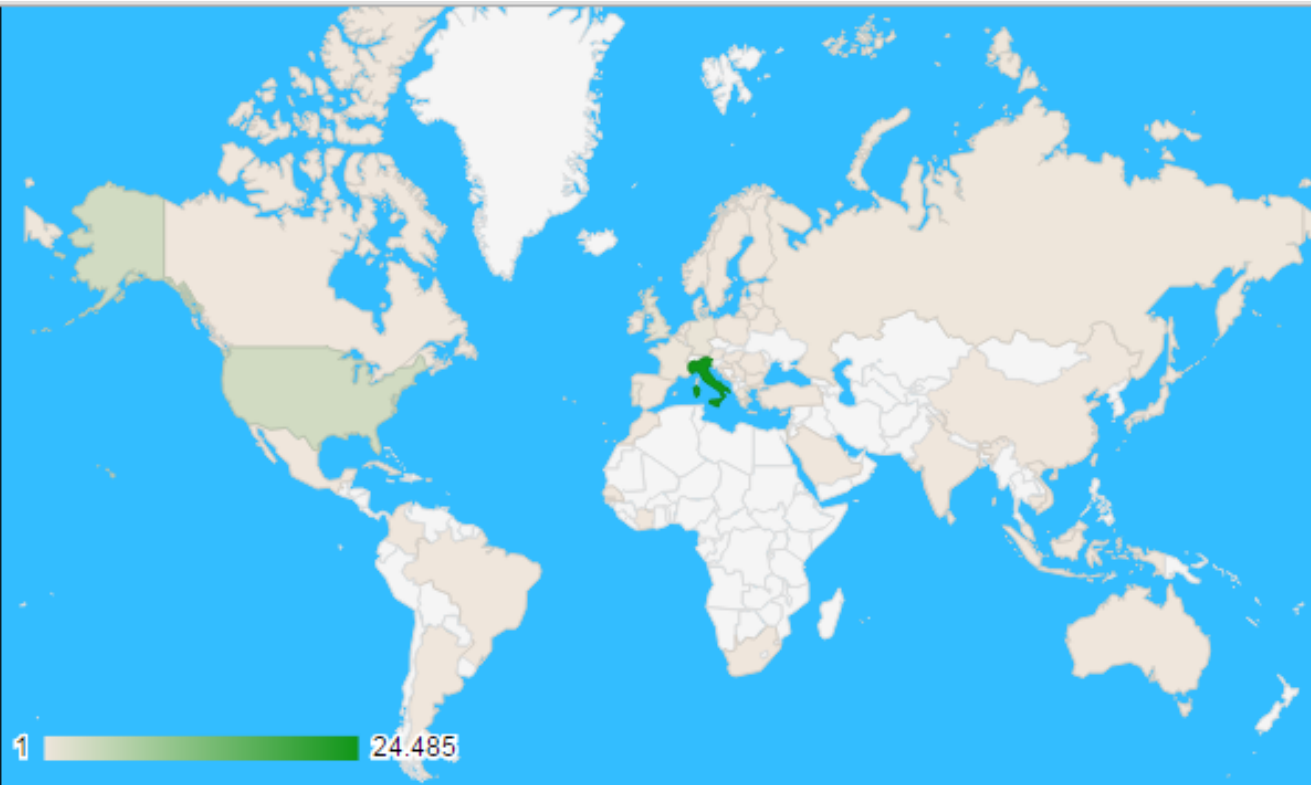
noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Publication and access

- RESTFul interface
- URIs
 - <http://vocab.nerc.ac.uk/collection/>
- SOAP web service
- Methods
 - getCollections
- SPARQL endpoint
- Queries
 - PREFIX dc:<<http://purl.org/dc/elements/1.1/>>
PREFIX skos:<<http://www.w3.org/2004/02/skos/core#>>
PREFIX rdf:<<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>
SELECT ?s ?p WHERE {
?s rdf:type skos:Collection . ?s dc:title ?p
}

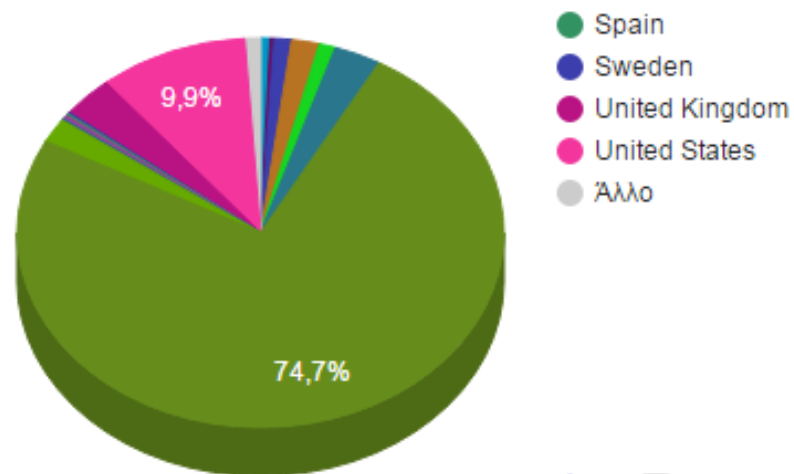


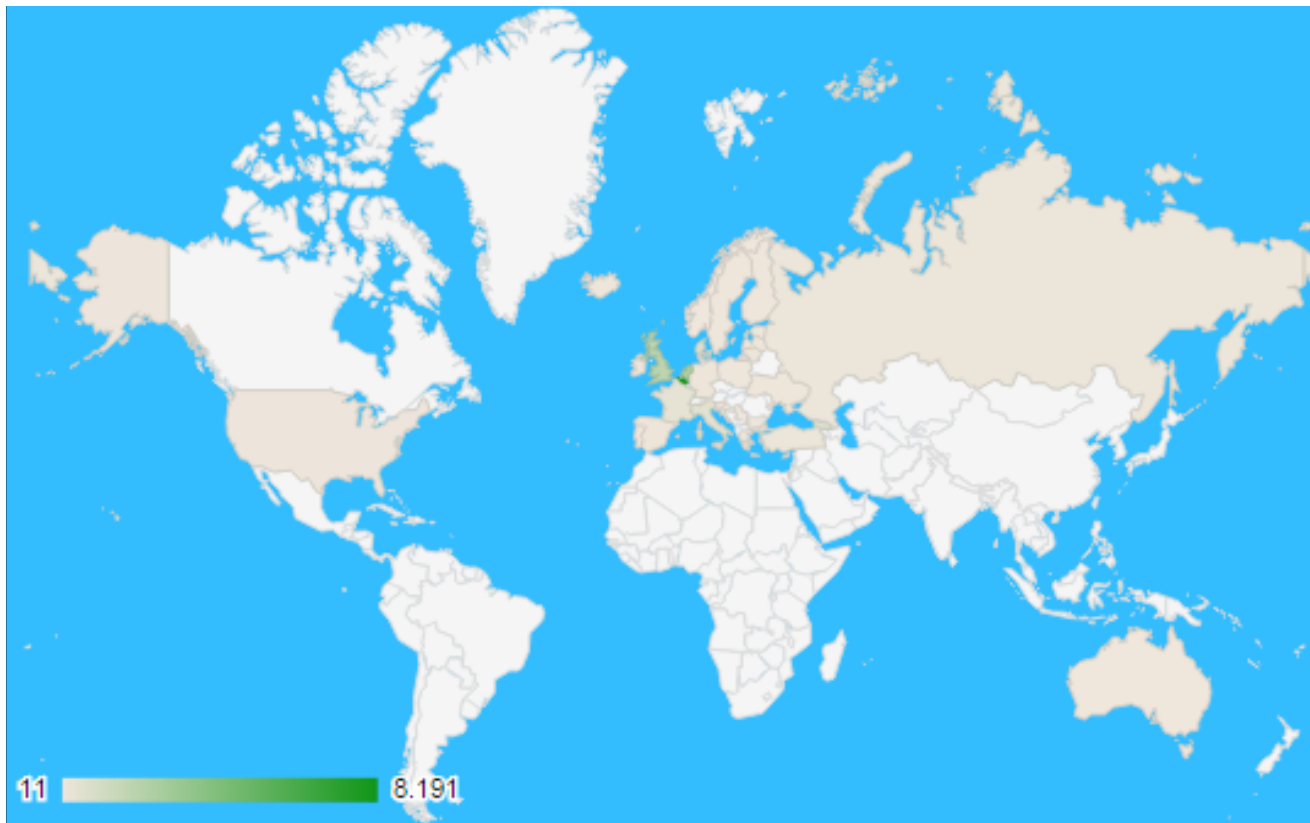


1	Italy	24.485
2	United States	3.250
3	United Kingdom	1.129
4	Israel	1.023
5	Portugal	666
6	Germany	606
7	France	377
8	Greece	359
9	Belgium	210
10	Russia	106
11	Denmark	88
12	Spain	54
13	Romania	53
14	Sweden	47
15	Netherlands	44
16	Taiwan	41
17	Senegal	36
18	Australia	35
19	Lithuania	16
20	Brazil	14

◀ ▶ 1 2 3

RESTful hits

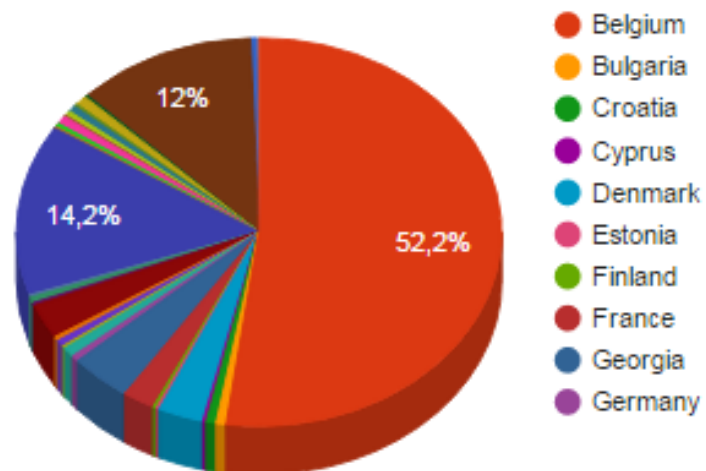


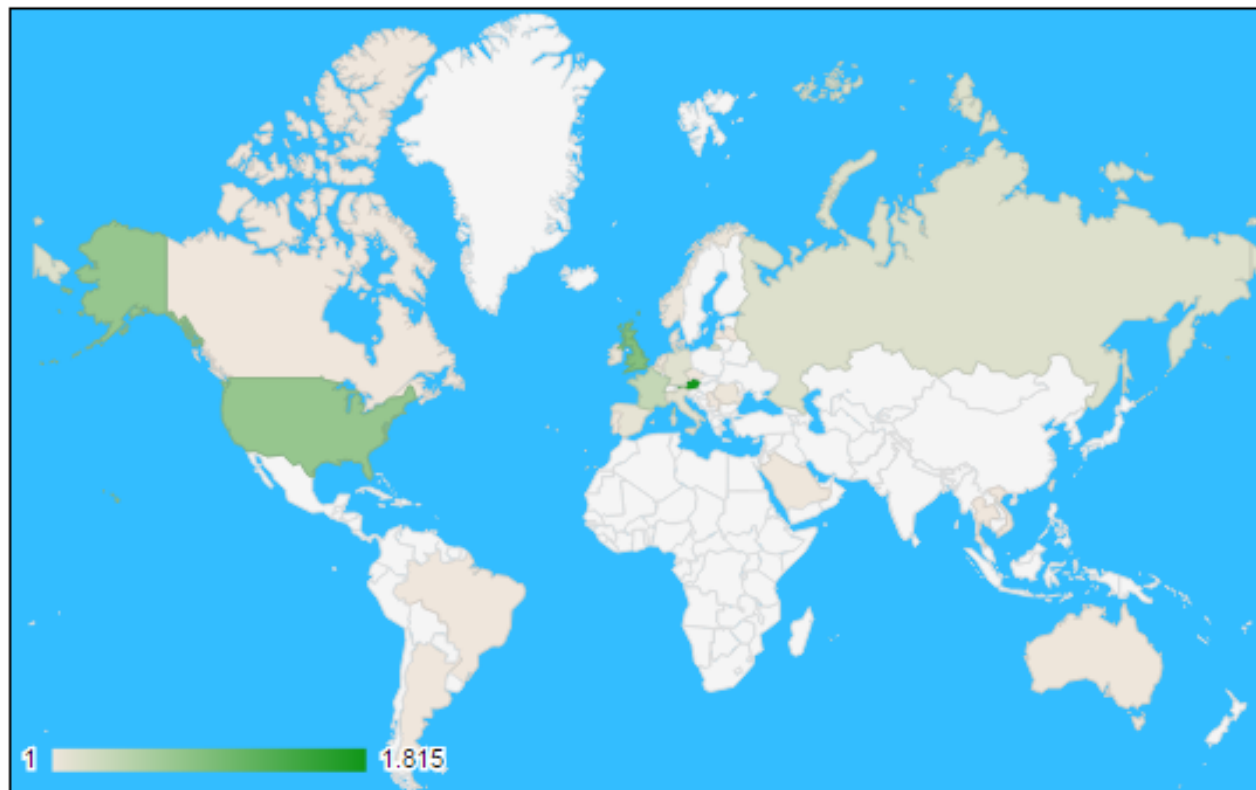


	Country	SOAP hits ▼
1	Belgium	8.191
2	Netherlands	2.234
3	United Kingdom	1.878
4	Georgia	647
5	Denmark	475
6	Italy	457
7	France	336
8	Greece	129
9	Turkey	125
10	Russia	107
11	Bulgaria	105
12	Croatia	102
13	Ireland	85
14	Spain	83
15	Germany	80
16	Lithuania	76
17	United States	69
18	Slovenia	64
19	Israel	58
20	Poland	47

◀ ▶ 1 2

SOAP hits

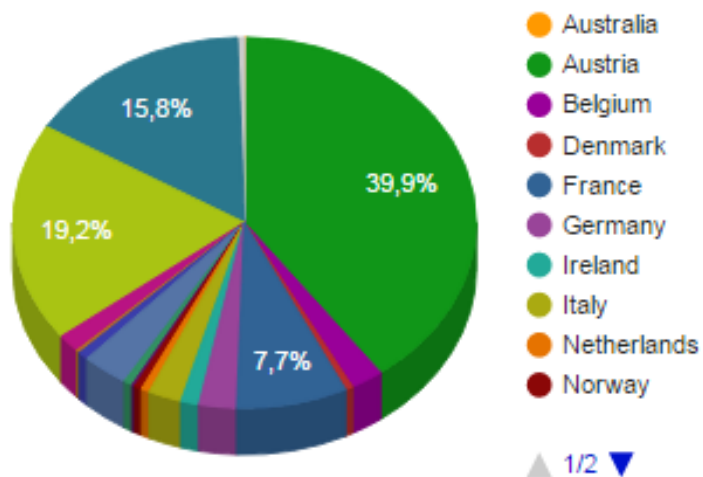




	Country	SPARQL hits ▼
1	Austria	1.815
2	United Kingdom	873
3	United States	721
4	France	351
5	Russia	147
6	Germany	119
7	Italy	106
8	Belgium	104
9	Spain	73
10	Ireland	58
11	Réunion	33
12	Romania	28
13	Netherlands	27
14	Denmark	24
15	Norway	23
16	Saudi Arabia	12
17	Portugal	10
18	Australia	7
19	Argentina	4
20	Brazil	4

◀ ▶ 1 2

SPARQL hits



TOOLS

- [Vocab Editor](#)
- [NVS2 search](#)
- [Vocab Builder](#)



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Technologies

- Oracle: backend office
- Java as programming language
- Jena TDB: TDB is a component of [Jena](#) for RDF storage and query.
- Jena Fuseki: SPARQL server



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Vocabulary Governance



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Types of Governance

Content Governance

- Determines the concepts to be included in controlled vocabularies and their labels (URIs, text terms and definitions)

Technical Governance

- Storage, versioning and serving of controlled vocabularies



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Types of Governance

Content Governance

- The history of oceanographic data management is littered with examples of poor vocabulary content governance
- Decisions made by individuals – even students
- Decisions taken with inadequate thought
- Lazy governance (saying “yes” to all requests)



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Types of Governance

Content Governance Models

- Benign dictator
 - Can work, but doesn't scale and what happens if the dictator goes under a bus?
- Moderated caucus
 - Commonly adopted based on e-mail list servers and can work well
 - Risks associated with the model
 - Caucus apathy
 - Caucus intransigence or moderator too weak
 - Moderator overload (tooling can help)
 - Moderator too strong
- Formal committee
 - Very slow



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Types of Governance

Technical Governance

- The history of oceanographic data management is littered with examples of poor vocabulary technical governance
- Multiple copies
 - One site served two versions of a list
 - » Both were different
 - » Neither was the master
- Weak management
 - Scrappy files scattered over several FTP sites
 - Local copies taken
 - Like Galapagos finches these evolve becoming similar but significantly different
- No version labelling or timestamps
 - Evolving local copies had no idea how they related to their origin



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

NVS 2 Governance Model

Each collection is allocated a content governance authority
These include:

- BODC internal governance
- SeaDataNet Technical Task Team
- SeaVoX (Intergovernmental Oceanographic Commission caucus)
- ICES
- CF Standard Names Committee
- NASA Global Change Master Directory
- International Hydrographic Bureau
- ISO



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

SeaVoX

How SeaVoX operates

The aim of the mailing list is to reach consensus through discussion on the topics raised.

- A week of silence on a topic is taken as 'subject closed' and the decisions made on that topic will be implemented.
- If consensus cannot be reached then the decision becomes the responsibility of the MarineXML Steering Group Chair.
- Once loaded, vocabulary terms cannot be deleted and changes to term definitions can only broaden meanings.



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Governance Best Practice

Content Governance Do's and Don'ts

- Do process requests for change as quickly as possible
- Do ensure that change requests don't redefine the collection
- Do establish managed deprecation
- Do insist on definitions for terms
- Don't change the semantics mapped to a URN
- Don't delete concepts



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Governance Best Practice

Technical governance Do's and Don'ts

- Do provide on-line access to the latest version
- Do provide URIs to concepts and collections
- Do provide access to previous versions
- Do empower content governance with tools
- Do provide both clients and service APIs
- Don't allow URNs to change
- Don't make unnecessary change to URLs



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Challenges

- http URIs (are they going to be there forever?)
- Persistent URIs for Vocabularies
- Collaborative environments
- Vocabulary Mappings and Alignments
- Multilingualism
- Reasoning/Inferencing using vocabularies



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

