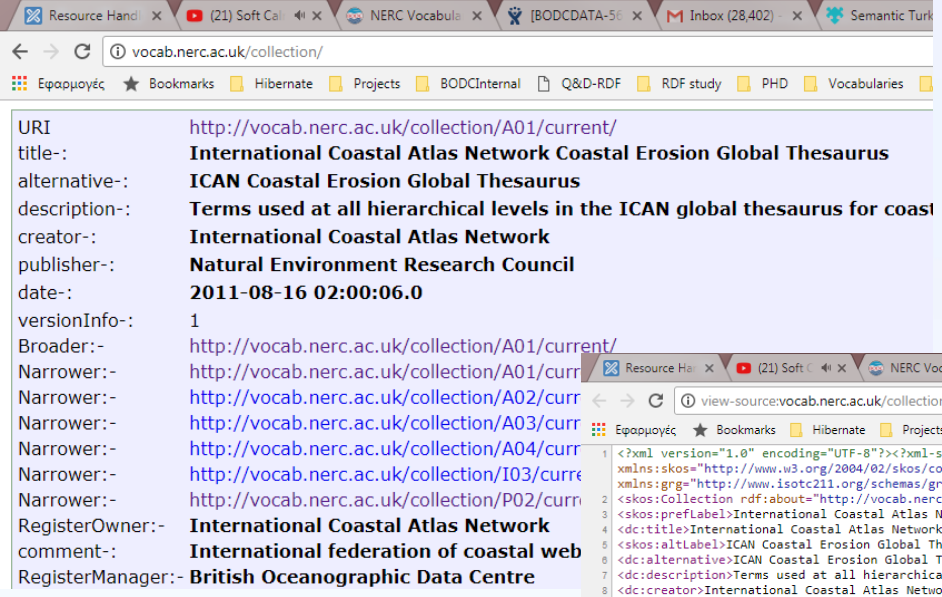
Common Vocabulary Services

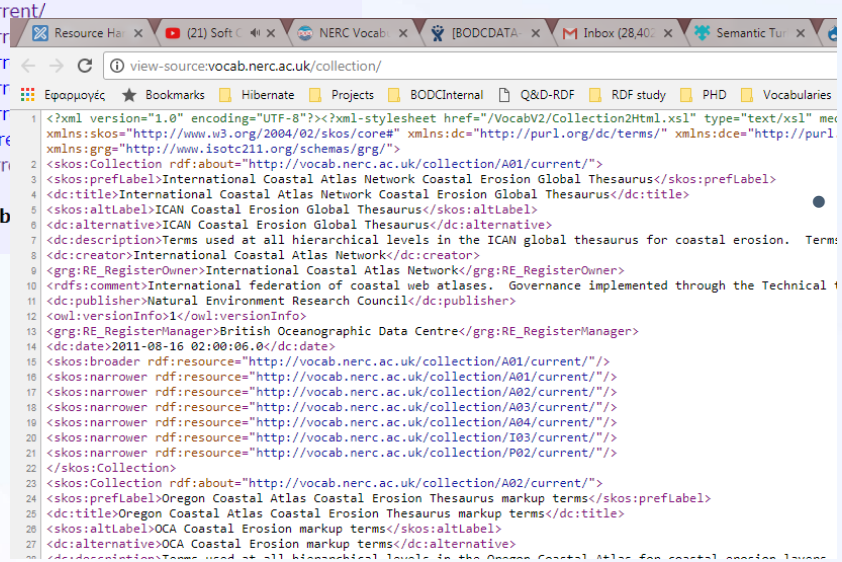
Alexandra Kokkinaki (alexk@bodc.ac.uk)

British Oceanographic Data Centre, National Oceanography Centre,
National Oceanography Centre, UK

Controlled vocabularies

- In information science controlled vocabularies are carefully selected lists 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.
- Roles:
  - capture expertise in agreed, well-defined, standardised descriptions
  - enable population of a given field in a metadata model with standardised unambiguous terms
  - promote consistency, harmonization and interoperability
  - enhance discoverability





- **Concepts**
  - 159311 valid concepts including:
  - P01 parameter codes: 37368
  - L22 instrument codes :1198
- **Mappings**
  - 617777 internal mappings between concepts
  - 21059 mappings of NVS concepts to external resources
- **Schemes**
  - 21 SKOS schemes

- **Collections**
  - 239 vocabulary collections
  - 71 are owned and governed by BODC
  - 55 are under SDC/EMODnet/SeaVox/SWE content governance
  - Remainder (113) are owned by 25 different governing bodies

NERC Vocabulary Server

sdn-userdesk@seadatanet.org – www.seadatanet.org

Vocab Management Group

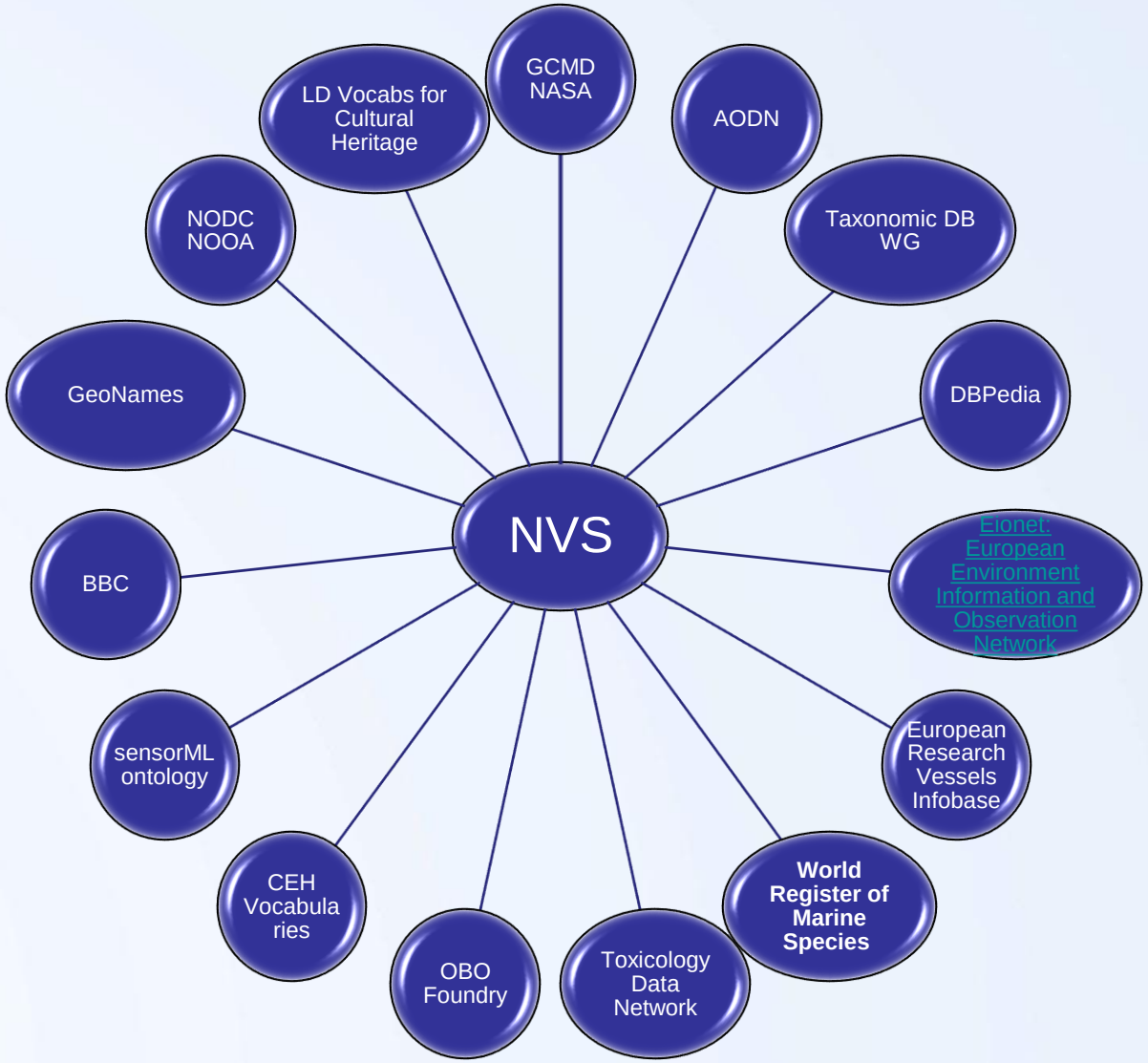
- Technical lead: **Alexandra Kokkinaki**
- Content lead and overall group coordination: **Gwen Moncoiffé**
- Instruments (L22 and L05) and SWE Vocab lead: **Louise Darroch**
- Vocab schema management: **Sean Gaffney**
- Consultant: **Roy Lowry**
- Overlap of expertise and training of new members to ensure every domain is covered by at least 2 members of staff:
- Chemical domain: Gwen Moncoiffé and Lou Darroch
- Biological domain: Gwen Moncoiffé and **Arwen Bargery**
- Physical domain: **Mark Hebden** and **Dani Edgar**
- Instruments and platforms: Lou Darroch, **Vi Paba** and Dani Edgar
- Gatekeepers: Gwen and Sean
- New requests through enquiries@bodc.ac.uk

NVS Facts

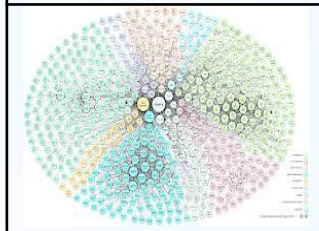
- NVS is a SKOS-vocabulary
- Linked Data
  - Unique http URIs

    http://vocab.nerc.ac.uk/collection/XXX/current/
    http://vocab.nerc.ac.uk/collection/XXX/current/YYYYY/
    http://vocab.nerc.ac.uk/scheme/

  - Content negotiation
    - RDF/XML
    - Human readable interface
  - Mappings
    - Internal (NVS 2 NVS)
    - External (NVS to ext Vocabularies)



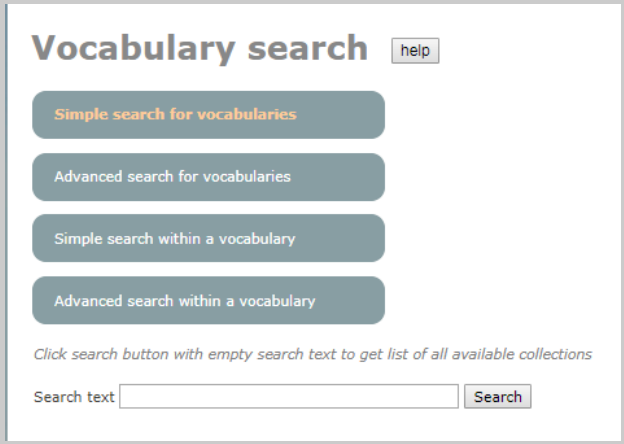
Publication

	http://vocab.nerc.ac.uk/sparql/
	http://vocab.nerc.ac.uk/collection/ http://vocab.nerc.ac.uk/collection/XXX/current/ http://vocab.nerc.ac.uk/collection/XXX/current/YYY/ http://vocab.nerc.ac.uk/scheme/
	http://vocab.nerc.ac.uk/vocab2.wsdl

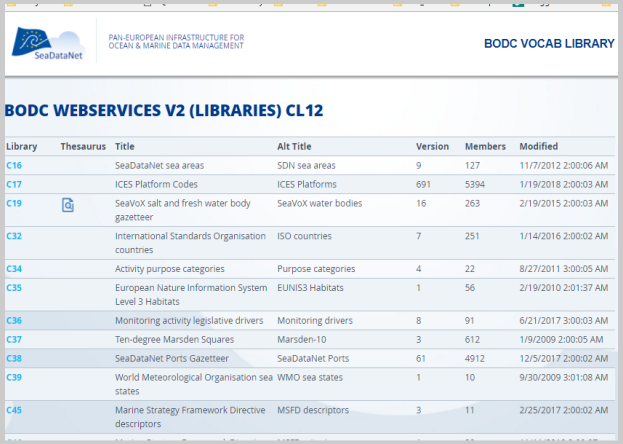
Principles of operation

- Governance
  - managed by the BODC rigorous content governance principles
  - Each controlled vocabulary its content governance authority
  - its content governance model
    - totally internal, collaborative or totally external.
- Deletion vs Deprecation
  - Never delete, only deprecate
  - All deprecated terms are replaced by a new term: skos:replacedBy
- Versioning
  - Version URLs currently on Collection level
  - Version number on terms and collection

Searching

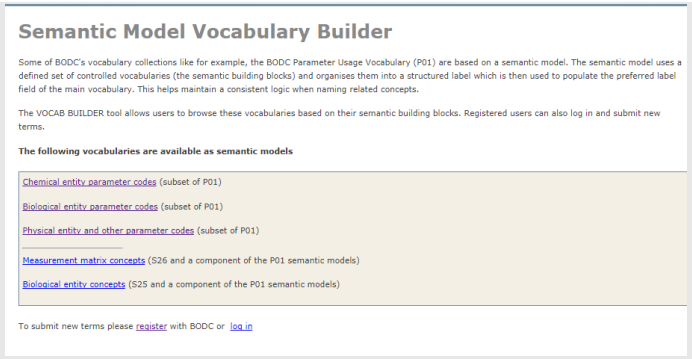


https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/



http://seadatanet.maris2.nl/v_bodc_vocab_v2/welcome.asp

(Search+Build)-P01

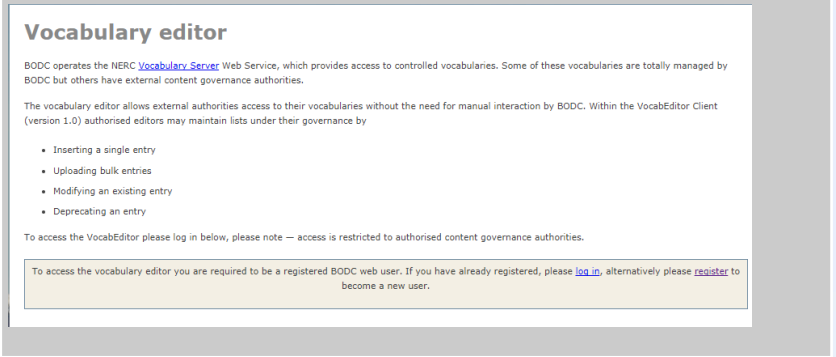


https://www.bodc.ac.uk/resources/vocabularies/vocabulary_builder/

Search P01

http://seadatanet.maris2.nl/bandit/browse_step.php

Edit



https://www.bodc.ac.uk/resources/vocabularies/vocabulary_edittor/

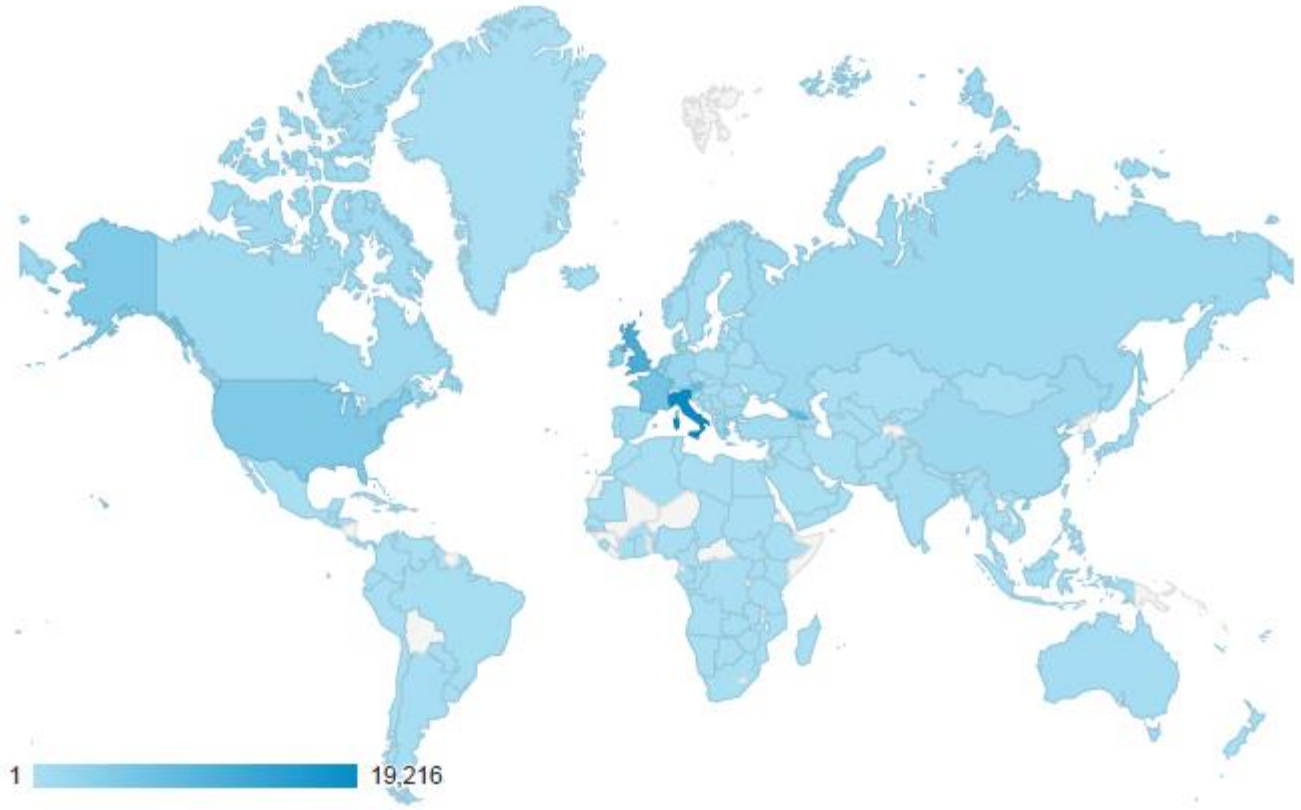
NVS Usage

All Users
100.00% Sessions

Jan 1, 2017 - Dec 31, 2017

Map Overlay

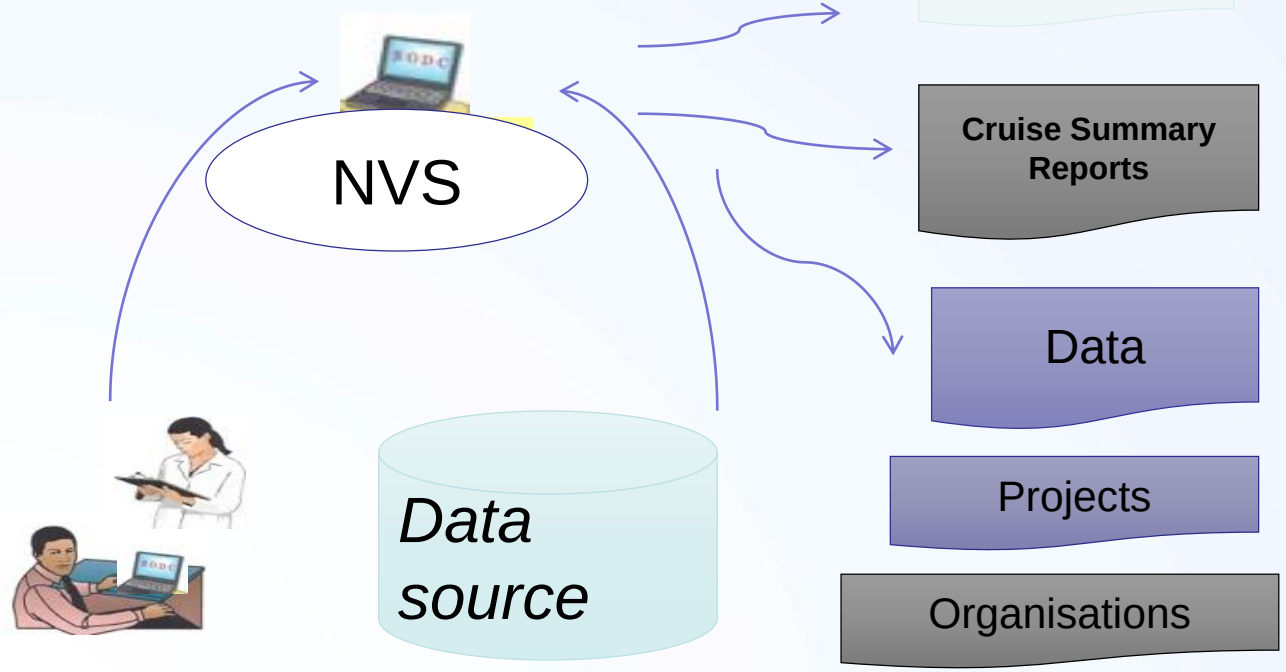
Summary



Country	Sessions
	92,510 % of Total: 100.00% (92,510)
1. Italy	19,355 (20.92%)
2. United Kingdom	10,167 (10.99%)
3. Georgia	8,364 (9.04%)
4. Austria	8,233 (8.90%)
5. France	6,467 (6.99%)
6. United States	4,764 (5.15%)
7. Netherlands	3,490 (3.77%)
8. Ireland	3,192 (3.45%)
9. Belgium	2,591 (2.80%)
10. Germany	2,170 (2.35%)

NVS & EU SeaDataNet-2

Metadata creation Tool



Monitoring Systems

Datasets

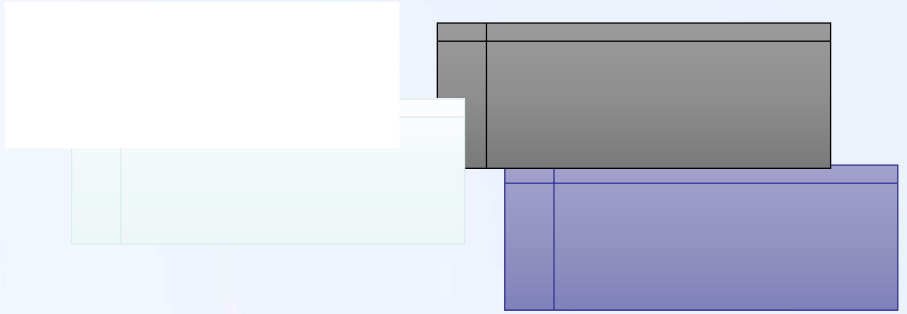
Cruise Summary Reports

Data

Projects

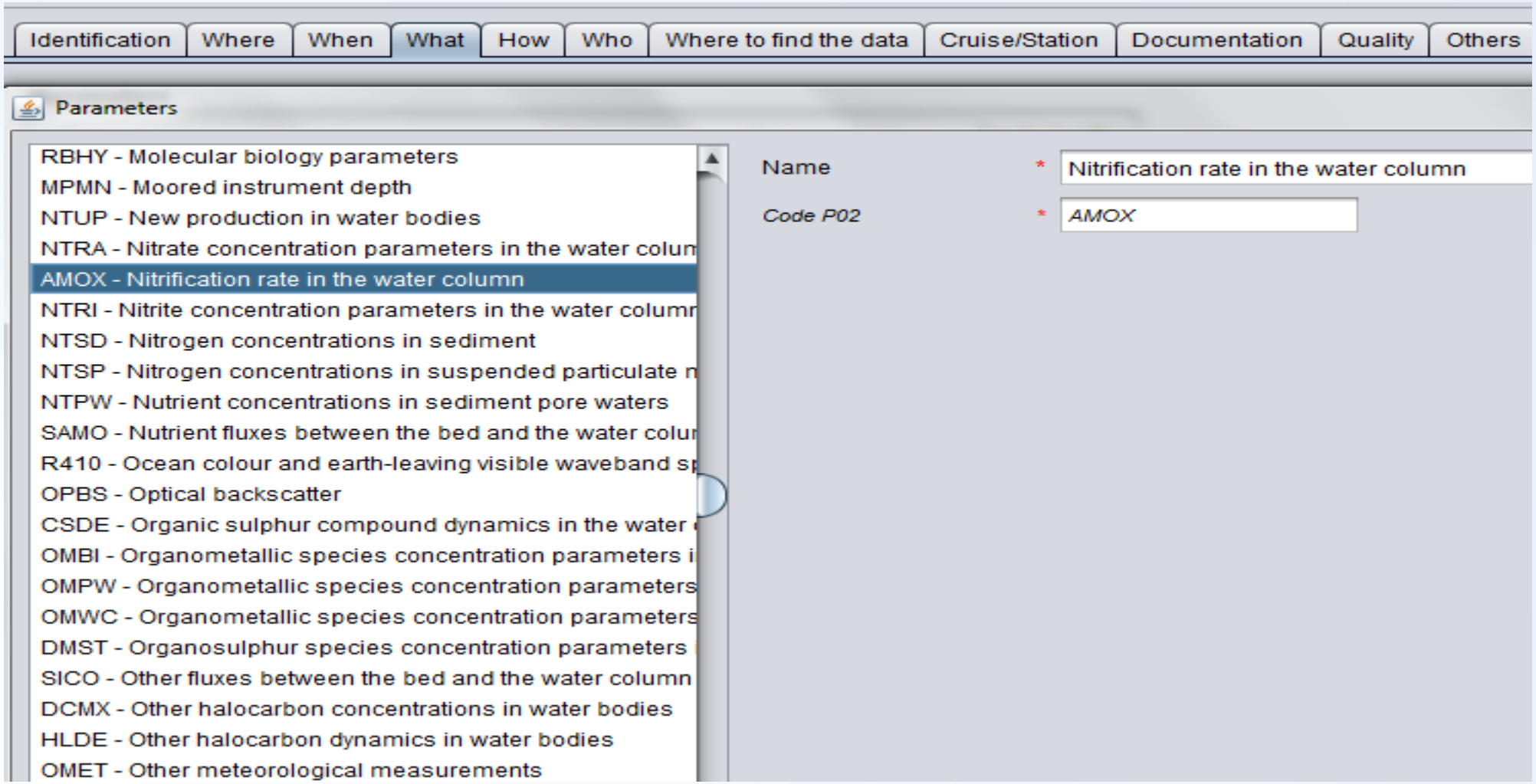
Organisations

Metadata Discovery



Map from http://www.seadatanet.org/Overview/Partners

NVS2.0 & EU SeaDataNet-2



Administration and dimensions
> Administration and dimensions
Atmosphere
> Atmospheric chemistry

Discovery parameters

All
Acoustic backscatter in the water column
Acoustic noise in the water column
Air pressure
Air temperature

Cruise/Station name

Projectname

Datasetname

Instrument type

All
>2000 Hz top-bandwidth single-channel seisn
>2000 Hz top-bandwidth sub-bottom penetrat
1000 Hz top-bandwidth multi-channel seismic

Instrument depth (m) from ____ to ____

Sea regions

All
World
> Arctic Ocean
> > Baffin Bay

Platform type

All
aeroplane
autonomous surface water vehicle
autonomous underwater vehicle

Waterdepth (m) from ____ to ____

Measuring area type: All

Originator: All

Temporal resolution: All

CDI partner: All

Date (yyyymmdd) from ____ to ____

Country: All

Duration ____ to ____ Unit Hou

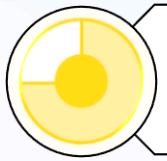
Access restriction

All
academic
by negotiation
commercial charge

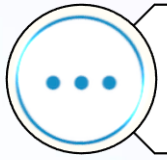
Progress

WP8.1.3 Further develop the Vocabulary Builder tool

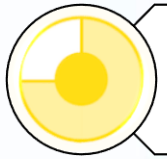
W8.1.1 Improve the transparency of the vocabulary governance model



WP8.1.2 Develop new vocabularies



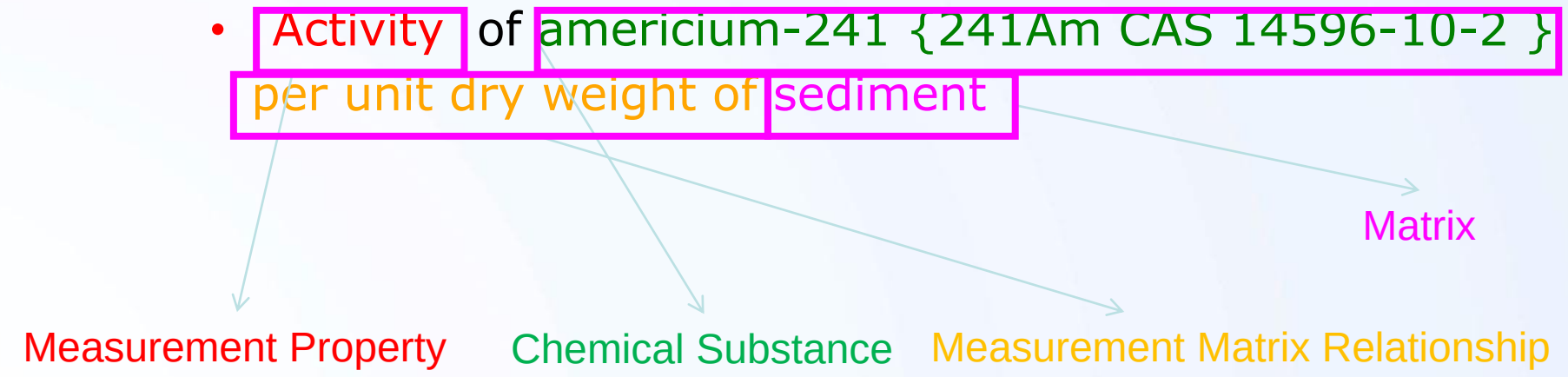
WP8.1.5 Make versioning and history of concepts visible through NVS



WP8.1.6 Document the provenance of mappings

Background

- P01 concept labels are created from a concatenation of concept labels from other vocabularies following an underlying semantic model.



P01 Biological Entity Parameter Code Builder

help

Preferred label

show/hide exact results | reset all

Found **7354** exact matches

- Select a measurement property S06
- Select a statistical qualifier (if applicable) S07
- Select a primary biological entity S25
- Select a secondary biological entity (if applicable) S25
- Select a measurement-matrix relationship S30
- Select a matrix S26
- Select a sample preparation (if applicable) S03
- Select an analytical method (if applicable) S04
- Select a post-analysis processing (if applicable) S05

Return to index

Login to enable new concept creation

Inventories
Vocabularies
NVS search tool
NVS editor
NVS vocabulary builder
BODC parameter codes
SeaVoX
Delivery formats
Products
Help and hints
Portals and links
Search

Semantic Model Vocabulary Builder

Some of BODC's vocabulary collections like for example, the BODC Parameter Usage Vocabulary (P01) are based on a semantic model. The semantic model uses a defined set of controlled vocabularies (the semantic building blocks) and organises them into a structured label which is then used to populate the preferred label field of the main vocabulary. This helps maintain a consistent logic when naming related concepts.

The VOCAB BUILDER tool allows users to browse these vocabularies based on their semantic building blocks. Registered users can also log in and submit new terms.

The following vocabularies are available as semantic models

Chemical entity parameter codes (subset of P01)

Biological entity parameter codes (subset of P01)

Physical entity and other parameter codes (subset of P01)

Measurement matrix concepts (S26 and a component of the P01 semantic models)

Biological entity concepts (S25 and a component of the P01 semantic models)

Search: No Login

Submit terms: Login

To submit new terms please register with BODC or log in

Content progress

- New flow cytometry technical codes added for SDC to P01
- New F02 vocabulary set up and populated with reference names for flow cytometry groupings as agreed by group of FC experts
- A couple of Sampling metadata terms added to P01 GP model (see e.g. http://vocab.nerc.ac.uk/collection/P01/current/MTHAREA1/ and http://vocab.nerc.ac.uk/collection/P01/current/MSHSIZE1/) for Emodnet_chem microplastic request; the physical entity is defined as the "sample collector" i.e. uses the same terminology as that used in L21.
- New C17 platform codes populated
- New contaminant codes created for Emodnet chem

Improve the transparency of the vocabulary governance model

- Keep the governance information on the collection level
- Create github page for each collection and associate the collection with the page through an rdfs:seeAlso

Example: P02

<skos:Collection rdf:about="http://vocab.nerc.ac.uk/collection/P02/current/">
<dc:creator>SeaDataNet</dc:creator>
<grg:RE_RegisterOwner>SeaDataNet</grg:RE_RegisterOwner>
<rdfs:comment>Governance for vocabularies used in the EU SeaDataNet project implemented as consultation between BODC and the members of the SeaDataNet Technical Task Team</rdfs:comment>
<rdfs:seeAlso>https://github.com/gwemon/P02-Biological-terms-review/tree/master</rdfs:seeAlso>

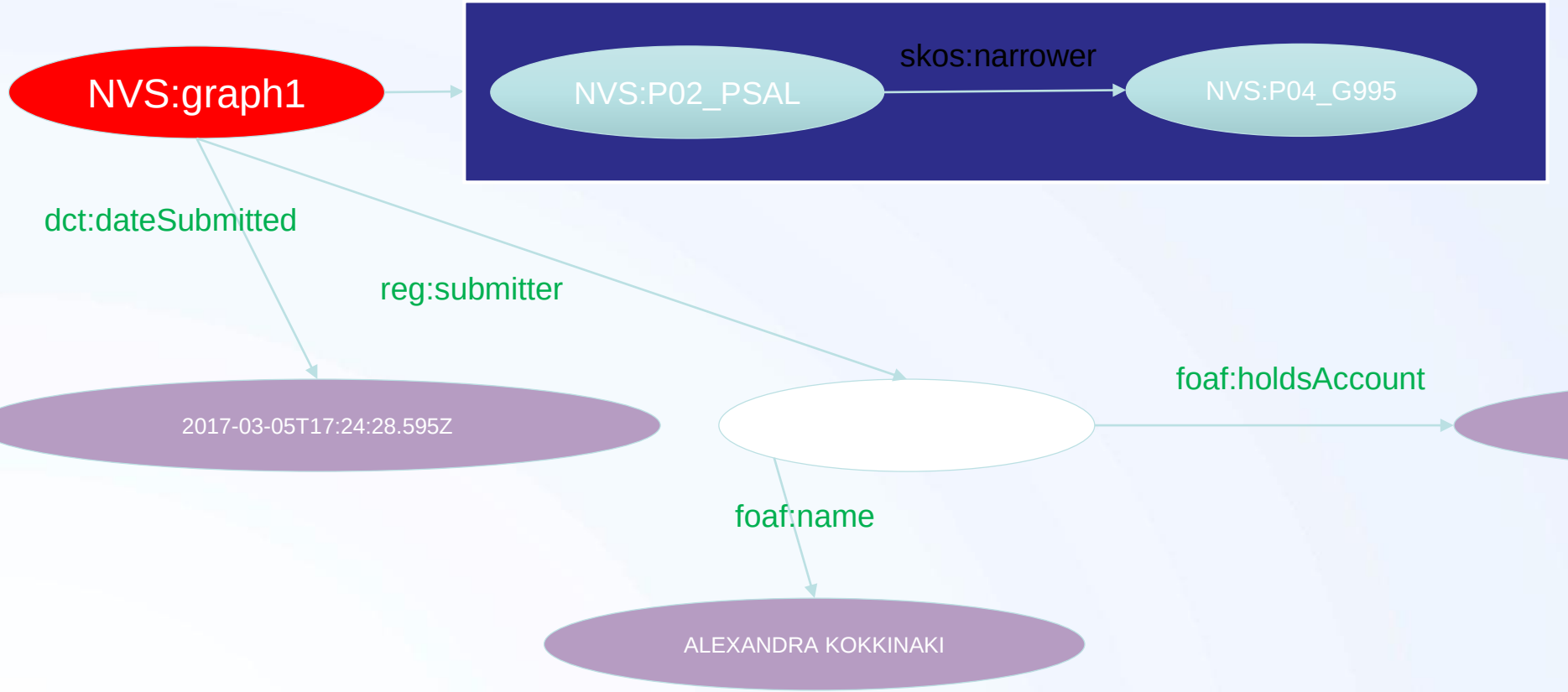
Versioning of concepts- URI design

- http://vocab.nerc.ac.uk/collection/P07/current/CF12N86/ Brings the current concept version
- http://vocab.nerc.ac.uk/collection/P07/current/CF12N86/1/ Brings concept version 1 without mappings
- http://vocab.nerc.ac.uk/collection/P07/current/CF12N86/2/ Brings concept version 2 without mappings
- http://vocab.nerc.ac.uk/collection/P07/current/CF12N86/3/ Brings concept version 3 without mappings

- http://vocab.nerc.ac.uk/collection/P07/current/CF12N86/ dct:hasVersion
- http://vocab.nerc.ac.uk/collection/P07/current/CF12N86/1/, http://vocab.nerc.ac.uk/collection/P07/current/CF12N86/2/, http://vocab.nerc.ac.uk/collection/P07/current/CF12N86/3/

WP8.1.6: Provenance of mappings - Progress

@prefix reg: http://purl.org/linked-data/registry#
@prefix foaf: <http://xmlns.com/foaf/0.1/>
@prefix skos: <http://xmlns.com/foaf/0.1/>




Thanks!
QUESTIONS?