Further Developments of Common Vocabularies within SeaDataCloud and related european Projects

ODIP II Final Workshop
Galway, Ireland - 02 October 2017

Gwenaëlle Moncoiffé
VOCAB MANAGEMENT GROUP
BRITISH OCEANOGRAPHIC DATA CENTRE
Online Access through the NERC Vocab Server (NVS)

Machine-to-machine

- RDF restful API  http://vocab.nerc.ac.uk
- SOAP API       http://vocab.nerc.ac.uk/vocab2.wsdl
- SPARQL end point http://vocab.nerc.ac.uk/sparql/

Browser interfaces

- **NVS Search** – Search across and within vocabularies
  [https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/](https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/)

- **NVS Editor** – Allow external editors to submit and edit new terms and mappings to specific vocabularies (access control)
  [https://www.bodc.ac.uk/resources/vocabularies/vocabulary_editor/](https://www.bodc.ac.uk/resources/vocabularies/vocabulary_editor/)

- **NVS Vocabulary Builder** (V2 due for release next month) -- Search and submit parameter codes using the semantic building blocks of the BODC parameters.
  [https://www.bodc.ac.uk/resources/vocabularies/vocabulary_builder/](https://www.bodc.ac.uk/resources/vocabularies/vocabulary_builder/)
Vocabularies hosted at BODC
Status as of September 2017

- **Collections**
  - 231 vocabulary collections accessible through the NVS
  - 70 (30%) are owned and governed by BODC
  - 48 (21%) are under SDC/EMODnet/SeaVox/SWE content governance
  - Remainder (113) are owned by 25 different governing bodies

- **Concepts**
  - 151977 valid concepts including:
    - P01 parameter codes: 37368
    - L22 instrument codes: 1198

- **Mappings**
  - 595073 internal mappings between concepts
  - 20804 mappings of NVS concepts to external resources

- **Calls to NVS server**
  - Between 37000 and 90000 calls per months in 2017

- **Requests for new codes**
  - In 2017: regularly over 20 requests per month every month
  - Often 1/3 external but 50:50 external:internal in August and September
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
SeaDataCloud Vocabulary deliverables

- Improve the transparency of the vocabulary governance model
  → not yet started

- Develop new vocabularies
  → Started

- Undertake a review of P02 and P03 vocabularies
  → Started

- Make versioning and history of concepts visible through NVS
  → not yet started

- Document the provenance of mappings
  → started (see presentation by A. Kokkinaki)

- Further develop the Vocabulary Builder tool
  → started and near completion
Improve the transparency of the vocabulary governance model

Broaden and extend the current SeaVox governance group. A combined SeaDataNet and MarineXML Vocabulary Content Governance Group moderated by BODC
Improve the transparency of the vocabulary governance model

• Adopt best practices as determined by RDA Vocabularies Services Interest Group

• Add contact details to the governance authority
• Track governance discussions and decisions following models from ISO 19135 and ISO/IEC 11179-3
• Tie a mailing list conversation with a version number
• Archived mailing lists

• Will support a more open governance, greater collaboration and input from the user community and domain experts

Not yet started
Collaborations in developing new vocabularies

- **Sensor Web Enablement (SWE) Marine Profile Group**
  - 7 New collections have been created and population has started
  - Governance group is active
  - BODC Vocab contact is Louise Darroch
  - Governance via the Vocab Editor is being set up

- **Flow Cytometry Vocabularies WG (SDC)**
  - Domain expert consultation
  - Discussions between domain experts, data managers and vocab experts on-going
  - BODC Vocab contact is Gwen Moncoiffé

- **Marine Micro-litter Vocabularies (EMODnet Chemistry)**
  - Discussions between domain experts, data managers and vocab experts have taken place
  - Set of collections and concepts ready for publication
  - BODC Vocab contact is Gwen Moncoiffé
New Vocabularies e.g. SWE

SWE Marine profiles group lead by 52North
OGC SWE Vocabularies governed by SWE Marine Profile group
– http://vocab.nerc.ac.uk/collection/W03/current/ --> SensorML Event Types
– http://vocab.nerc.ac.uk/collection/W04/current/ --> SensorML Capabilities
– http://vocab.nerc.ac.uk/collection/W05/current/ --> SensorML Characteristics
– http://vocab.nerc.ac.uk/collection/W06/current/ --> SensorML Classification
– http://vocab.nerc.ac.uk/collection/W07/current/ --> SensorML Identification
– http://vocab.nerc.ac.uk/collection/W08/current/ --> SensorML Contacts

• New vocabularies as requested
• OGC vocabularies as suggested by the SWE Marine Profiles group
• Improve the governance and transparency of the group based on experience from SeaVox
• Ongoing
Review of P02 and P03 Vocabularies

- P02 and P03 SDN vocabularies need to be reviewed and some terms will need to be deprecated.

- Upgrade to existing systems and software so that the deprecation of concepts and terms is supported. Under the coordination of the TTG (Maris), development work will be required to CDI import (Maris), Mikado software (Ifremer), Nemo software (Ifremer), Download Manager software (Ifremer), CSR XML import (BSH), CSR data input webform (BSH), EDIOS import (BODC), EDMED import (BODC).

- A draft proposal for rationalising P02 and P03 (mainly terms related to biological codes) is being drafted and will be submitted for consultation to the SDC-TT.

Started
Make Versioning of Concepts Visible

- Individual concepts have version control stored internally and a mechanism will be developed to allow users access to the version history of individual concepts through the NVS.

<table>
<thead>
<tr>
<th>ID</th>
<th>Code</th>
<th>Description</th>
<th>Parameters</th>
<th>Dates</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>P02</td>
<td>PSAL</td>
<td>Salinity of the water column</td>
<td>WC_Sal</td>
<td>29-Jul-04</td>
<td>accepted</td>
</tr>
<tr>
<td>P02</td>
<td>PSAL</td>
<td>Salinity of the water column</td>
<td>WC_Sal</td>
<td>29-Jul-04</td>
<td>accepted</td>
</tr>
</tbody>
</table>

Not yet started
Vocabulary Builder Tool

- Also known as the "one-armed bandit" or fruit-machine concept
- Wheels = semantic elements of the P01 parameter preferred labels
- Developed as part of EMODnet Chemistry > chemical codes
- Built as a search and submission tool requiring the user to log in
- Re-deployed as a search tool with option to log in to submit new terms
- Functionality extended to search any type of parameter using 3 entry points corresponding to the 3 P01 semantic models
  - Chemical Entity Parameters
  - Biological Entity Parameters
  - Physical Entity and Other Parameters
- Remodellisation work to expose the whole P01 parameter dictionary through these 3 semantic models
- Modellisation work to develop a structured semantics to support some of the descriptive elements of the SeaVoX Device Catalogue (L22) is also being investigated
Semantic Model Vocabulary Builder

Some of BODC's vocabulary collections like for example, the BODC Parameter Usage Vocabulary (PO1) 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 PO1)
- Biological entity parameter codes (subset of PO1)
- Physical entity and other parameter codes (subset of PO1)
- Measurement matrix concepts (S26 and a component of the PO1 semantic models)
- Biological entity concepts (S25 and a component of the PO1 semantic models)

To submit new terms please register with BODC or log in
P01 Physical Entity and Other Parameter Code Builder

Preferred label

Found 3714 exact matches:

- Select a measurement property
- Select a statistical parameter
- Select a physical entity (if applicable)
- Select a measurement-matrix relationship
- Select a matrix
- Select a sample preparation (if applicable)
- Select an analytical method (if applicable)
- Select a post-analysis processing (if applicable)

Return to Index

Login to enable new concept creation
P01 Physical Entity and Other Parameter Code Builder

Preferred label
Practical salinity

Found 31 exact matches

Select a measurement property

Approved

Select approved concept

Found 0 | Showing (1 - 0) | 1
Practical salinity
Clear selection

The quantity of dissolved ions (predominantly salt in seawater) expressed on a scale (PSU-78) based on the conductivity ratio of a seawater sample to a standard KCl solution.
### P01 Physical Entity and Other Parameter Code Builder

**Preferred label:**

- Practical salinity not specified water body

---

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Alternative label</th>
<th>Preferred label</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSlT201</td>
<td>$p_{_sl}$</td>
<td>Practical salinity of the water body</td>
</tr>
<tr>
<td>PSlC0202</td>
<td>$p_{_sl}$,CTD2</td>
<td>Practical salinity of the water body by CTD (second sensor) and computation using UNESCO 1983 algorithm</td>
</tr>
<tr>
<td>PSlC0202</td>
<td>$p_{_sl}$,CTD,calib2</td>
<td>Practical salinity of the water body by CTD (second sensor) and computation using UNESCO 1983 algorithm and NO calibration against independent measurements</td>
</tr>
<tr>
<td>PSlC0202</td>
<td>$p_{_sl}$,CTD,calib2</td>
<td>Practical salinity of the water body by CTD (second sensor) and computation using UNESCO 1983 algorithm and calibration against independent measurements</td>
</tr>
<tr>
<td>PSlC0202</td>
<td>$p_{_sl}$,CTD2</td>
<td>Practical salinity of the water body by CTD and computation using UNESCO 1983 algorithm</td>
</tr>
</tbody>
</table>

- Select a measurement property
- Select a statistical parameter
- Select a physical entity (if applicable)
- Select a measurement-matrix relationship
- Select a matrix
- Select a sample preparation (if applicable)
- Select an analytical method (if applicable)
- Select a post-analysis processing (if applicable)
P01 Biological Entity Parameter Code Builder

- Preferred label:

Found 7116 exact matches

- Select a measurement property
- Select a primary biological entity
- Select a secondary biological entity (if applicable)
- Select a matrix
- Select a measurement-matrix relationship
- Select a statistical qualifier (if applicable)
- Select a sample preparation (if applicable)
- Select an analytical method (if applicable)
- Select a post-analysis processing (if applicable)

Return to index

Login to enable new concept creation
P01 Biological Entity Parameter Code Builder

Preferred label:

Abundance Skeletonema costatum (ITIS: 2402: WoRMS 140074)

Found 2 exact matches | Showing (1 - 2) | 1

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Alternative label</th>
<th>Preferred label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2709928</td>
<td>Abundance(Skeletonema costatum (ITIS: 2402: WoRMS 140074) per unit volume of aquarium water sample by optical microscopy) A. Salmon (1996) Ecosystems 7: 1273.</td>
<td></td>
</tr>
<tr>
<td>P1010101</td>
<td>P10101</td>
<td>Abundance of Skeletonema costatum (ITIS: 2402: WoRMS 140074) per unit volume of the water body by optical microscopy.</td>
</tr>
</tbody>
</table>
P01 Chemical Entity Parameter Code Builder

Found 13085 exact matches

- Select a measurement property
- Select a measurement statistical qualifier
- Select a chemical substance
- Select a measurement-matrix relationship
- Select a matrix
- Select a select a sample preparation (if applicable)
- Select a select an analytical method (if applicable)
- Select a select a post-analysis processing (if applicable)
P01 Chemical Entity Parameter Code Builder

Preferred label
Activity not specified of americium-241 (241Am CAS 14596-10-2) per unit dry weight of sediment

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Alternative label</th>
<th>Preferred label</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA241Cont</td>
<td>Activity of americium-241 (241Am CAS 14596-10-2) per unit dry weight of sediment by alpha plus beta spectroscopy</td>
<td></td>
</tr>
<tr>
<td>GA241Cont</td>
<td>Activity of americium-241 (241Am CAS 14596-10-2) per unit dry weight of sediment by gamma spectroscopy</td>
<td></td>
</tr>
</tbody>
</table>

Found 2 exact matches | Showing ( 1 - 2 ) | ▼

- Select a measurement property
- Select a measurement statistical qualifier
- Select a chemical substance
- Select a measurement-matrix relationship
- Select a matrix
- Select a select a sample preparation (if applicable)
- Select a select an analytical method (if applicable)
- Select a select a post-analysis processing (if applicable)
Activity of americium-241 (241Am CAS 14596-10-2) per unit dry weight of sediment by alpha plus beta spectroscopy

URI: http://vocab.nerc.ac.uk/collection/P01/current/A41CSPXT/
Identifier (en): SDN:P01::A41CSPXT
Preferred label (en): Activity of americium-241 (241Am CAS 14596-10-2) per unit dry weight of sediment by alpha plus beta spectroscopy
Definition (en): Unavailable
Version Info (): 2
Deprecated (): False
Broader (en): http://vocab.nerc.ac.uk/collection/P02/current/SRAD/
Broader (en): http://vocab.nerc.ac.uk/collection/S06/current/S0600043/
Broader (en): http://vocab.nerc.ac.uk/collection/S26/current/MAT00136/
Broader (en): http://vocab.nerc.ac.uk/collection/S27/current/CS002761/
Related (en): http://vocab.nerc.ac.uk/collection/P06/current/UBQK/
Related (en): http://vocab.nerc.ac.uk/collection/S02/current/S041/
Related (en): http://vocab.nerc.ac.uk/collection/S04/current/S04117/
Date (): 2015-07-28 11:45:23.0
Questions?