The BODC Parameter Usage Vocabulary (PUV) semantic model exposed

Gwenaëlle Moncoiffé
Alexandra Kokkinaki

Vocabulary Management Group at BODC
National Oceanography Centre, Liverpool and Southampton

Adapted from a presentation given at IMDIS, 5-7 November 2018, Barcelona, Spain
Last reviewed: 27/08/2019
BODC Parameter Usage Vocabulary (PUV)

Codename: P01

A controlled vocabulary for labelling data streams and fields in oceanographic databases and data files

- Has underpinned BODC data management systems since 1980s
- From <20 codes in the 1980s to ~5000 codes in 2000 to >40,000 today
- Growth in diversity and complexity as it incorporates concepts from biology, biogeochemistry, and geophysics
- Accessible online in 2005 as part of the NERC DataGrid and NERC-funded EnParDis projects (NERC Vocabulary Server NVS)
- Elements of the semantic model put in place but not exposed
- Further standardisation and growth as part of European-funded projects SeaDataNet, NETMAR, SeaDataCloud, and EMODnet chemistry
- Incorporates accepted SKOS and RDF standards in 2012
BODC PUV = P01 collection

Opaque 8-byte identifier

Structured label based on a semantic model
The semantic model is based on the conceptualisation of what constitutes a measurement and the atomisation into its constituent parts.

Elements constrained against controlled vocabularies
Advantages of exposing the model

- Easier to search
- Easier to align to other semantic resources
- Easier to maintain
- Each element becomes a semantic resource

- Resources can be shared, linked to, re-used, and grown collaboratively
The property element

PROPERTY of an OBJECT in RELATION to a MATRIX by a METHOD

- quantitative (Concentration, Practical salinity, Production rate, Abundance)
- qualitative: binary (Presence or absence), ordinal (Abundance category), or nominal (Colour class, Shape class)

- All PROPERTY terms are defined in collection S06

http://vocab.nerc.ac.uk/collection/S06/current/
https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/S06/
The object element

A PROPERTY of an OBJECT in RELATION to a MATRIX by a METHOD

- Physical
  - phenomenon (waves, wind)
  - object (measurement platform, particles)
  - element (water, air)

- Biological
  - substance
  - group of substances
  - chemical element

- Chemical
  - organism or any component parts (including organs)
  - an association of biological entities (predator-prey or parasite-host relationships)
The chemical entity object

A property of an object in relation to a matrix by a method

Physical  Chemical  Biological

Chemical objects referred to in a P01 label are defined in collection S27

http://vocab.nerc.ac.uk/collection/S27/current/
https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/S27/
Secondary compound vocabularies

The physical and biological entities are defined in the S25 and S29 SKOS collections. These are themselves structured compound vocabularies meaning that the preferred label of the concepts within those collections is constructed from the combination of concepts defined in separate controlled vocabularies.
The physical entity object

a PROPERTY of an OBJECT in RELATION to a MATRIX by a METHOD

Example of P01 label

Proportion by dry weight of particles (180-300um) in the sediment by sieving and settling tube method

http://vocab.nerc.ac.uk/collection/P01/current/PRSC0217/
a **PROPERTY** of an **OBJECT** in **RELATION** to a **MATRIX** by a **METHOD**

**Physical**

- **physical entity name**
  - S29
  - water current

- **physical entity sub-group**
  - S19
  - moving platform

- **datum**
  - S20
  - not applicable

---

**Example of P01 label**

Eastward velocity of water current relative to moving platform in the water body by shipborne acoustic doppler current profiler (ADCP)

**If an element is set to “not applicable” then this element is ignored when the compound name is compiled.**

http://vocab.nerc.ac.uk/collection/P01/current/LREWAS01/
The biological entity object

Example of S25 label

Halichoerus grypus (ITIS: 180653: WoRMS 137080) [Stage: post-weaned pup Sex: female Subcomponent: blubber Subgroup: dead]
Halichoerus grypus (ITIS: 180653; WoRMS 137080) [Stage: post-weaned pup] [Sex: female] [Subcomponent: blubber] [Subgroup: dead]

In this example the object of interest on which the set of measurements or observations was made is the blubber of a dead female post-weaned grey seal pup.

http://vocab.nerc.ac.uk/collection/S25/current/BE006418/
In some situations, the biological object of interest cannot be identified to a high level of taxonomic precision but the scientist has used keywords to describe its shape and the fact that the organism had photosynthetic pigments – this information is preserved in the label by using terms from controlled vocabularies concerned with morphology and sub-grouping of biological organisms.

http://vocab.nerc.ac.uk/collection/S25/current/BE001841/
Compatibility with Darwin Core standard

A Property of an Object in Relation to a Matrix by a Method

Abundance of biological entity specified elsewhere per unit volume of the sediment

Track duration of biological entity specified elsewhere

Count (January) {midwinter count} of biological entity specified elsewhere

etc.

Examples of P01 labels

S25 = “biological entity specified elsewhere”
Links to authoritative name registries

- **Physical**
- **Biological**
- **Chemical**

A PROPERTY of a **OBJECT** in RELATION to a **MATRIX** by a **METHOD**

- S29
- S27
- S25: biological entity specified elsewhere

- ChEBI
- Aphiaid+lsid+itis

- World Register of Marine Species (WORMS)

- CHEMBL: A TOXNET DATABASE
  - Lite, Browse, Advanced
The matrix element

a PROPERTY of an OBJECT in RELATION to MATRIX by a METHOD

MATRIX is the environment in which the measurement is made or in which the object of interest is embedded

- MATRIX is defined in S26
- MATRIX is a structured compound vocabulary
The matrix element

A property of an object in relation to a matrix by a method.

- **sphere name**
  - Mainly used for sub-grouping of sediments and suspended particulate material based on size.
- **sphere sub-group**
- **phase name**
  - PARTICULATE
    - dissolved plus reactive particulate
    - aerosol
    - Etc.
- **phase sub-group**
  - GF/F-10um
  - slow sinking >GF/F
  - <0.2um
  - 127nm
  - Etc.

- **S21**
  - water body
  - sediment
  - sediment pore water
  - atmosphere
  - Etc.

- **S22**
- **S23**
- **S24**
The biota matrix variant

Example of P01 label

Concentration of hexachlorobenzene \{HCB CAS 118-74-1\} per unit wet weight of biota \{Halichoerus grypus (ITIS: 180653: WoRMS 137080) [Subcomponent: blubber]\}

If the matrix is the biota then the matrix element is the combination of the S26 term “biota” and a biological entity (S25)

http://vocab.nerc.ac.uk/collection/P01/current/IC003116/
The relationship element

a **PROPERTY** of an **OBJECT** in **RELATION** to a **MATRIX** by a **METHOD**

- **RELATION** is the LINK between the **PROPERTY** of the **OBJECT** and the **MATRIX**
- It contains important information about the multiple ways of expressing a measured quantity in relation to its environment
- It forces us to be explicit about the way the measurement is reported e.g.
  - per unit volume of the water body…
  - per unit wet weight of biota…
  - integrated over depth in the water body…

- The relation terms are defined in [S02](#)

Examples of P01 labels

- Count of *Halichoerus grypus* (ITIS: 180653: WoRMS 137080) out of the water body
- Count of *Halichoerus grypus* (ITIS: 180653: WoRMS 137080) in the water body
The method element

- The method fields are optional
- P01 codes with a method defined are mapped to the broader non-method specific codes when it exists
- The broader terms are used for aggregation or when the information is stored elsewhere in a schema or when the information is simply not available (as can be the case when dealing with legacy data)
Properties can be associated with a statistical term to create P01 concepts of the form

a PROPERTY STATISTIC of an OBJECT in RELATION to a MATRIX by a METHOD

• example: Concentration standard deviation of aluminium {Al CAS 7429-90-5} per unit volume of the water body [particulate 0.8-51um phase]

• STATISTIC is defined in S07

http://vocab.nerc.ac.uk/collection/S07/current/
https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/S07/
The semantic elements are defined in the S01 collection

http://vocab.nerc.ac.uk/collection/S01/current/

Mappings connect the S01 concept to concepts held in its controlled vocabulary

For example
S01 concept “Parameter entity” is mapped to every concepts in S06
S01 concept “Chemical entity” is mapped to every concepts in S27
S01 concept “Matrix” is mapped to every concepts in S26
etc.