The BODC Parameter Usage Vocabulary (PUV) semantic model exposed

Gwenaëlle Moncoiffé
Alexandra Kokkinaki

Vocabulary Management Group at BODC
National Oceanography Centre, Liverpool and Southampton
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
- <20 codes in the 1980s to ~5000 codes in 2000 to >40000 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
- Incorporated accepted SKOS and RDF standard in 2012 (NVS2)
BODC PUV = P01 collection

opaque 8-byte identifier

Structured label based on a semantic model
Core elements

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

a PROPERTY of an OBJECT in RELATION to a MATRIX by a METHOD
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/
Properties can be associated with a statistical term to become a **PROPERTY STATISTIC** of an **OBJECT** in **RELATION** to a **MATRIX** by a **METHOD**

- examples: standard deviation, mean, annual mean, hourly mean, etc.
- **STATISTIC** is defined in S07

http://vocab.nerc.ac.uk/collection/S07/current/
https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/S07/
The object element

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

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

- Chemical
  - substance
  - group of substances
  - chemical element

- Biological
  - organism or any component parts (including organs)
  - an association of biological entities (predator-prey or parasite-host relationships)
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

Example of P01 label

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

http://vocab.nerc.ac.uk/collection/P01/current/LREWAS01/
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]

The biological entity is:
the blubber of a dead female post-weaned pup grey seal

http://vocab.nerc.ac.uk/collection/S25/current/BE006418/
**Dinophyceae** (ITIS: 9874: WoRMS 19542)  
[Morphology: banana-shaped]  
[Subgroup: autotrophic]

Unidentified **autotrophic dinoflagellate** (banana-shape)

Source: Figueroa et al 2010

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”

BIOLOGY

EMODnet

BIS OCEAN BIOGEOGRAPHIC INFORMATION SYSTEM

National Oceanography Centre

European Marine Observation and Data Network

British Oceanographic Data Centre

NERC SCIENCE OF THE ENVIRONMENT
Links to authoritative name registries

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

- **S29**
- **S27**
- **S25** biological entity specified elsewhere

- **ChemiDplus**
- **ChE BI**
- **WoRMS**

- **aphiaid:lsid:itis**
The matrix element

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

- Sphere name
  - S21
    - water body
    - sediment
    - sediment pore water
    - atmosphere
    - etc

- Sphere sub-group
  - S22
    - particulate
    - dissolved plus reactive particulate
    - aerosol
    - etc

- Phase name
  - S23
    - particulate
  - GF/F-10um
    - slow sinking >GF/F
    - <0.2um
    - 127nm
    - etc

- Phase sub-group
  - S24
Matrix biota variant

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

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

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]}

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
- 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
- The broader terms are used for aggregation; when the information is stored elsewhere in a schema; or when the information is not available
Anatomy of a BODC Parameter Code

Parameter entity

Statistical entity

Object entity

Matrix entity

Method entity

WHAT

Measurement – matrix relationship

WHERE

HOW

- Biological
- Chemical
- Physical

Sample preparation
Analysis
Data processing
The semantic model on the NERC Vocab Server (NVS)

The semantic elements are defined in the S01 collection
http://vocab.nerc.ac.uk/collection/S01/current/

Mappings connect the S01 concept to 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.
SPARQL endpoint vocab.nerc.ac.uk/sparql/
Tools using the semantic model

- BODC Vocab builder tool
- MARIS Facet search tool

**CONTENT GOVERNANCE**

Content governance of the vocabularies is very important and is done by a combined SeaDataNet and MarineXML Vocabulary Content Governance Group (SeaVoX), moderated by BODC, and including experts from SeaDataNet, MMI, MOTIVE, JCOMMOPS and more international groups. SeaVoX discussions are based on an e-mail list server.

NVS2.0
Ontology expression

sosa:ObservableProperty

cpm:ObservableProperty

rdf:type

cpm:property

P01/AMONAATX

cpm:objectOfInterest

S06/S0600045

rdf:type

cpm:Property

S27/CS026908

rdf:type

cpm:Substance

S26/MAT00633

rdf:type

cpm:Matrix

S02/S053

rdf:type

cpm:Constraint

sosa:usedProcedure

S04/S04137

rdf:type

sosa:Procedure

NERC Science of the Environment

National Oceanography Centre
British Oceanographic Data Centre
Annotating data with controlled vocabularies is time-consuming. Standardized fine-grain annotation ensures optimum re-use of the data.

Entities tagged with precise concepts can always be aggregated. Entities tagged with broad concepts will always have a more limited re-use value.
Thank you

Roy Lowry - BODC former technical director and creator of the dictionary and the NVS

Former BODC staff members Adam Leadbetter and Rob Thomas (both now at Marine Institute in Ireland)

The Vocab Management team at BODC including Lou Darroch, Dani Edgar, Mark Hebden, Vi Paba, Arwen Bargery, DBAs Paul McGarrigle, Sean Gaffney, and software developer Quyen Luong

The user community

NERC funding and in particular National Capability funding

European Union funding (SeaDataNet, NETMAR, and SeaDataCloud)
Any questions?

vocab.services@bodc.ac.uk

GitHub

https://github.com/nvs-vocabs/P01