



SeaDataCloud

SeaDataCloud First Training Course
Vocabulary Training Session

Alexandra Kokkinaki and James Ayliffe (BODC)

First Training, Ostend, 22nd - 27th June 2018
sdn-userdesk@seadatanet.org – www.seadatanet.org

Which Vocabularies?

- Parameter Usage Vocabulary (P01) → Focus on this
- Data Production Tools (L22) →
- Units of measurement (P06) →
- Parameter Discovery Vocabulary (P02) →

The P01 Parameter Usage Vocabulary

- A **controlled vocabulary** for labelling data channels and data value fields (i.e. recorded parameters, observations and measurement variables in environmental datasets)
- Became a structured compound vocabulary in 2004 supported by a number of semantic models
- The semantic model is the conceptualisation of what is being measured and the management of its constituent parts as separate controlled vocabularies.
- P01 has underpinned BODC data management systems since the 1990s
- Adopted by SeaDataCloud and more recently by OBIS (De Pooter et al, 2017 <https://doi.org/10.3897/bdj.5.e10989>) to facilitate data exchange, interoperability and discovery

BODC P01 collection

← → ↻ [Natural Environment Research Council \[GB\] | https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/P01/](https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/P01/)

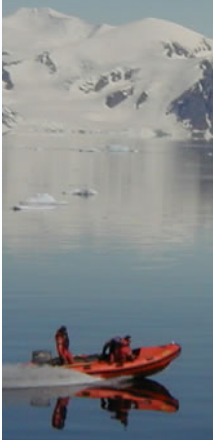
NVS editor
 NVS vocabulary builder
 BODC parameter codes
 SeaVoX

▶ Delivery formats
 ▶ Products
 ▶ Help and hints
 ▶ Portals and links
 Search

Search text Vocabulary [advanced options](#)

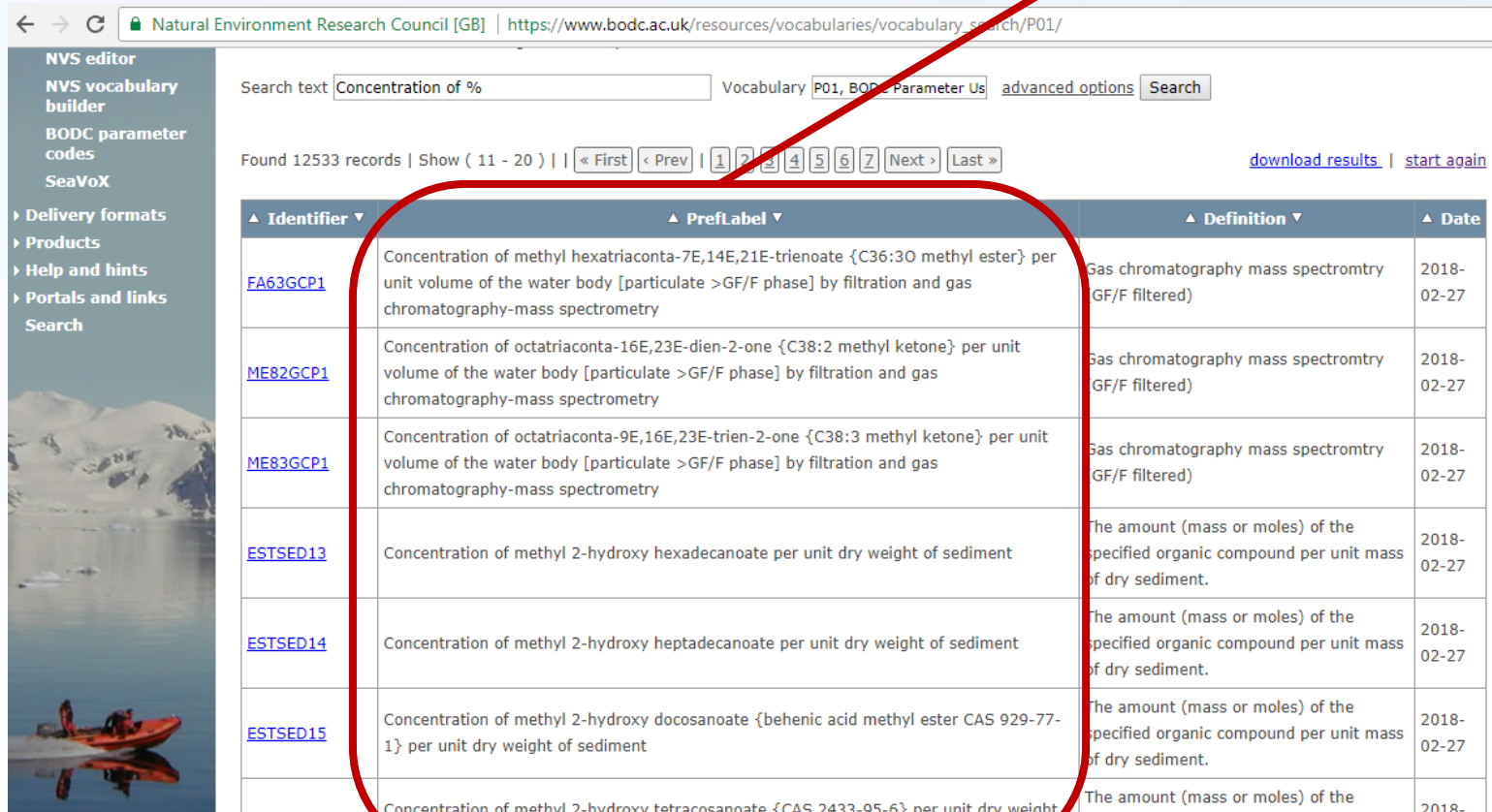
Found 12533 records | Show (11 - 20) | | | [download results](#) | [start again](#)

▲ Identifier ▼	▲ PrefLabel ▼	▲ Definition ▼	▲ Date
FA63GCP1	Concentration of methyl hexatriaconta-7E,14E,21E-trienoate {C36:30 methyl ester} per unit volume of the water body [particulate >GF/F phase] by filtration and gas chromatography-mass spectrometry	Gas chromatography mass spectrometry (GF/F filtered)	2018-02-27
ME82GCP1	Concentration of octatriaconta-16E,23E-dien-2-one {C38:2 methyl ketone} per unit volume of the water body [particulate >GF/F phase] by filtration and gas chromatography-mass spectrometry	Gas chromatography mass spectrometry (GF/F filtered)	2018-02-27
ME83GCP1	Concentration of octatriaconta-9E,16E,23E-trien-2-one {C38:3 methyl ketone} per unit volume of the water body [particulate >GF/F phase] by filtration and gas chromatography-mass spectrometry	Gas chromatography mass spectrometry (GF/F filtered)	2018-02-27
ESTSED13	Concentration of methyl 2-hydroxy hexadecanoate per unit dry weight of sediment	The amount (mass or moles) of the specified organic compound per unit mass of dry sediment.	2018-02-27
ESTSED14	Concentration of methyl 2-hydroxy heptadecanoate per unit dry weight of sediment	The amount (mass or moles) of the specified organic compound per unit mass of dry sediment.	2018-02-27
ESTSED15	Concentration of methyl 2-hydroxy docosanoate {behenic acid methyl ester CAS 929-77-1} per unit dry weight of sediment	The amount (mass or moles) of the specified organic compound per unit mass of dry sediment.	2018-02-27
	Concentration of methyl 2-hydroxy tetracosanoate {CAS 2433-95-6} per unit dry weight	The amount (mass or moles) of the	2018-



BODC P01 collection

Structured Label based on a semantic model



Search text: Vocabulary: [advanced options](#)

Found 12533 records | Show (11 - 20) |

[download results](#) | [start again](#)

Identifier	PrefLabel	Definition	Date
FA63GCP1	Concentration of methyl hexatriaconta-7E,14E,21E-trienoate {C36:30 methyl ester} per unit volume of the water body [particulate >GF/F phase] by filtration and gas chromatography-mass spectrometry	Gas chromatography mass spectrometry (GF/F filtered)	2018-02-27
ME82GCP1	Concentration of octatriaconta-16E,23E-dien-2-one {C38:2 methyl ketone} per unit volume of the water body [particulate >GF/F phase] by filtration and gas chromatography-mass spectrometry	Gas chromatography mass spectrometry (GF/F filtered)	2018-02-27
ME83GCP1	Concentration of octatriaconta-9E,16E,23E-trien-2-one {C38:3 methyl ketone} per unit volume of the water body [particulate >GF/F phase] by filtration and gas chromatography-mass spectrometry	Gas chromatography mass spectrometry (GF/F filtered)	2018-02-27
ESTSED13	Concentration of methyl 2-hydroxy hexadecanoate per unit dry weight of sediment	The amount (mass or moles) of the specified organic compound per unit mass of dry sediment.	2018-02-27
ESTSED14	Concentration of methyl 2-hydroxy heptadecanoate per unit dry weight of sediment	The amount (mass or moles) of the specified organic compound per unit mass of dry sediment.	2018-02-27
ESTSED15	Concentration of methyl 2-hydroxy docosanoate {behenic acid methyl ester CAS 929-77-1} per unit dry weight of sediment	The amount (mass or moles) of the specified organic compound per unit mass of dry sediment.	2018-02-27
	Concentration of methyl 2-hydroxy tetracosanoate {CAS 2433-95-6} per unit dry weight	The amount (mass or moles) of the	2018-

Conceptual model criteria

- Standardisation with minimum loss of information
- Consistency
- Unambiguous names and definitions
- Meaningful to the human user

BUT

- Semantic rigour necessary to enable machine interpretation
- **Enablement of automated processes must come first**
- Interoperability by application of Semantic Web standards and mapping to authoritative vocabulary digital resources.



How to find a P01 parameter code?

How to find a P01 parameter code?

1. Understand the BODC semantic models including components, structure and logic
2. Analyse what you are trying to map and match to the P01 semantic structure
 1. What is the object of interest?
 2. What is the property?
 3. What is the matrix?
3. Search using your favourite tool or a combination of online tools
4. Use match on external identifiers such as CAS numbers or CheBi IDs or Aphia IDs or ITIS TSN when appropriate
5. This is not easy
 - scientific measurements are complex and extremely diverse
 - attempt to model them can be subjective
 - Optimise for machine-to-machine communication
 - Errors are made so provide feedback by email to vocab.services@bodc.ac.uk

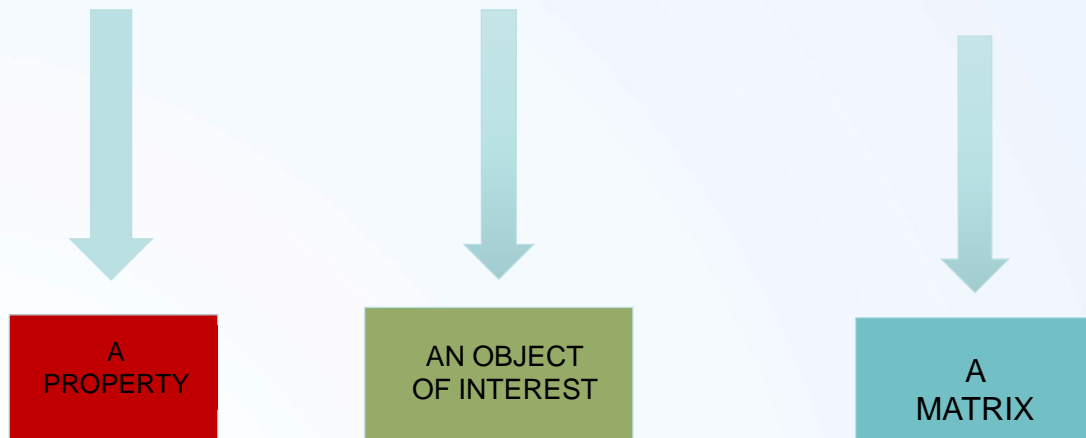
Understand the semantic models

- The Parameter semantic model
 - The Measurement Matrix semantic model
 - Matrix for Physical entities
 - Matrix for other entities
 - The Biological entity semantic model

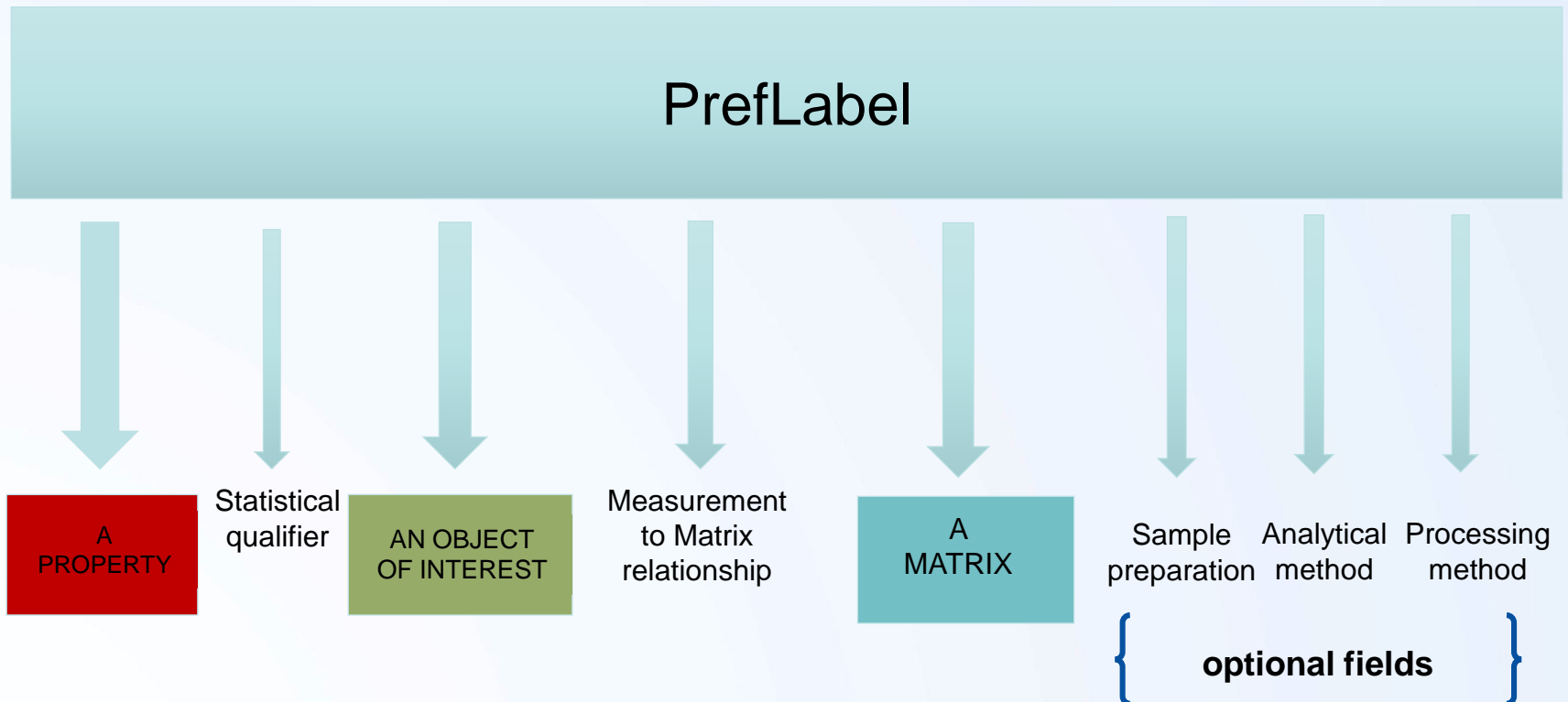
The Parameter semantic model

Parameter semantic model - 3 main elements

prefLabel

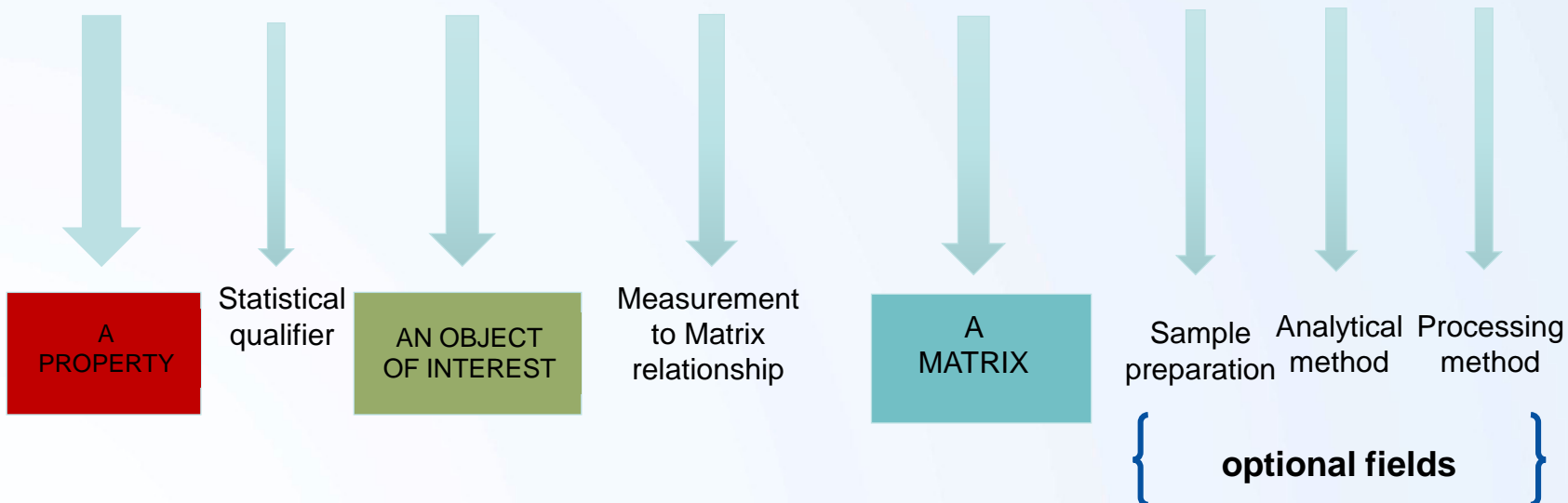


Parameter semantic model - all component elements



Examples

PrefLabel



Concentration of ammonium {NH₄⁺ CAS 14798-03-9} per unit mass of the water body [dissolved plus reactive particulate]

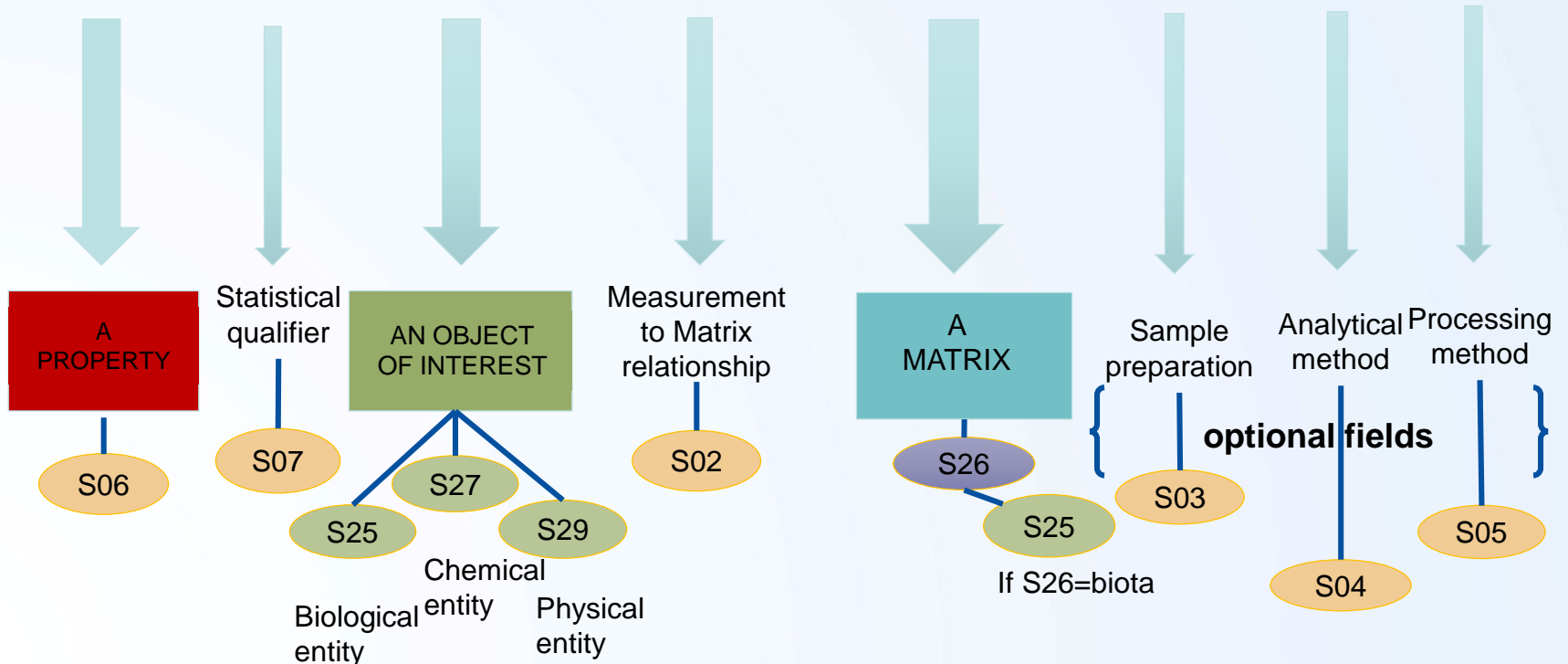
Concentration standard deviation of ammonium {NH₄⁺ CAS 14798-03-9} per unit mass of the water body [dissolved plus reactive particulate]

Rigorous semantics

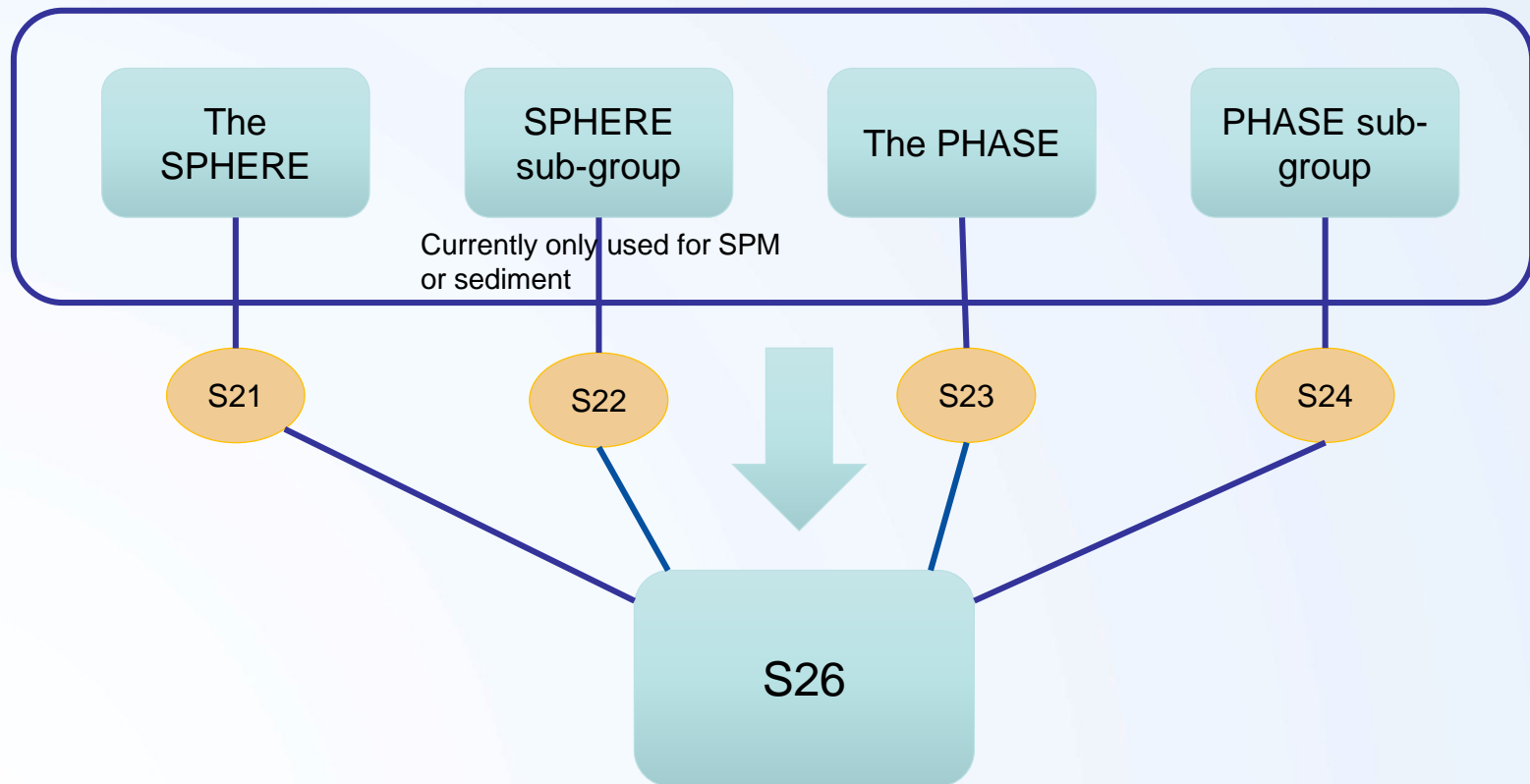
- 1) The parameter entity (i.e. property of interest) = a quantitative or nominative property of an object of interest
- 2) The object of interest = a chemical, biological or physical entity;
- 3) The matrix = the environmental entity to which the chemical, biological or physical entity relates or in which it is embedded;
- The statistical entity allows us to create separate concepts for any statistical parameters associated with the measured property (e.g. standard deviation, uncertainty, standard error, annual mean, etc.).
- If the statistical parameter is set to "not specified" then the reported property is the parameter entity (e.g. "Concentration");
- If it is specified then the reported property is the statistical parameter of the parameter entity (e.g. "Concentration standard deviation")

The Parameter semantic model vocabularies

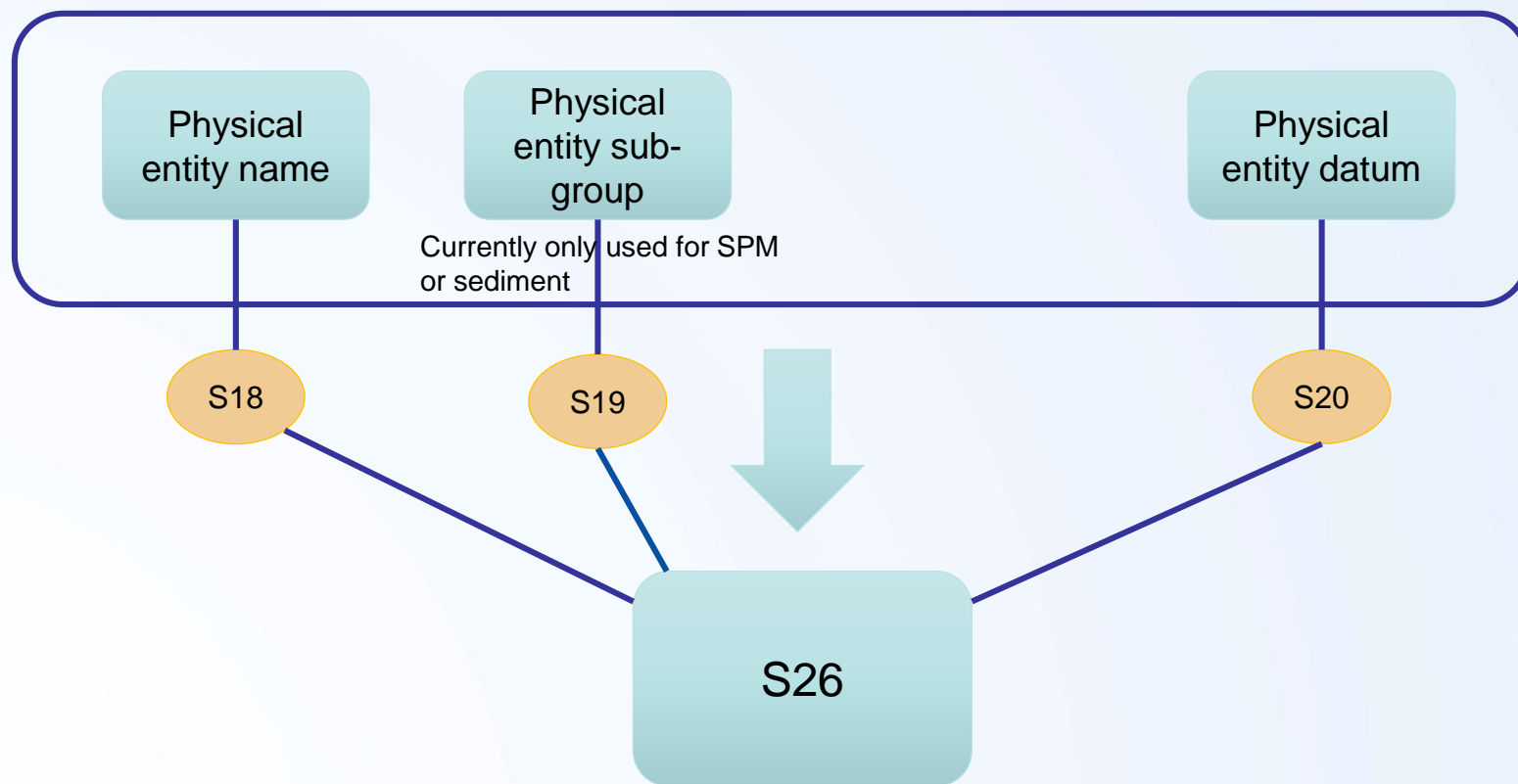
Parameter Semantic Model



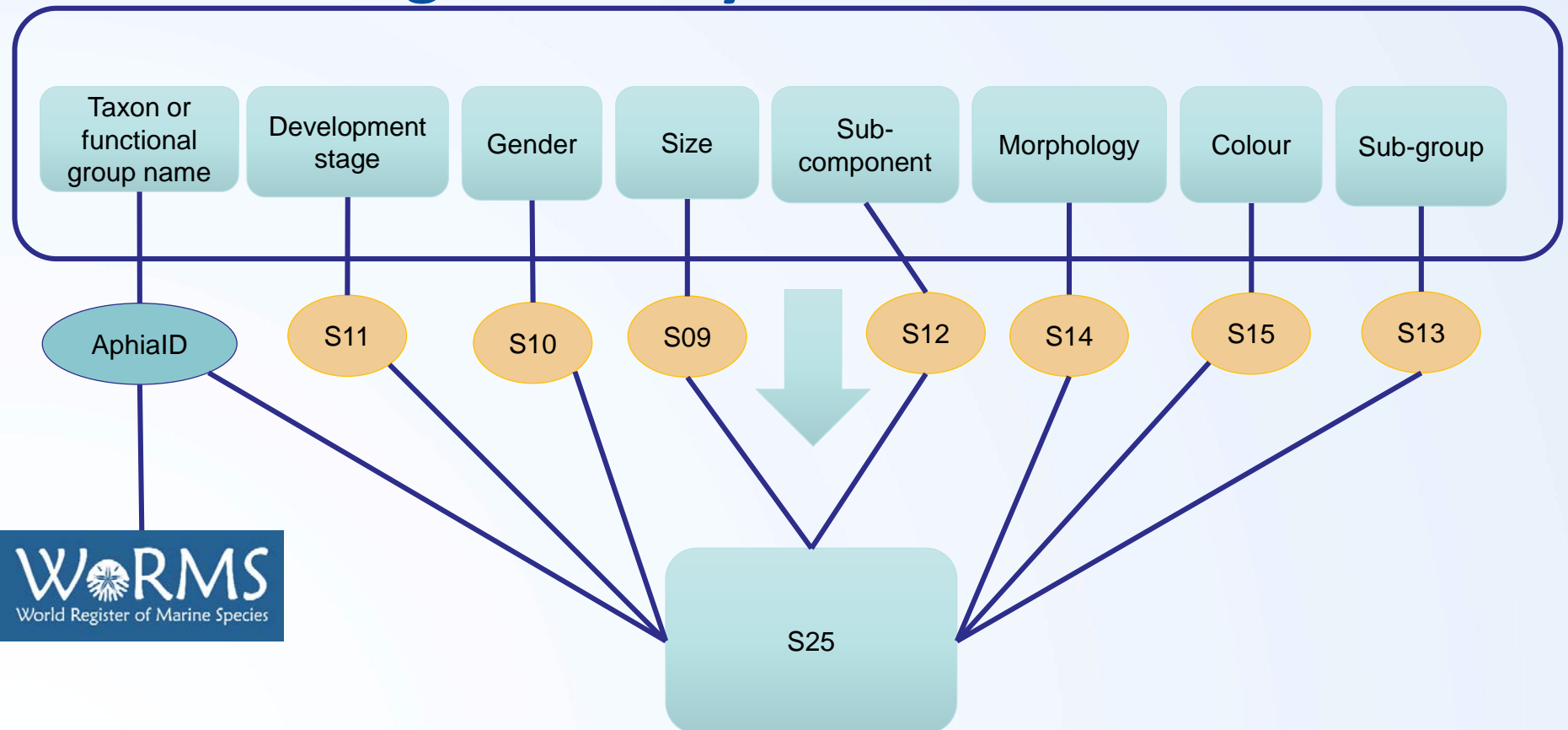
The MATRIX semantic model



The MATRIX semantic model



The Biological Entity semantic model



Mapping the source parameter to a P01 code – key questions

1. What is the object of interest?
2. Is it a chemical substance? A biological entity? A physical entity? Not sure?
3. What kind of quantity or property was measured or reported?
4. In what environment or on what substrate was the observation made?
5. How is the measurement reported? What basis was used? what units?
6. How was the measurement made?
7. Was the sample filtered? If so what was the pore size or filter type?
8. Does the object of interest need to be sub-setted into classes like for example electromagnetic radiation wavelengths or particle size-classes?



First Training, Ostend, 22nd - 27th June 2018

Search Tools

Hierarchical search

- Drill down P08 > P03 > P02 > P01
- http://seadatanet.maris2.nl/v_bodc_vocab_v2/vocab_relations.asp?lib=P08
- Best for
 - - exploring / browsing parameter discovery terms
 - - inexperienced user to get a feel for the classification of parameter codes

Keyword search using the BODC Vocab Search tool

- https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/p01/
- e.g. search for Concentration%chlorophyll-a%water body%spectrophotometry
- Best for
 - - experienced users who know the P01 parameter dictionary and its semantic models
- Main advantages:
 - can search on the identifier, the preferred label, the alternative label or the definition
 - displays the definitions on the screen
 - Option to output full record of selected parameter to a file
 - link to full record and mappings. This is an easy way to find the P02 group code associated with a P01 code for example.

Semantic component facet search using the MARIS tool

- http://seadatanet.maris2.nl/bandit/browse_step.php
- Best for – users who are not very familiar with the P01 parameter dictionary and its semantic construction
- Start with a keyword then drill down the list of possible codes selecting terms relevant to the search

e.g. start by entering chlorophyll-a in the search box

select concentration

P02= Chlorophyll pigment concentrations in water bodies

select the appropriate matrix (e.g. water body [particulate >GF/F phase])

the analytical method if wanted (e.g. HPLC)

Search and submit using the BODC Vocab Builder tool

- https://www.bodc.ac.uk/resources/vocabularies/vocabulary_builder/
- Can be used as a search tool if not logged in
- or search and submit if logged in
- Best for
 - - exploring biological entity, matrix and parameter semantic models
 - - experienced users who want to submit a new P01 code, a new biological entity or a new matrix code or any new terms related to any of these models

P01 Chemical Entity Parameter Code Builder help

Preferred label

Concentration of chlorophyll-a {chl-a CAS 479-61-8} water body [particulate >GF/F phase]

[show/hide exact results](#) | [reset all](#)

Found 17 exact matches

- ✓ Select a measurement property
- Select a measurement statistical qualifier
- ✓ Select a chemical substance
- Select a measurement-matrix relationship
- ✓ Select a matrix

[Return to index](#)

[Login to enable new concept creation](#)



Subjectivity examples

Submitted as:

“Concentration of tar per unit dry weight of sand by sieving and gravimetry”

Became:

“Dry weight of tar residues per unit area of the sediment by sieving, picking and gravimetry”

Biota entries:

“Number Phoronida (brachiopod larvae)”

Became

“Abundance of Brachiopoda (ITIS: 156755: WoRMS 1803) [Stage: larvae] per unit volume of the water body by optical microscopy”

P01=ABR08136

MARIS DEMO 1

Using the MARIS facet search tool

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

Find:

Chlorophyll-a extracted from a water sample, filtered on GF/F, extracted in acetone, analysed by fluorometry and expressed in micrograms per litre

MARIS DEMO 1

Main keyword=Chlorophyll-a > enter in input string box

Property > concentration

Matrix > water body [particulate >GF/F phase]

Analytical method > fluorometry

Sample preparation > filtration, acetone extraction

Measurement-Matrix relationship > per unit volume of the

BODC DEMO 1

Use the BODC keyword search tool

https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/p01/

Use the Advanced search

Find:

Chlorophyll-a extracted from a water sample,
filtered on GF/F, extracted in acetone, analysed by
fluorometry and expressed in micrograms per litre



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

BODC DEMO 1

Concentration of chlorophyll-a per unit
volume of water
by GF/F acetone fluorometry

Be Aware

- Ensure use of the “%” as a wildcard
- Use the wildcard to account for the plural forms
- A P01 Label always starts with an Upper case
- The spelling is British English spelling

Using the BODC Vocab Builder

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

P01 Chemical Entity Parameter Code Builder

Preferred label

Concentration not specified of chlorophyll-a {chl-a CAS 479-61-8} per unit volume of the water body [particulate >GF/F phase]

[show/hide exact results](#) | [show/hide close match results](#) | [reset all](#)

Found 13 exact matches | Showing (1 - 10) |

Submit New BODC Code DEMO 1

You need a standard deviation code for:

Concentration of chlorophyll-a {chl-a CAS 479-61-8}
per unit wet volume of sediment by acetone
extraction and spectrophotometry

<http://vocab.nerc.ac.uk/collection/P01/current/COCHSE04/>

**Concentration of chlorophyll-a {chl-a CAS 479-61-8}
per unit wet volume of sediment by acetone
extraction and spectrophotometry**

Additional Mappings

ICES Codes to BODC P01

ICES: <http://vocab.ices.dk/>

PARAM:SPECI:MATRX:SEXCO:BASIS

HG:Pleuronectes platessa: LI: F: W

Mercury:127143:Liver:Female:Wet weight

BODC = HGMMCF21

Concentration of mercury {Hg} per unit wet weight of biota

{Pleuronectes platessa (ITIS: 172902: WoRMS 127143) [Sex: female

Subcomponent: liver]}

Examples – MARIS and BODC

#1 Exercise to practice selection of complex matrix/phases

Find the P01 and P02 codes for a dataset containing size-fractionated chlorophyll-a extracted in acetone and analysed by fluorometry reported in $\mu\text{g/l}$ for the following size fractions:

1. 0.2-2,
2. 2-20,
3. 20-200
4. >0.2 microns obtained by summing the concentrations in the size-fractions.

Examples – MARIS and BODC

#2 Exercise to practice finding a suite of related measurements and illustrate use of synonyms or multiple accepted spellings.

Find the P01 and P02 codes for parameters associated with an HPLC pigment speciation dataset. The pigments to be mapped are chl-a, divinyl chlorophyll-a, alpha-carotene, and pheophytin a. The samples were filtered on 0.2um nuclepore membrane, extracted in acetone, analysed by HPLC reported in units of micrograms and picograms per litre.

Examples – MARIS and BODC

#3 Exercise

Find the P01 and P02 codes for chlorophyll concentrations measured using an in situ fluorometer deployed on a CTD profiling package

#4 Exercise

Find the P01 and P02 codes for a nutrient dataset for silicate, nitrate plus nitrite, nitrite, phosphate concentrations measured on unfiltered water samples

Answers

#1 Exercise

- <http://vocab.nerc.ac.uk/collection/P01/current/SCHLFLPF/>
- <http://vocab.nerc.ac.uk/collection/P01/current/SCHLFLPG/>
- <http://vocab.nerc.ac.uk/collection/P01/current/SCHLFLPH/>
- <http://vocab.nerc.ac.uk/collection/P01/current/CPHLFLP4/>

- <http://vocab.nerc.ac.uk/collection/P02/current/CPWC/>

Answers

#2 Exercise

Tip: if using the BODC search tool, list the 61 pigment category types corresponding to concentration per unit volume of the water body [particulate >0.2um phase] by filtration, acetone extraction and high performance liquid chromatography (HPLC); then select the appropriate one from that by scrolling down or extracted to a file.

<http://vocab.nerc.ac.uk/collection/P01/current/CPHLHPP5/> P02=CPWC

<http://vocab.nerc.ac.uk/collection/P01/current/ACARHPP5/> or

<http://vocab.nerc.ac.uk/collection/P01/current/BECAHPP5/> P02=CARO

<http://vocab.nerc.ac.uk/collection/P01/current/DVCAHPP5/> P02=CPWC

<http://vocab.nerc.ac.uk/collection/P01/current/PTAXHPP5/> P02=PHWC

Answers

#3 Exercise

Tip: if using the BODC Vocab search tool use the additional filter to exclude anything that contains extraction. This gives you 25 returns. Simply scrolled down the list to find the one that fits your data.

In this case the preferred option is to go for the generic code for in situ fluorometer data:

<http://vocab.nerc.ac.uk/collection/P01/current/CPHLPR01/>

Answers

#4 Exercise (**N.B.** 2nd option is the wider option)

<http://vocab.nerc.ac.uk/collection/P01/current/SLCAMS01> P02=SLCA

Or <http://vocab.nerc.ac.uk/collection/P01/current/SLCAZZXX/> P02=SLCA

<http://vocab.nerc.ac.uk/collection/P01/current/NTRZYDZ> P02=NTRA

Or <http://vocab.nerc.ac.uk/collection/P01/current/NTRZZZXX/> P02=NTRA

<http://vocab.nerc.ac.uk/collection/P01/current/NTRIYYDZ> P02=NTRI

Or <http://vocab.nerc.ac.uk/collection/P01/current/NTRIZZXX/> P02=NTRI

<http://vocab.nerc.ac.uk/collection/P01/current/PHOSYYDZ> P02=PHOS

Or <http://vocab.nerc.ac.uk/collection/P01/current/PHOSZZXX/> P02=PHOS