

# The BODC Parameter Usage Vocabulary (PUV) semantic model exposed

Gwenaëlle Moncoiffé  
Alexandra Kokkinaki

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



**National  
Oceanography Centre**  
NATURAL ENVIRONMENT RESEARCH COUNCIL



**British Oceanographic  
Data Centre**  
NATURAL ENVIRONMENT RESEARCH COUNCIL

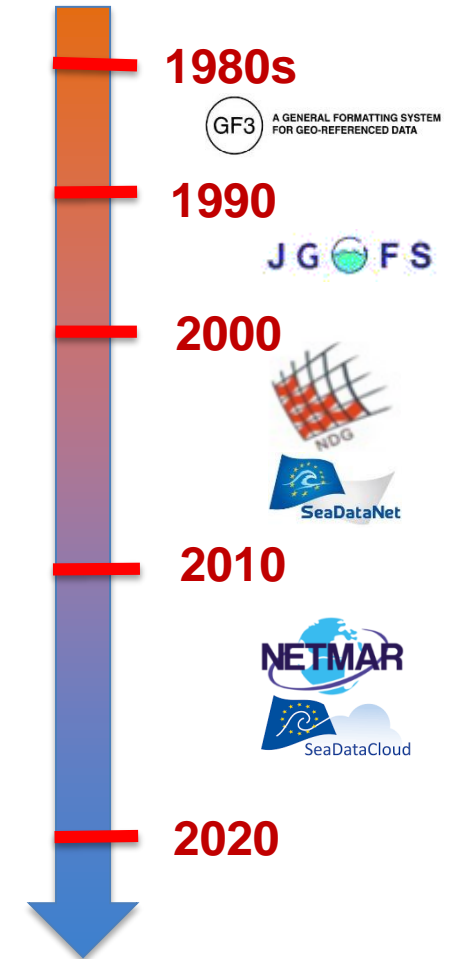
**NERC** SCIENCE OF THE  
ENVIRONMENT

# 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

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

Found 12533 records | Show ( 11 - 20 ) | [« First](#) [< Prev](#) | [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [Next >](#) [Last »](#) [download results](#) | [start again](#)

Identifie	PrefLabel	Definition	Date
<a href="#">FA63GCP1</a>	Concentration of methyl hexatriaconta-7E,14E,21E-trienoate {C36:3O 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
<a href="#">ME82GCP1</a>	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
<a href="#">ME83GCP1</a>	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
<a href="#">ESTSED13</a>	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
<a href="#">ESTSED14</a>	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
<a href="#">ESTSED15</a>	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 7433-95-6} per unit dry weight	The amount (mass or moles) of the	2018-

# Core elements

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

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



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



a **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/](https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/S06/)



# Statistical element

Properties can be associated with a statistical term to become

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

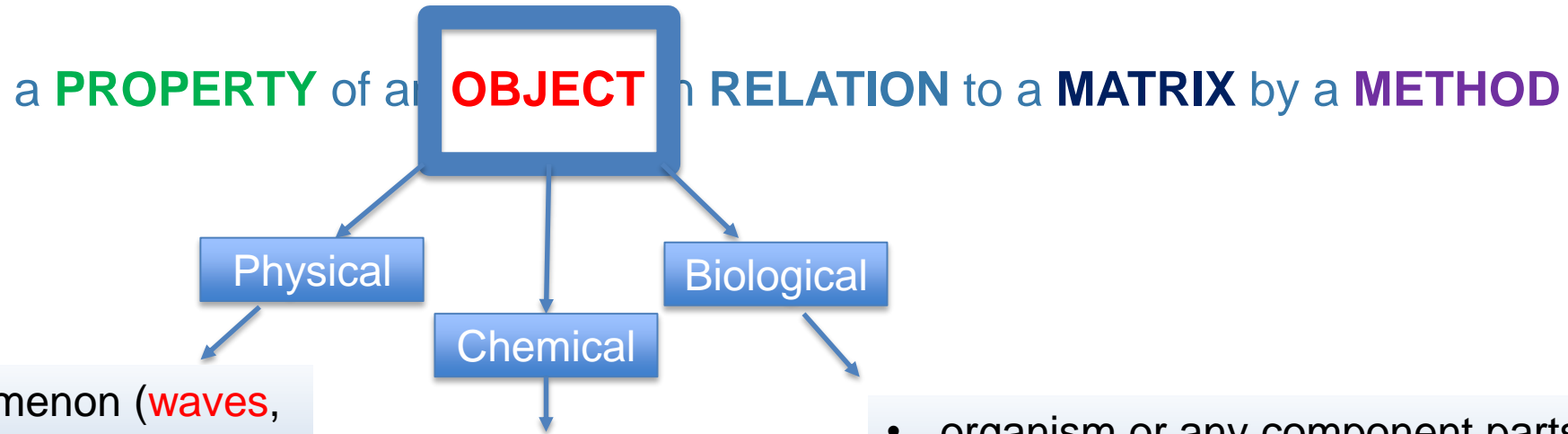
- examples: standard deviation, mean, annual mean, hourly mean, etc.
- STATISTIC is defined in S07

<http://vocab.nerc.ac.uk/collection/S07/current/>

[https://www.bodc.ac.uk/resources/vocabularies/vocabulary\\_search/S07/](https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/S07/)



# The object element



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

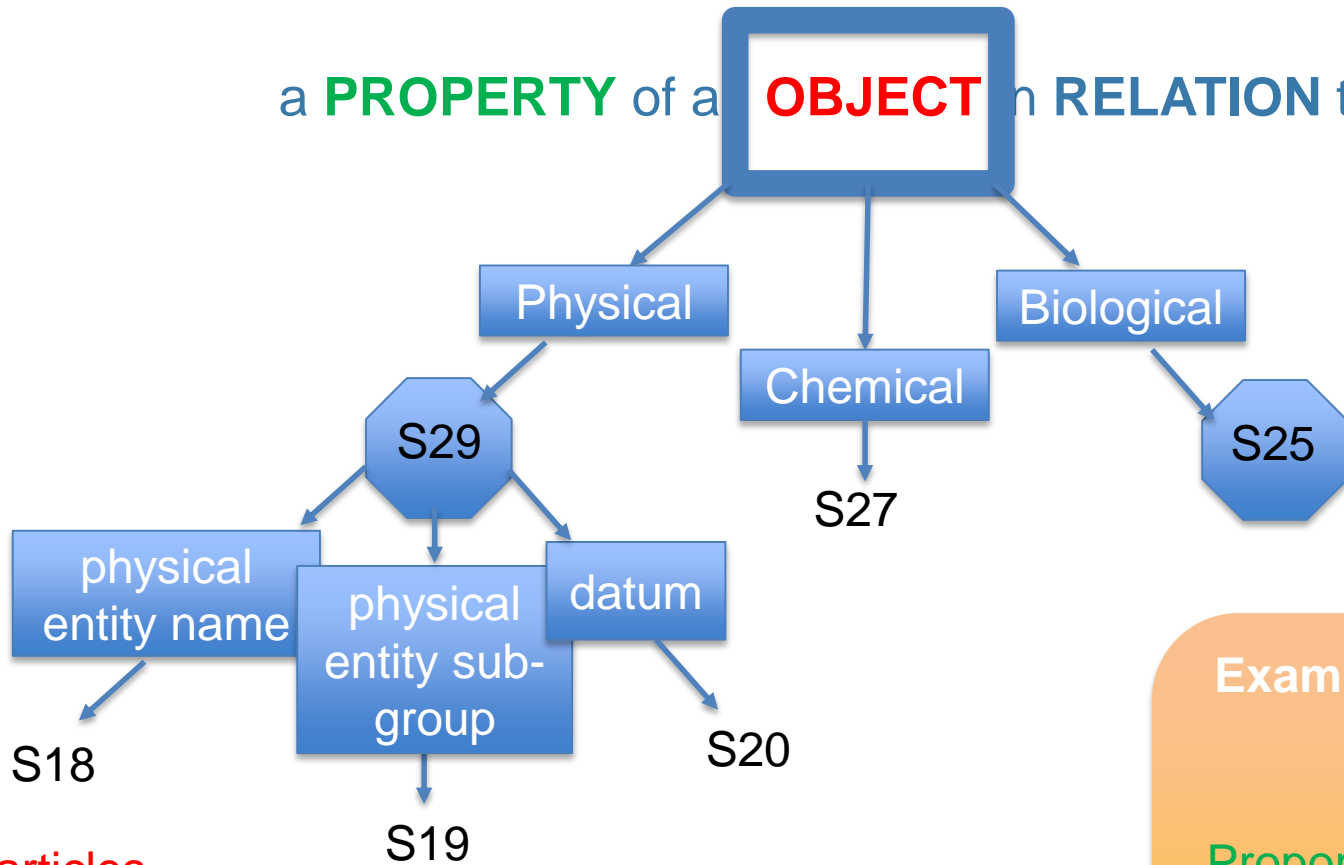
- substance
- group of substances
- chemical element

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





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



particles

+

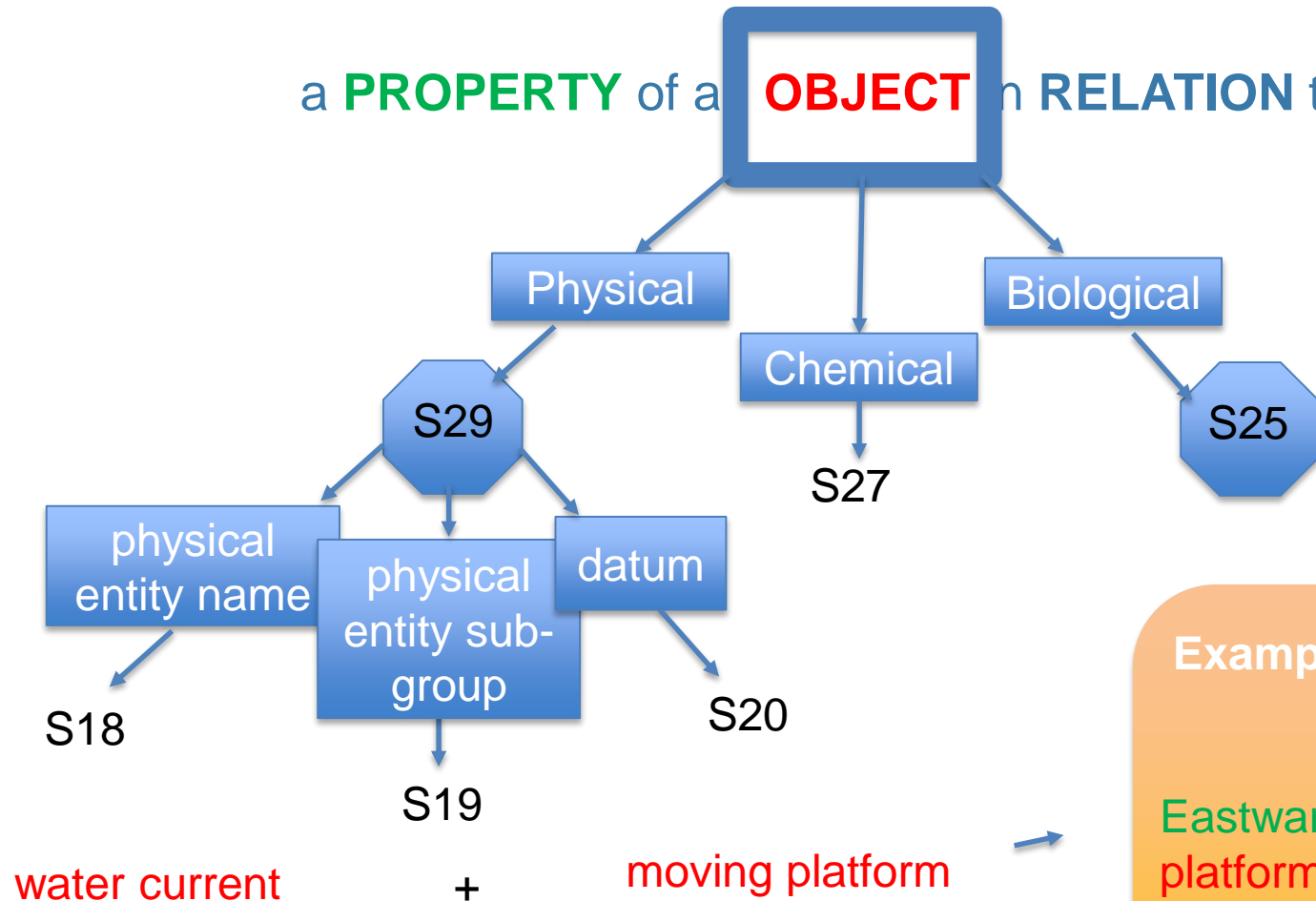
180-300um

### 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**



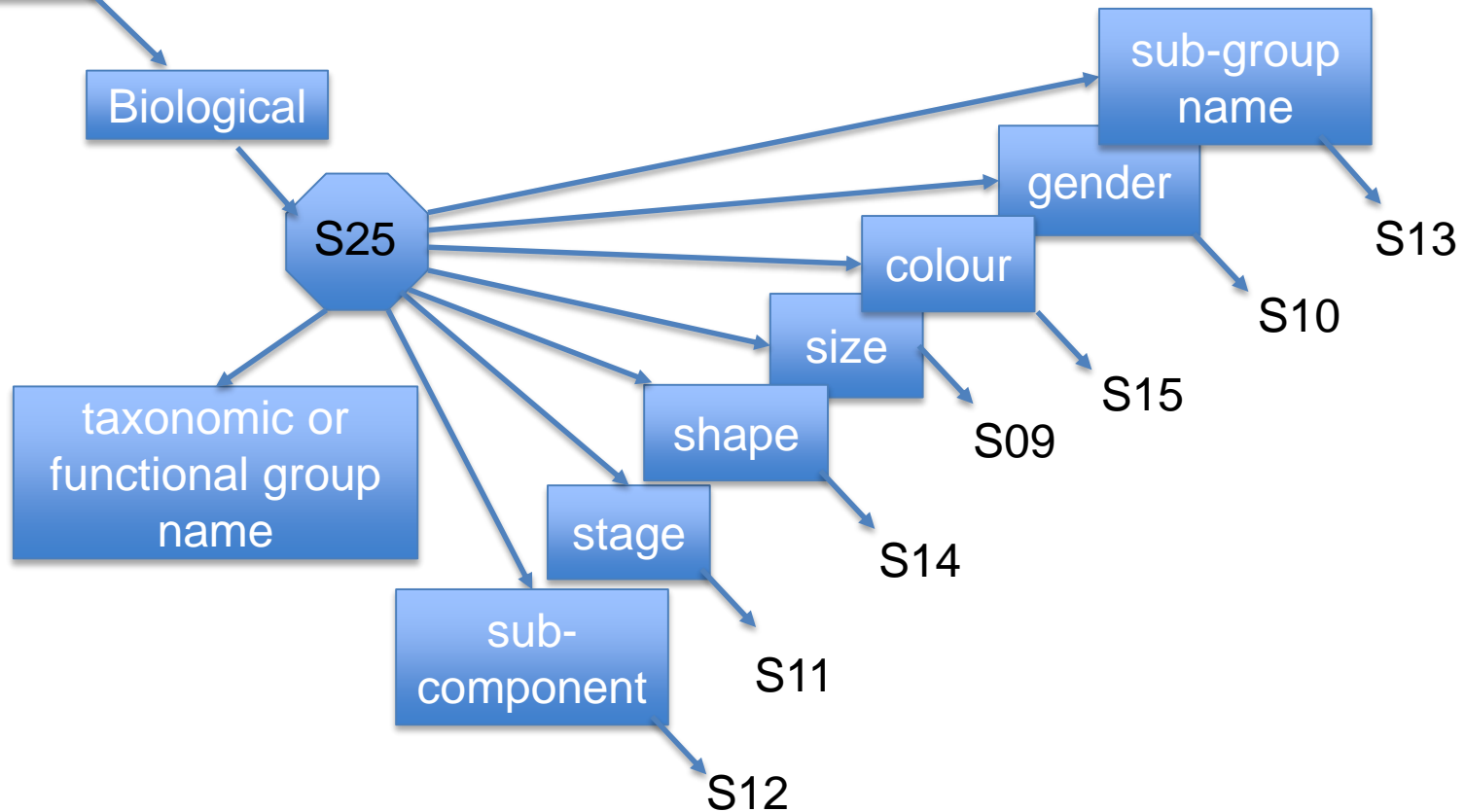
Example of P01 label

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/>



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



### Example of S25 label

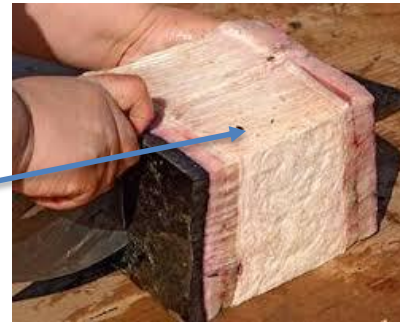
Halichoerus grypus (ITIS: 180653:  
WoRMS 137080) [Stage: post-  
weaned pup Sex: female  
Subcomponent: blubber  
Subgroup: dead]



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

Example of S25 label

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



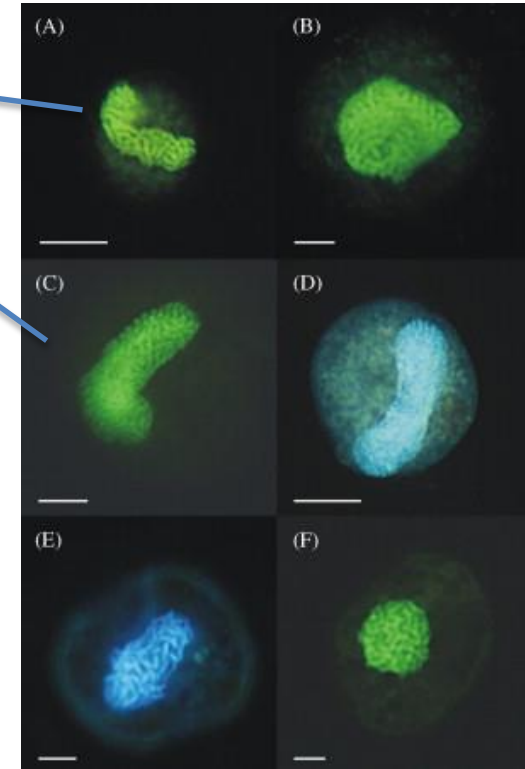
<http://vocab.nerc.ac.uk/collection/S25/current/BE006418/>

Unidentified autotrophic dinoflagellate (banana-shape)

Example of S25 label

Dinophyceae (ITIS: 9874; WoRMS 19542)

[Morphology: banana-shaped Subgroup: autotrophic]



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**

**OBJECT**

Biological

S25 = "biological entity specified elsewhere"

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

BIOLOGY



EMODnet

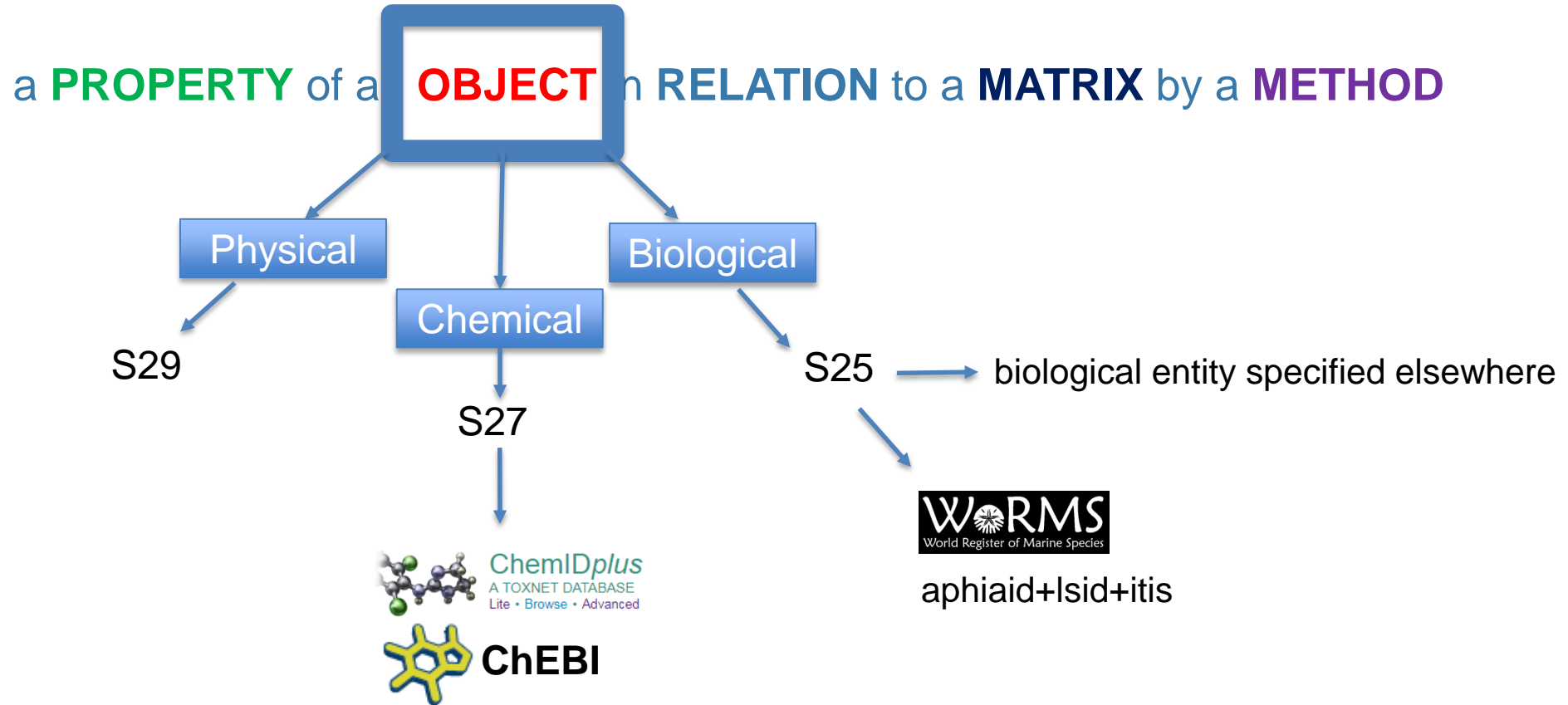
European Marine Observation and Data Network



OCEAN BIOGEOGRAPHIC INFORMATION SYSTEM



# Links to authoritative name registries



# The matrix element

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

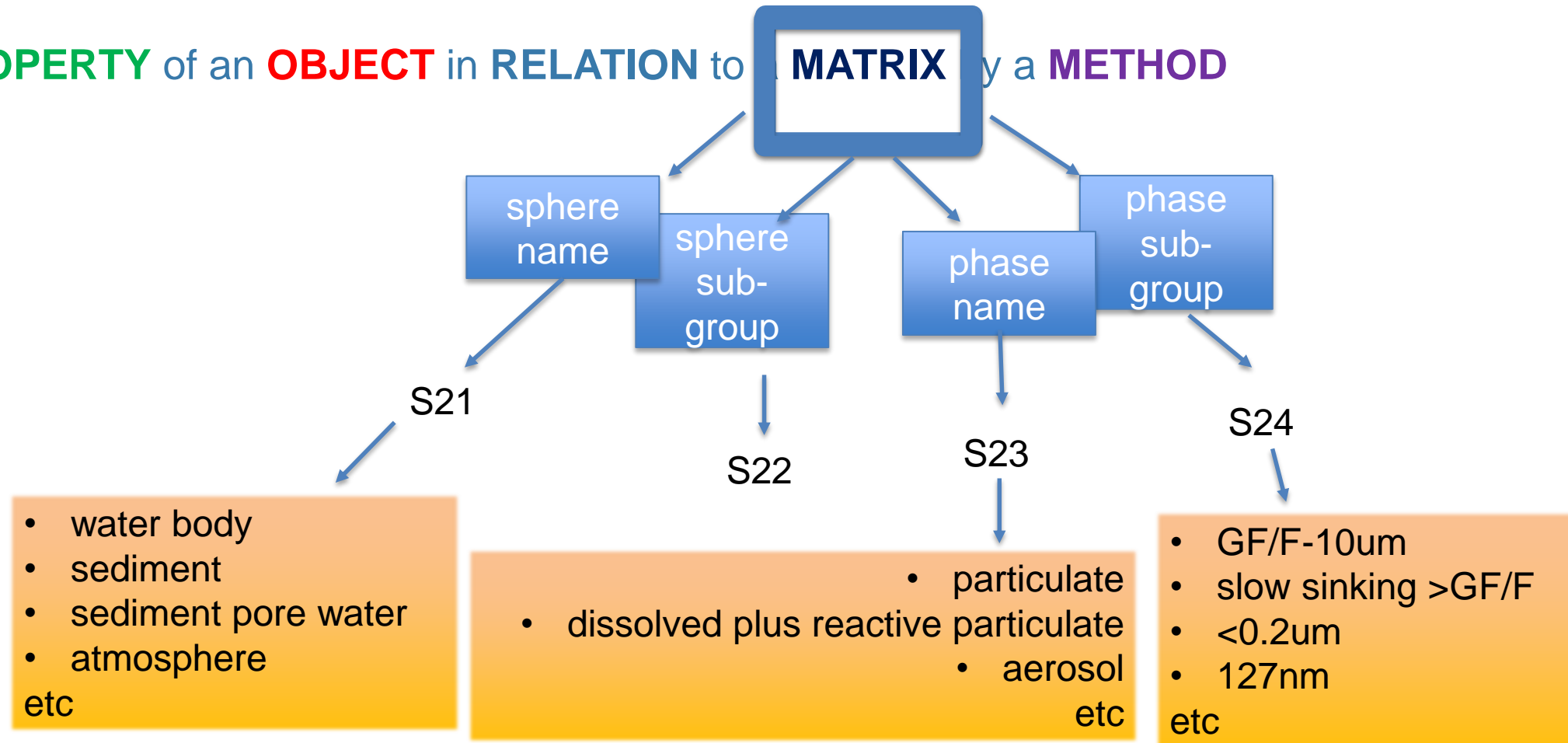
MATRIX is the environment in which the measurement is made or in which the object of interest is embedded

- MATRIX is defined in S26
- MATRIX is a structured compound vocabulary





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



# Matrix biota variant

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

## 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
- 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

Count of *Halichoerus grypus* (ITIS: 180653: WoRMS 137080) out of the water body

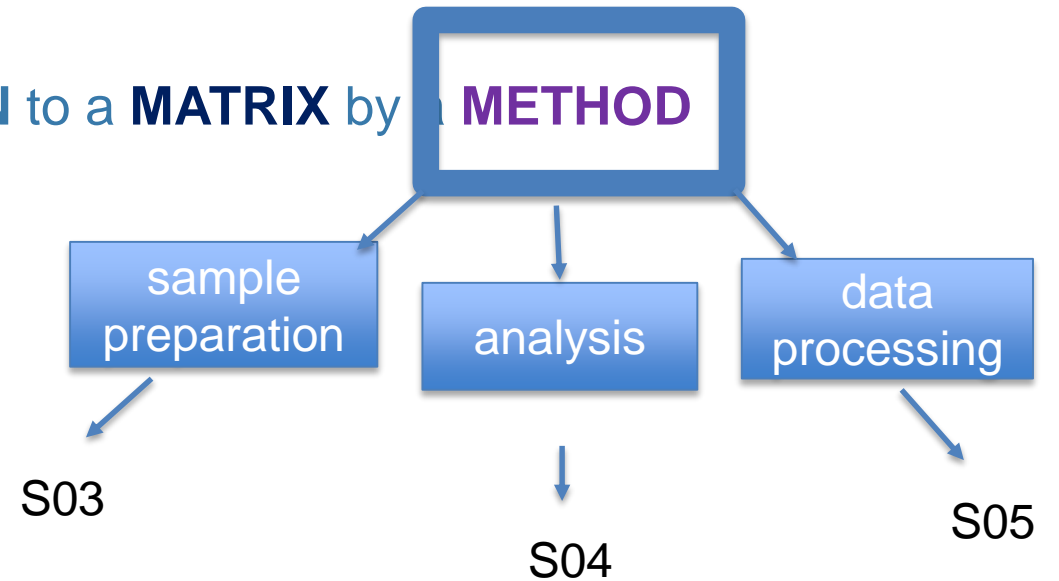
Count of *Halichoerus grypus* (ITIS: 180653: WoRMS 137080) in the water body

Examples of P01 labels



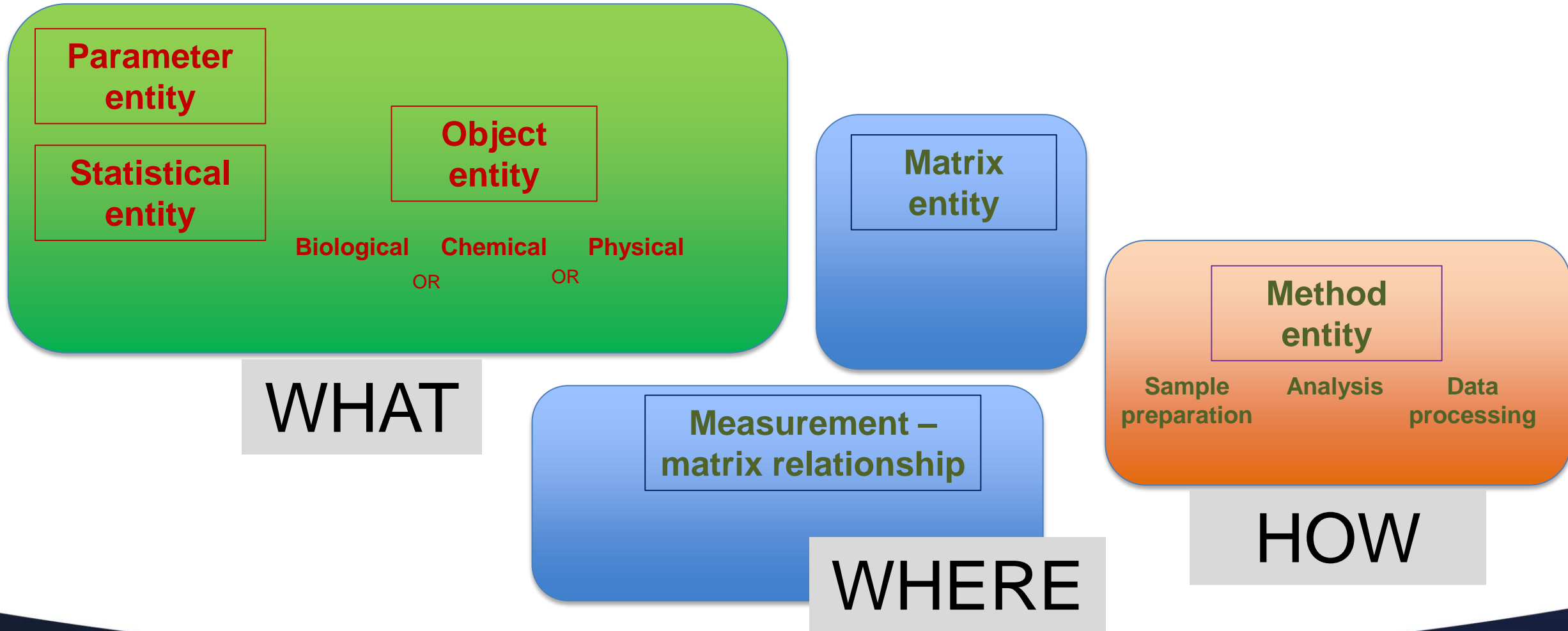
# The method element

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



- 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



# 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/

## NERC Vocabulary Fuseki Query

Dataset: No session

### SPARQL Query

```
prefix skos:<http://www.w3.org/2004/02/skos/core#> prefix rdf:
<http://www.w3.org/1999/02/22-rdf-syntax-ns#> prefix owl:
<http://www.w3.org/2002/07/owl#> prefix dc:
<http://purl.org/dc/terms/>

select distinct (?dci as ?Identifier) (?pl as ?PrefLabel) (?defx
as ?Definition) (?ver as ?Version) (?sr as ?related) ?Date (?dt as
?Url) where {<http://vocab.nerc.ac.uk/collection/P01/current/>
skos:member ?dt .
FILTER(regex(str(?pl),"Concentration of ammonium.per unit volume
```

Output:

If XML output, add XSLT style sheet (blank for none):

Force the accept header to text/plain regardless.





# Tools using the semantic model

a SOAP Application Programming Interface (API) for exchanging structured information across computer networks as the result of calls. It relies upon XML (eXtensible Markup Language) documents for passing messages. Moreover there is an operational SPARQL endpoint for machine interaction.

- [More about the Vocabulary Web services](#)
- [Query the NVS Common Vocabularies via the SeaDataNet interface](#)
- [Query the NVS Common Vocabularies via the BODC interface](#)
- [Semantic Model Vocabulary Builder for the P01 vocabulary](#)
- [P01 Vocabulary - Facet Search on Semantic Components](#)

BODC Vocab builder tool

MARIS Facet search tool

## CONTENT GOVERNANCE

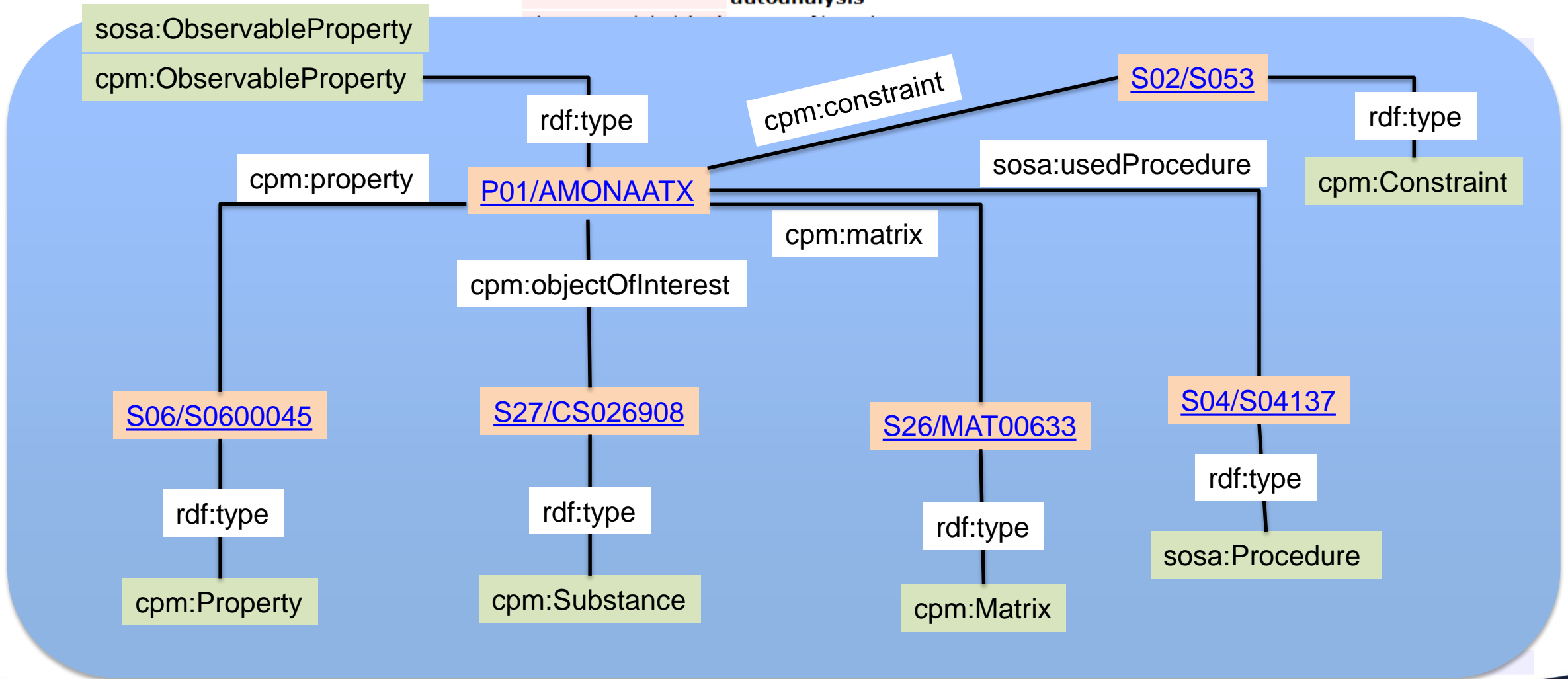
Content governance of the vocabularies is very important and is done by a combined SeaDataNet and MarineXML Vocabulary Content Governance Group (SeaVoX), moderated by **BODC**, and including experts from SeaDataNet, MMI, MOTIIVE, JCOMMOPS and more international groups. SeaVox discussions are based on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

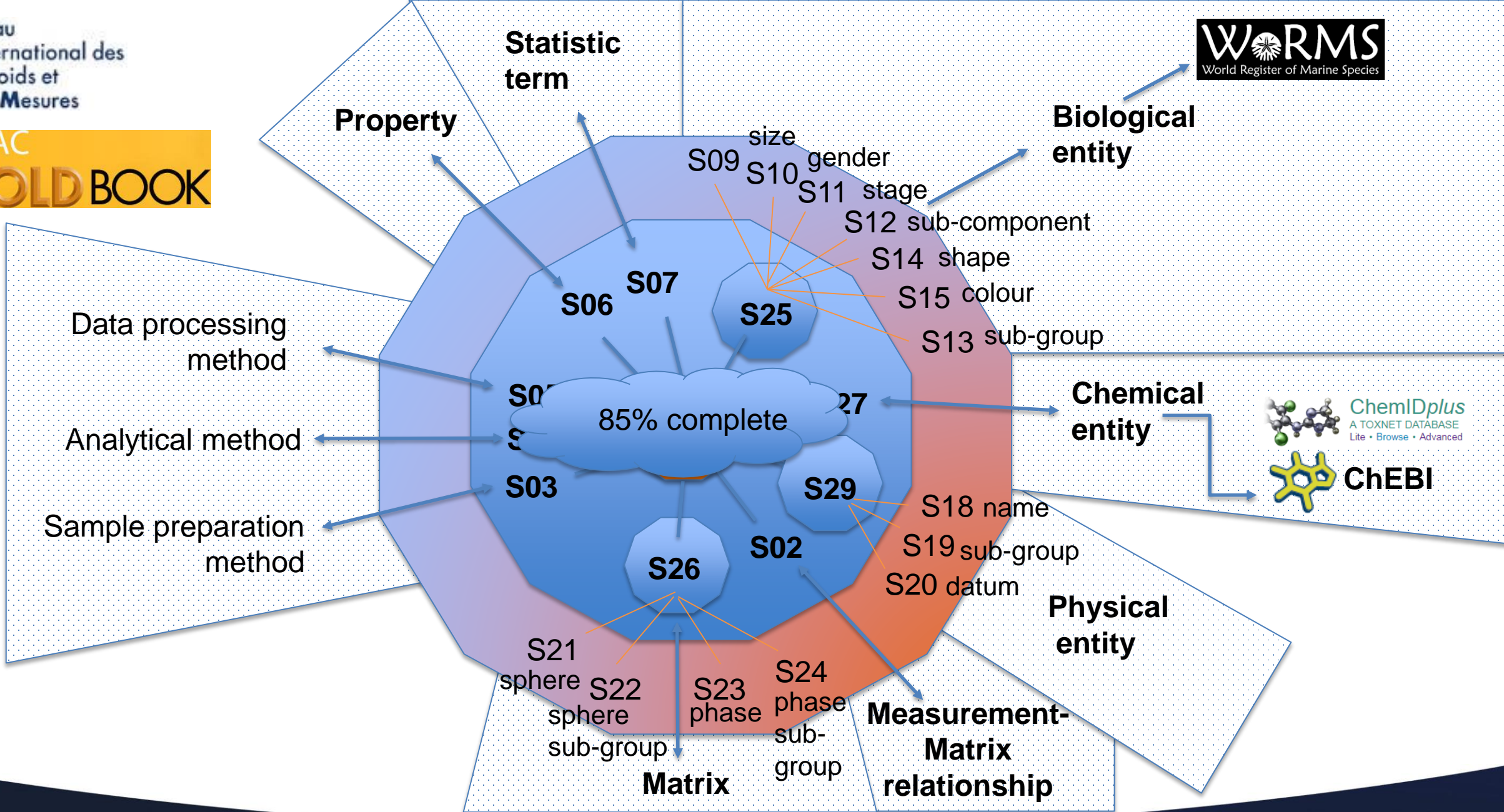
## NVS2.0



# Ontology expression

URI	<a href="http://vocab.nerc.ac.uk/collection/P01/current/AMONAATX/">http://vocab.nerc.ac.uk/collection/P01/current/AMONAATX/</a>
Identifier ( )	SDN:P01::AMONAATX
Preferred label (en)	Concentration of ammonium {NH4+ CAS 14798-03-9} per unit volume of the water body [dissolved plus reactive particulate phase] by colorimetric autoanalysis





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](mailto:vocab.services@bodc.ac.uk)

**GitHub**

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



**National  
Oceanography Centre**  
NATURAL ENVIRONMENT RESEARCH COUNCIL



**British Oceanographic  
Data Centre**  
NATURAL ENVIRONMENT RESEARCH COUNCIL

**NERC** SCIENCE OF THE  
ENVIRONMENT