

# 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



**National  
Oceanography Centre**  
NATURAL ENVIRONMENT RESEARCH COUNCIL



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

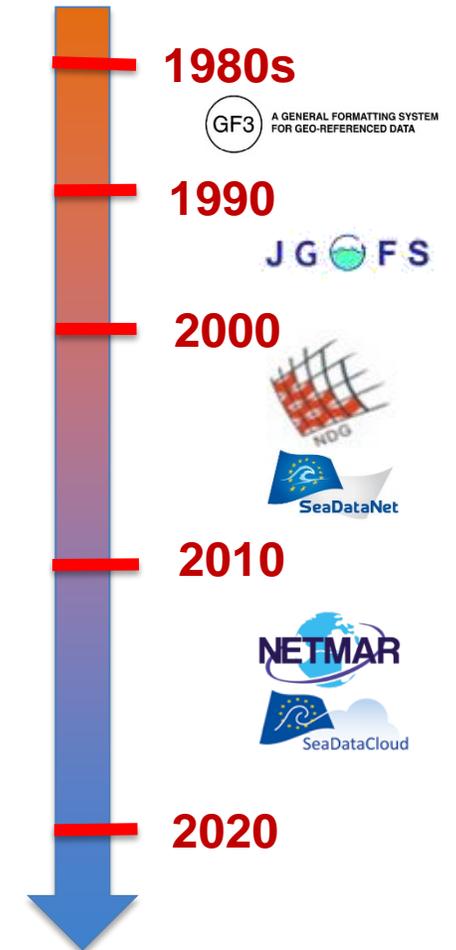
**NERC** SCIENCE OF THE  
ENVIRONMENT

# 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
- 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

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-

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.

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**

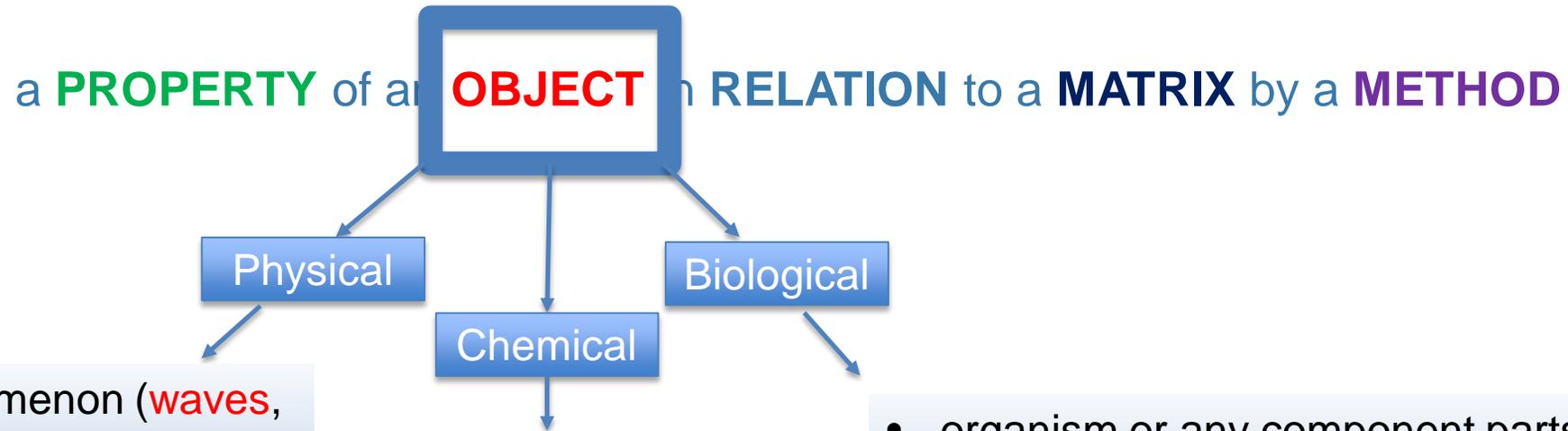
- quantitative (Concentration, Practical salinity, Production rate, Abundance)
- qualitative: binary (Presence or absence), ordinal (Abundance category), or nominal (Colour class, Shape class)
- All PROPERTY terms are defined in collection **S06**

<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/)



# The object element



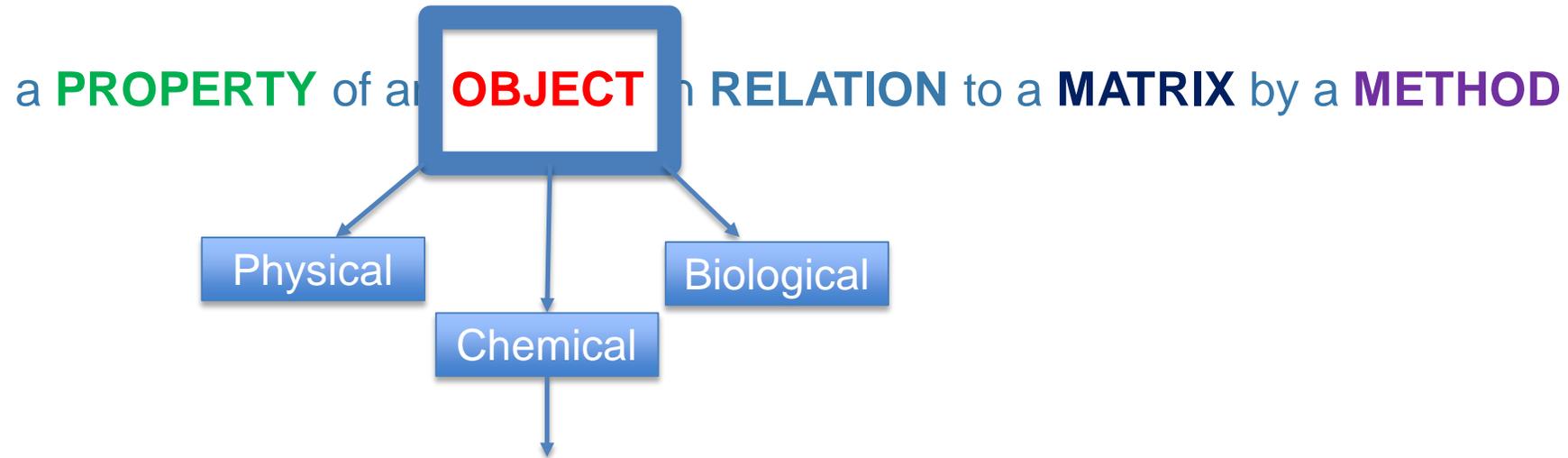
- phenomenon (**waves**, **wind**)
- object (**measurement platform**, **particles**)
- element (**water**, **air**)

- substance
- group of substances
- chemical element

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



# The chemical entity object



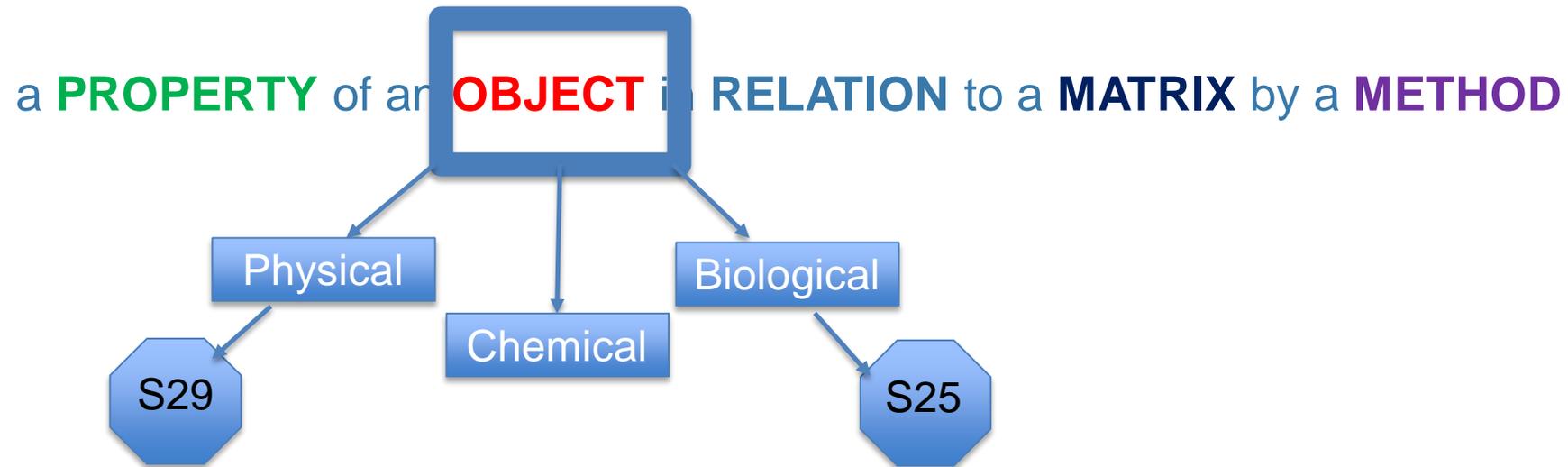
Chemical objects referred to in a P01 label are defined in collection S27

<http://vocab.nerc.ac.uk/collection/S27/current/>

[https://www.bodc.ac.uk/resources/vocabularies/vocabulary\\_search/S27/](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 a **OBJECT** in **RELATION** to a **MATRIX** by a **METHOD**



Physical

S29

physical entity name

physical entity sub-group

datum

S18

e.g. particles

S19

e.g. 180-300um

S20

e.g. not applicable



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/>



National Oceanography Centre  
NATURAL ENVIRONMENT RESEARCH COUNCIL



British Oceanographic Data Centre  
NATURAL ENVIRONMENT RESEARCH COUNCIL

NERC SCIENCE OF THE ENVIRONMENT

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



Physical

S29

physical entity name

physical entity sub-group

datum

S18

S20

S19

water current

moving platform

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.

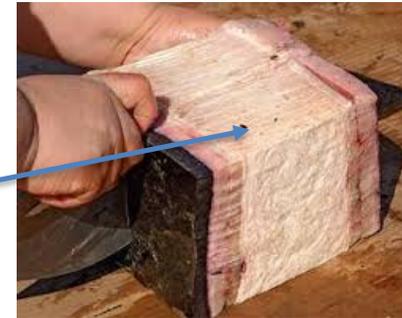
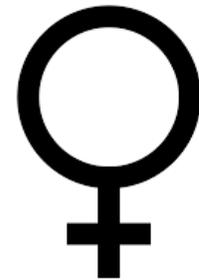
<http://vocab.nerc.ac.uk/collection/P01/current/LREWAS01/>





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

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)

Dinophyceae (ITIS: 9874: WoRMS 19542)

[Morphology: banana-shaped Subgroup: autotrophic]

Example of S25 label

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/>



National  
Oceanography Centre  
NATURAL ENVIRONMENT RESEARCH COUNCIL



British Oceanographic  
Data Centre  
NATURAL ENVIRONMENT RESEARCH COUNCIL

NERC SCIENCE OF THE  
ENVIRONMENT

# 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

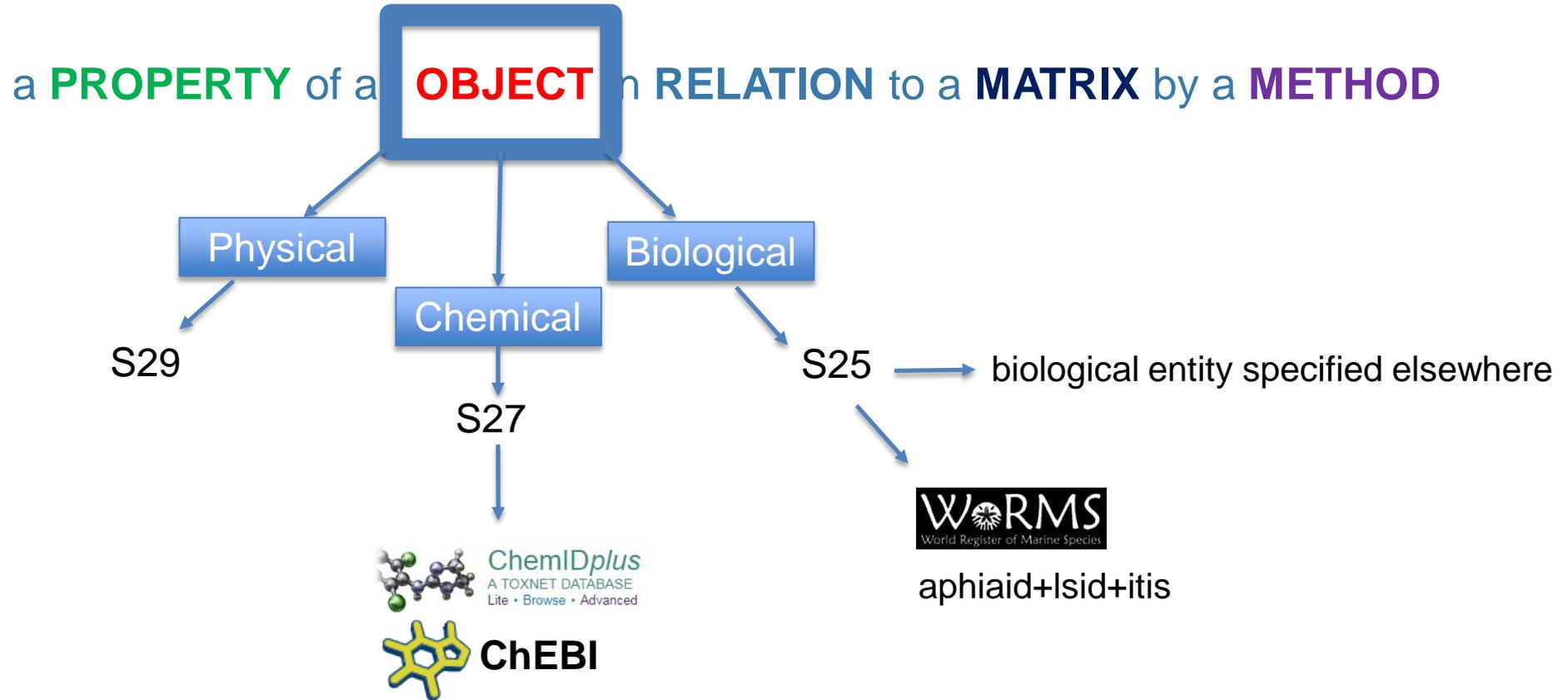


EMODnet



 **OBIS** OCEAN BIOGEOGRAPHIC INFORMATION SYSTEM

# Links to authoritative name registries



# The matrix element

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

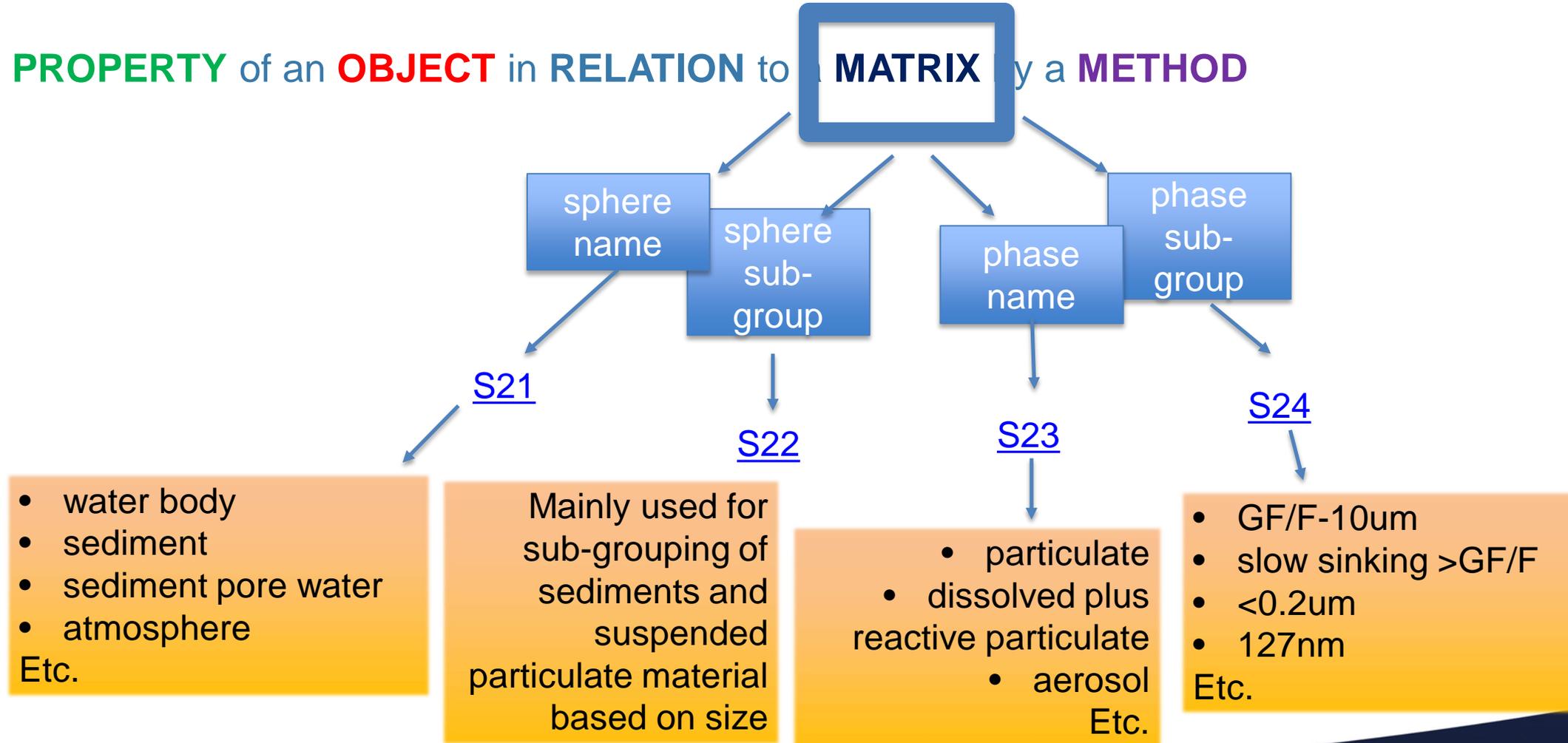
MATRIX is the environment in which the measurement is made or in which the object of interest is embedded

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



# The matrix element

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



# The biota matrix 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 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](#)

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

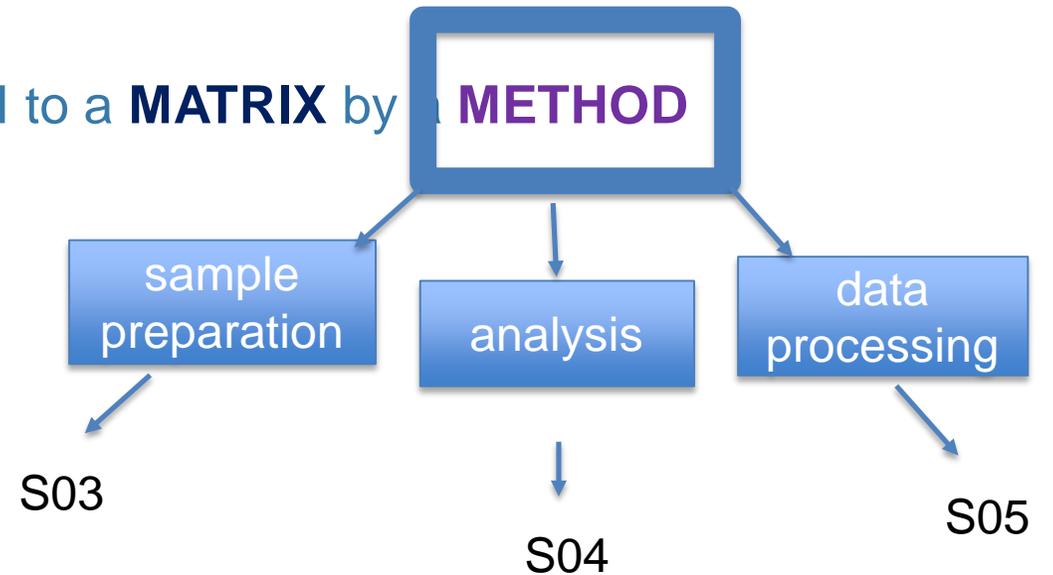
Count of *Halichoerus grypus* (ITIS: 180653: WoRMS 137080) in the water body

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 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)

# The statistical element

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/](https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/S07/)



# 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 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.

