

Towards Interoperability for Observed Parameters: Position Statement of an Emerging Working Group

Presented by Barbara Magagna
Environmental Agency Austria
barbara.magagna@umweltbundesamt.at

ICEI2018, Semantics for biodiversity and ecosystem research
September 27 2018





Barbara
Magagna

- Semantic Analyst and Data Manager at Environmental Agency Austria
- Co-developer of SERONTO (Socio-Ecological Research and Observation oNTology)
- Coordinator of EnvThes (LTER-Europe)
- Currently involved in ENVRIplus data theme (co-developer of RM and coordinator of provenance task)

Introducing me

- *RDA*: Research Data Alliance – international community driven organization to enable open sharing and re-use of data
- *VSSIG*: RDA Vocabulary and Semantic Services Interest Group lead by S. Cox, A. Shepherd, Y. L. Franc, J. Graybeal
- *Harmonizing measurement parameters*: VSSIG task group lead by Barbara Magagna and Michael Diepenbroek, targeting to become a working group

RDA – VSSIG task group

Involved Stakeholders:

- Research Infrastructures (RIs):
 - LTER: Barbara Magagna, Alessandro Oggioni, Philip Trambeth
 - ILTER: Kirstin Vanderbuilt
 - ICOS: Harry Lankreijer
 - AnaEE: André Chanzy, Christian Pichot
 - LifeWatch: Nicola Fiore, Ilaria Rosati, Paolo Tagliolato
 - AquaDiva: Friederike Klan,
- Data Centers:
 - PANGAEA: Michael Diepenbroek, Robert Huber, Melanie Buss, Uwe Schindler
 - BODC: Alexandra Kokkinaki, Gwen Moncoiffé
 - GFBio: Naouel Karam
- Technologies/Ontologies:
 - BioPortal: John Graybeal
 - ENVO: Pier Luigi Buttigieg, Chris Mungall
 - ePIC: Ulrich Schwardmann
- Others:
 - TIB: Markus Stocker
 - DLR: Sirko Schindler

Task Members

- Develop best practices and a generally accepted model for scientific observation and measurement parameters including possibly also measurement methods and devices by using agreed core terminologies
- Use: annotate research data with identified, unambiguous and machine readable vocabulary for describing data semantics
- Objective: improve interoperability for data discovery and data integration

Scope of the task group

Analysing ecological phenomena *across geographic, temporal, biological scales* requires a variety of existing data sets.

Observational data are often represented in *tabular form* but differ in:

- number of attributes,
- the relationships implied between attributes
- the coding conventions used for representing information within data sets.

The challenge

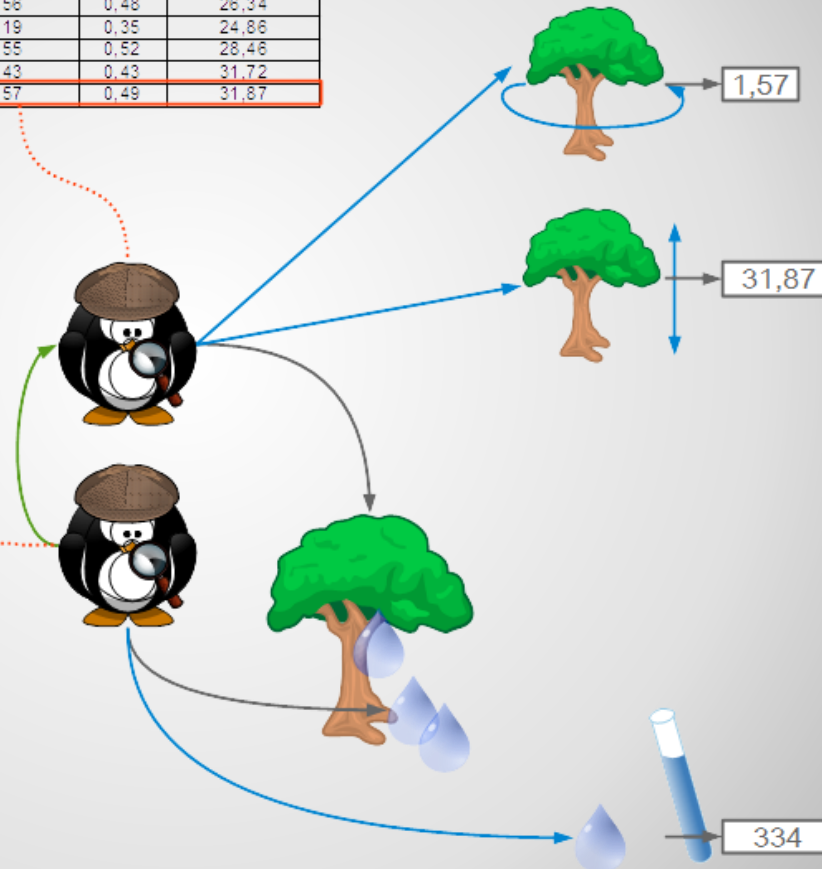
A number of different and incompatible schemas for describing research data exist because of diverse approaches in capturing data semantics. They differ in:

- Describing the complexity of parameters
- Providing data semantics for specific domains
- Indicating the admitted value of attributes
- Accounting or not for the specification of units

Most schemas capture insufficiently data semantics by conflating associated attributes and thus are not suitable to correctly describe unambiguously complex parameters

Shortcomings of schemas

SF_ml	Tree-No	Species	TreeCircum_m	BHD_m	TreeHeight_m
212	56	Birke	1,56	0,48	26,34
35	28	Eiche	1,19	0,35	24,86
62	34	Eiche	1,55	0,52	28,46
43	96	Buche	1,43	0,43	31,72
334	12	Buche	1,57	0,49	31,87



**What we
mean by
observations**

Monthly mean dissolved lead (ppb) in water taken from the river Thames by sampling

Monitored Property



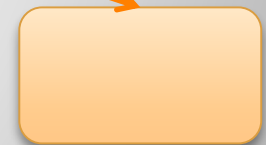
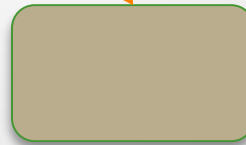
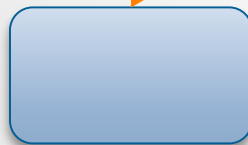
Feature



Observable Property



Process



What we mean by conceptualization

Monthly mean dissolved lead (ppb) in water taken from the river Thames by sampling

Monitored Property

River Thames

Feature

Monthly mean dissolved lead (ppb) in water

Observable Property

sampling

Process



What we mean by conceptualization

Monthly mean dissolved lead (ppb) in water taken from the river Thames by sampling

Monitored Property

River Thames

Feature

Monthly mean dissolved lead (ppb) in water

Observable Property

sampling

Process

monthly mean

?

dissolved

?

lead

?

concentration

?

ppb

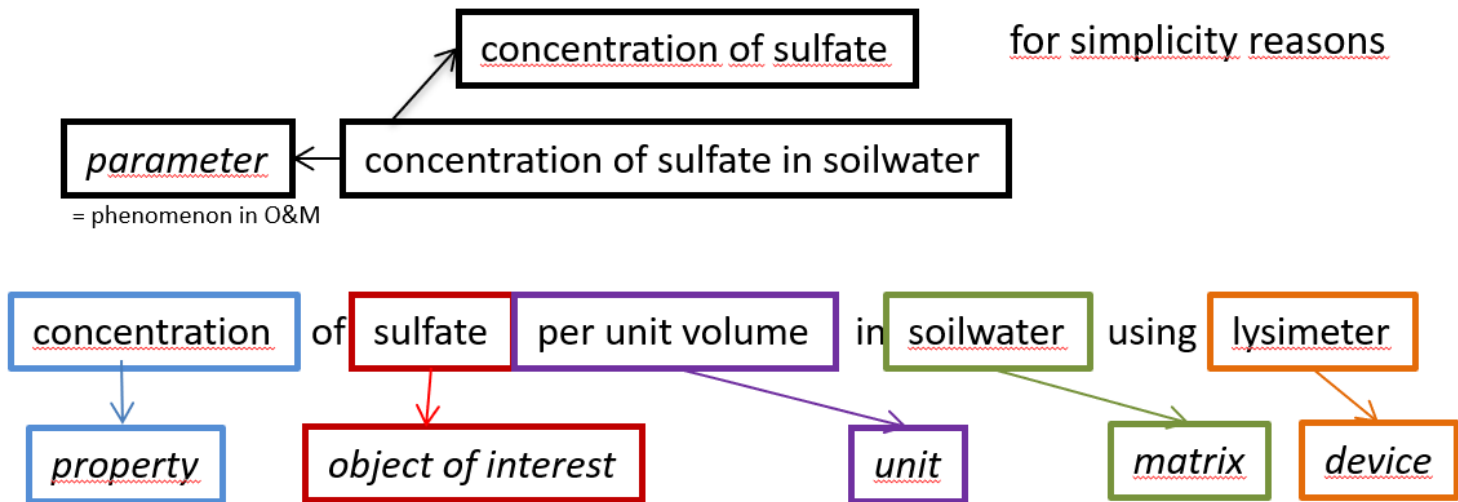
?

water

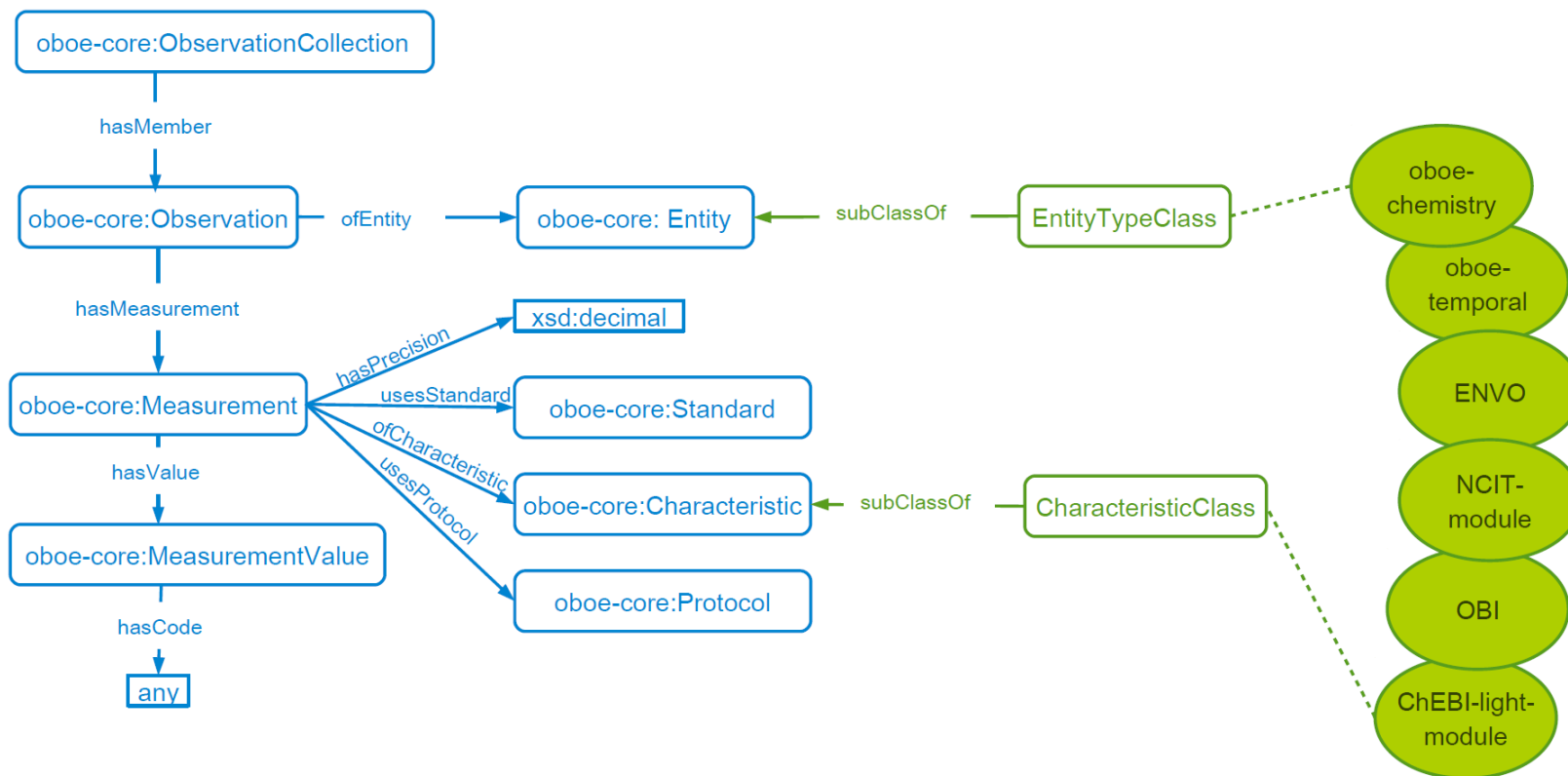
?

What we mean by conceptualization

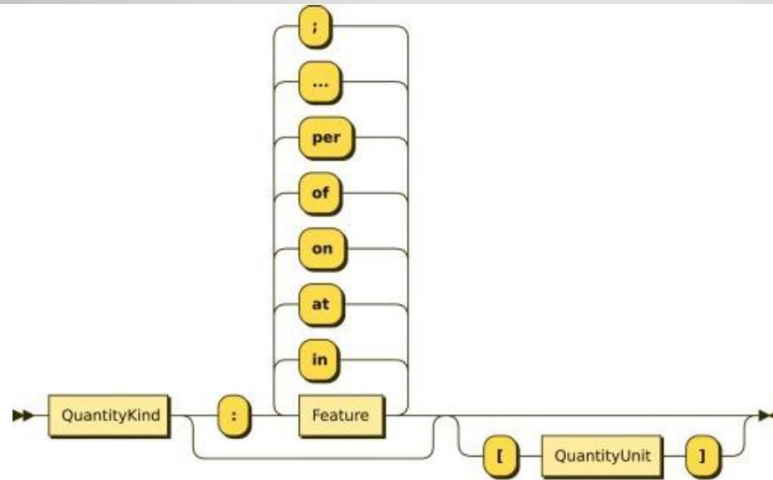
- **parameter** as a compound concept, determining an observable property for an object of interest as used by the LTER scientist.
- **atomic concepts** used for breaking down parameter in object of interest, property, matrix (media), (unit), device



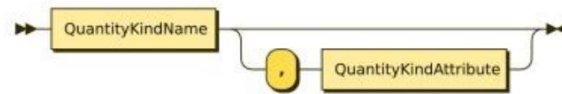
Observation parameters as defined in EnvThes/LTER



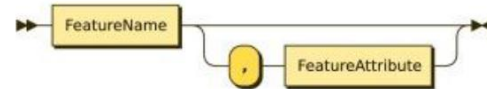
Parameter model OBOE used by AquaDiva/AnaEE



(a) ParameterInstance definition.



(b) QuantityKind definition.



(c) Feature definition.

<https://doi.org/10.1016/j.jbiotec.2017.07.016>

Parameter structure in PANGAEA

**The measurand or
'property'**

e.g. temperature,
uptake rate,
abundance,
concentration

Defined in:

[S06](#)

**An entity or object
of interest**

a biological entity,
a chemical entity
OR a physical quantity

[S25](#)

[S27](#)

[S29](#)

**An environmental
matrix or
compartment**

e.g. atmosphere and its sub-component phases,
water body or its sub-component phases,
sediments or its sub-components,
experimental environments

[S26](#)

**Measurement –
matrix relationship**

Contains logical connectors plus key information
about dimensionality when required
e.g. per unit volume of the

[S02](#)

BODC core elements

- 6 meetings so far (long summer break...)
- Presentations
 - LTER
 - AquaDiva
 - PANGAEA
 - BODC
 - ANAEE
 - ENVO
- Preparation of case statement (TG->WG)

Activities so far

- Agree on core terms
- Agree on terminologies for core elements
- Choose use cases with different complexities
- Describe the use cases by each schema used in the group individually
- Develop a common model, which must also be compatible to O&M
- Align the single approaches to the new agreed model
- Produce guidelines and write publications

Next steps

- Involve other important stakeholders like GBIF/TDWG, CF standard names, ...
- Convert to a RDA working group
- We use slack channel: vocabulary-services.slack.com, please subscribe to #tg-harmonize-measure
- All produced materials, meeting minutes and presentations:
<https://drive.google.com/open?id=1eZ2ypn2Q1SRqSZYOBiVMob3ZgkKG667e>
- Please join!

Communication channels