HAL open science

Developing semantic interoperability in ecology and ecosystem studies

Christian Pichot, Cécile Callou, Andre Chanzy, Philippe Clastre, Chloé Martin, Damien Maurice, Ghislaine Monet

To cite this version:

Christian Pichot, Cécile Callou, Andre Chanzy, Philippe Clastre, Chloé Martin, et al.. Developing semantic interoperability in ecology and ecosystem studies. RDA 17th Plenary Meeting, Apr 2021, Edinburgh, United Kingdom. . hal-03234155

HAL Id: hal-03234155

https://hal.inrae.fr/hal-03234155

Submitted on 25 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

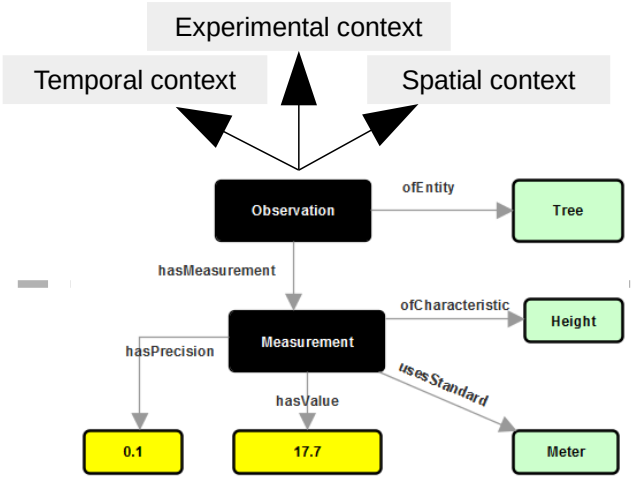
Developing semantic interoperability in ecology and ecosystem studies

RDA RESEARCH DATA ALLIANCE
17th Plenary

PICHOT Christian[1], CALLOU Cécile[2], CHANZY André[3], CLASTRE Philippe[1], MARTIN Chloé[2], MAURICE Damien[4], MONET Ghislaine[5]

1. INRAE URFM 228 route de l'Aérodrome 84914 Avignon
2. CNRS UMS BBEES 55 rue Buffon 75005 Paris
3. INRA UMR EMMAH 228 route de l'Aérodrome 84914 Avignon
4. INRAE UMR SILVA route d'Amance 54280 Champenoux
5. INRA UMR CARRTEL 75 avenue de Corzent 74200 Thonon-les-bains

Obs. & exp. on ecosystems



Graph data modeling

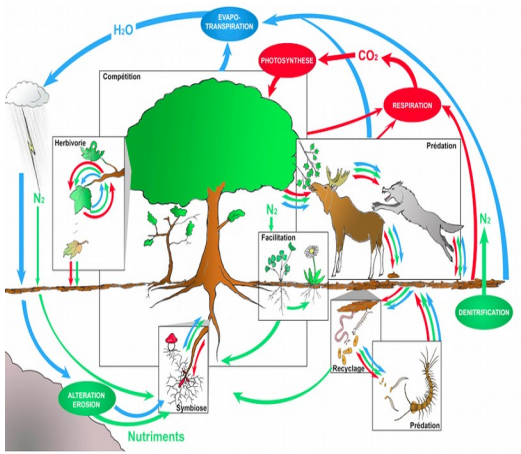
RDF

Data & metadata

Semantic interoperability

CC BY

Rationale

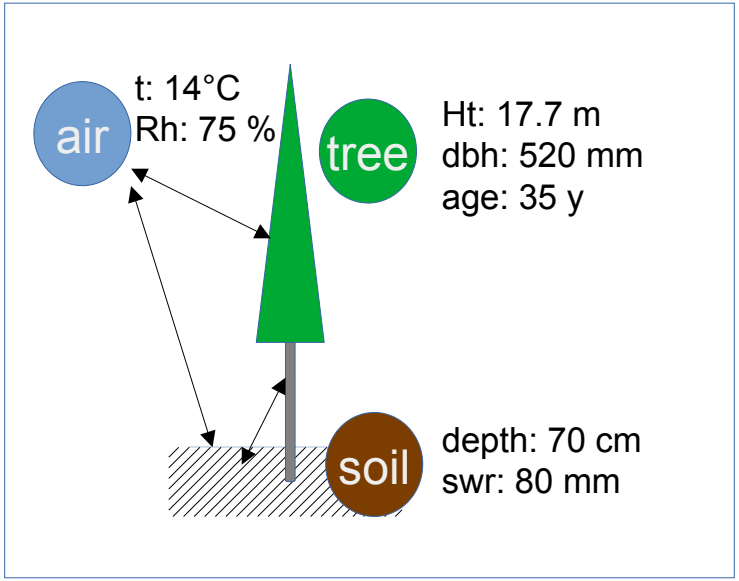


Ecosystem study requires complex research and deals with heterogeneous, varied and widespread data.

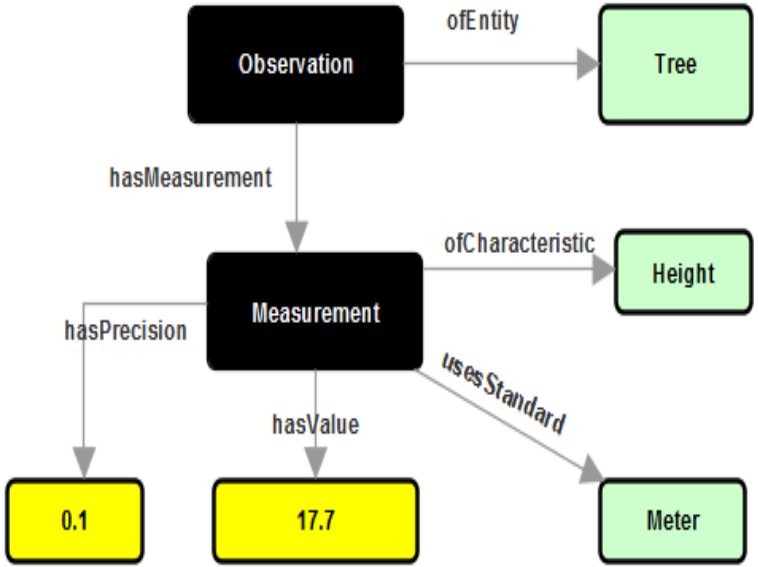
The proper understanding and interoperability of
the information sources remains one of the greatest challenges

Method

1) Identify
- the components of the system
- and their relationships



2) Model the system
using semantic vocabularies

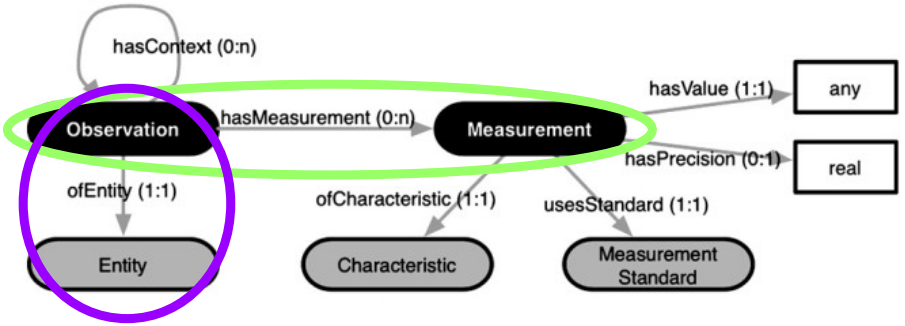


CC BY

Implementation

AnaEE* RI as scientific context:
The Research Infrastructure offers services for experimentation on continental ecosystems

OBOE* as ontological framework:
The ontology provides the atomic elements for modeling observations

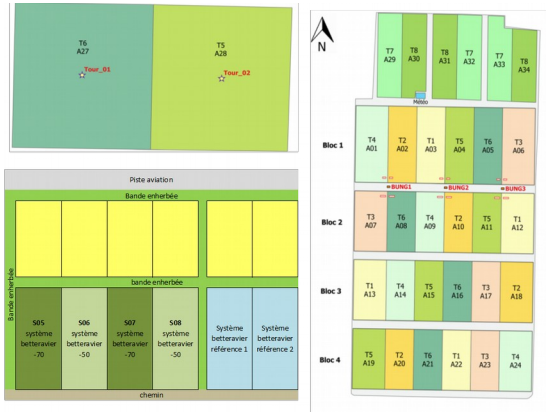


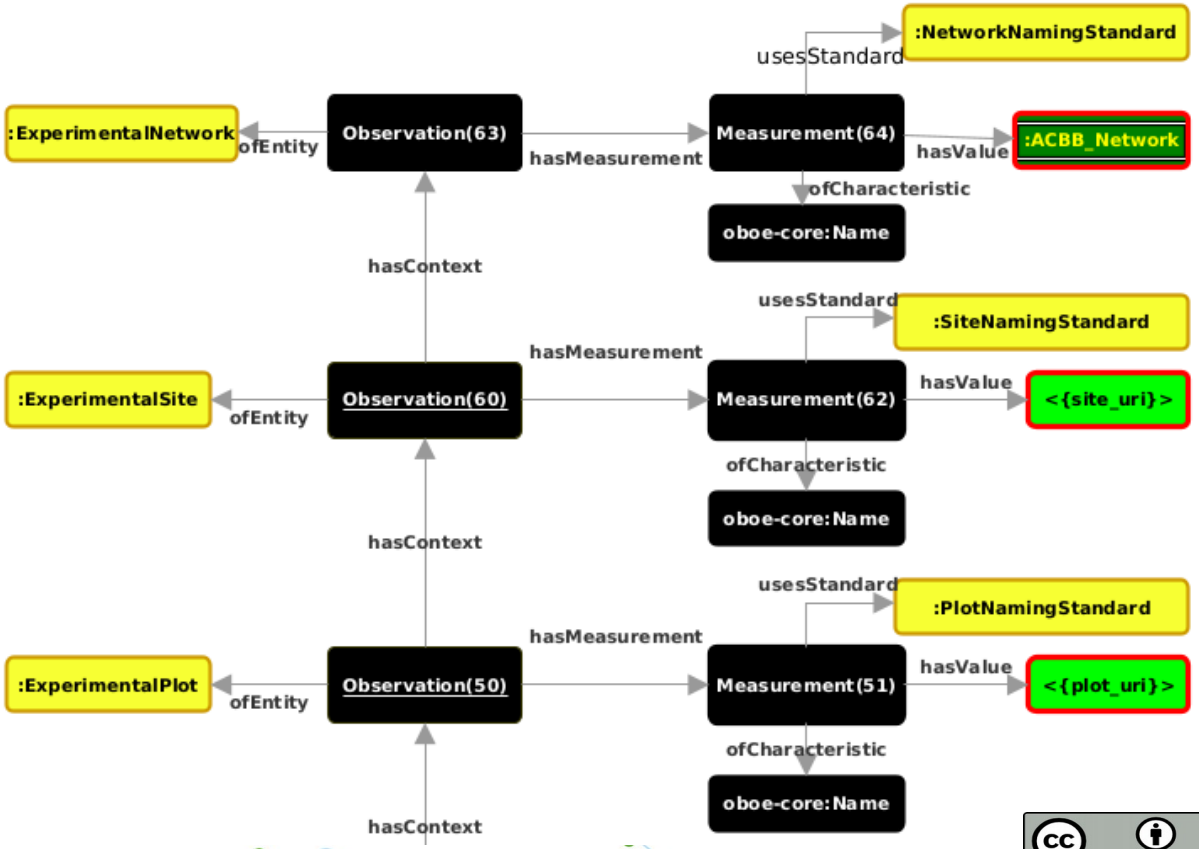
*Mark Schildhauer, Matthew B. Jones, Shawn Bowers, Joshua Madin, Sergeui Krivov, Deana Pennington, Ferdinando Villa, Benjamin Leinfelder, Christopher Jones, and Margaret O'Brien. 2016. OBOE: the Extensible Observation Ontology, version 1.2. KNB Data Repository. doi:10.5063/F1125R0F

CC BY

Implementation

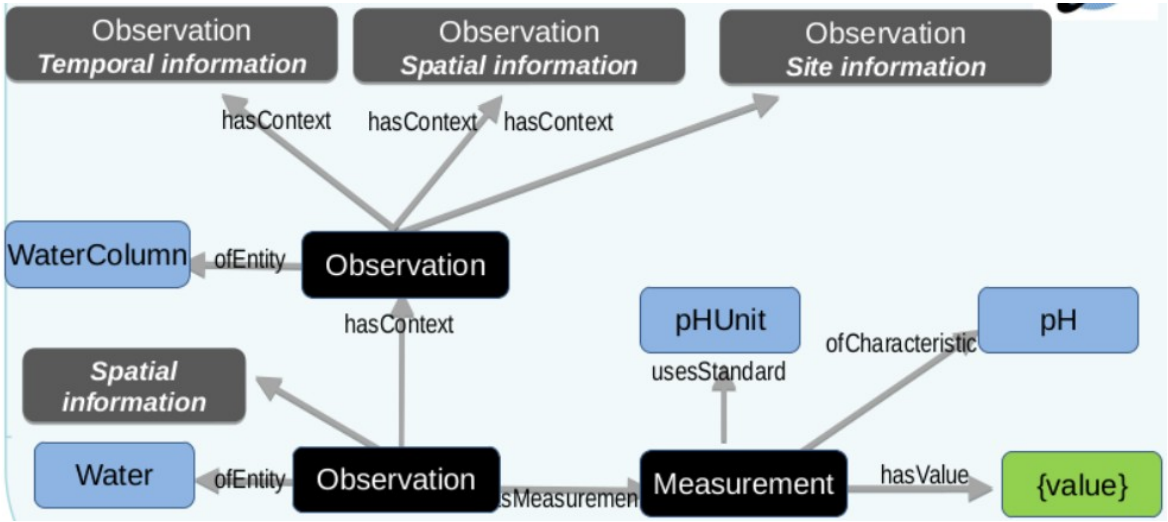
Modeling the experimental context





Implementation

Modeling the measured variables



Implementation

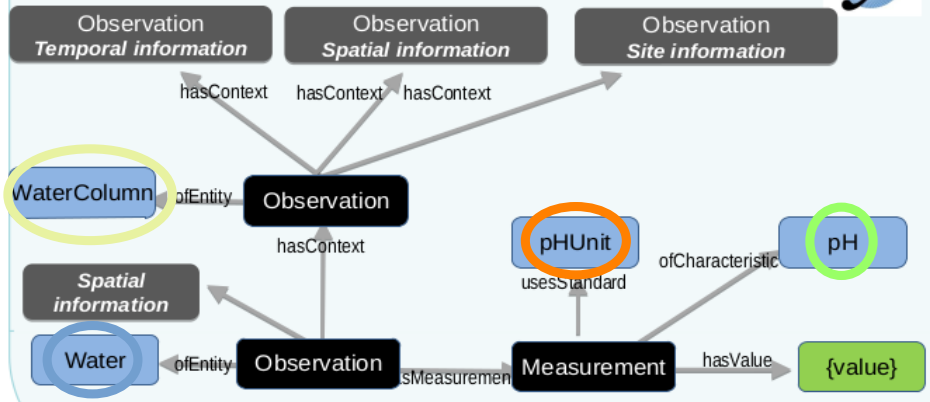
Graph pattern approach:

Design parametric graphs suitable for a set of variables

Variable semantic description .CSV

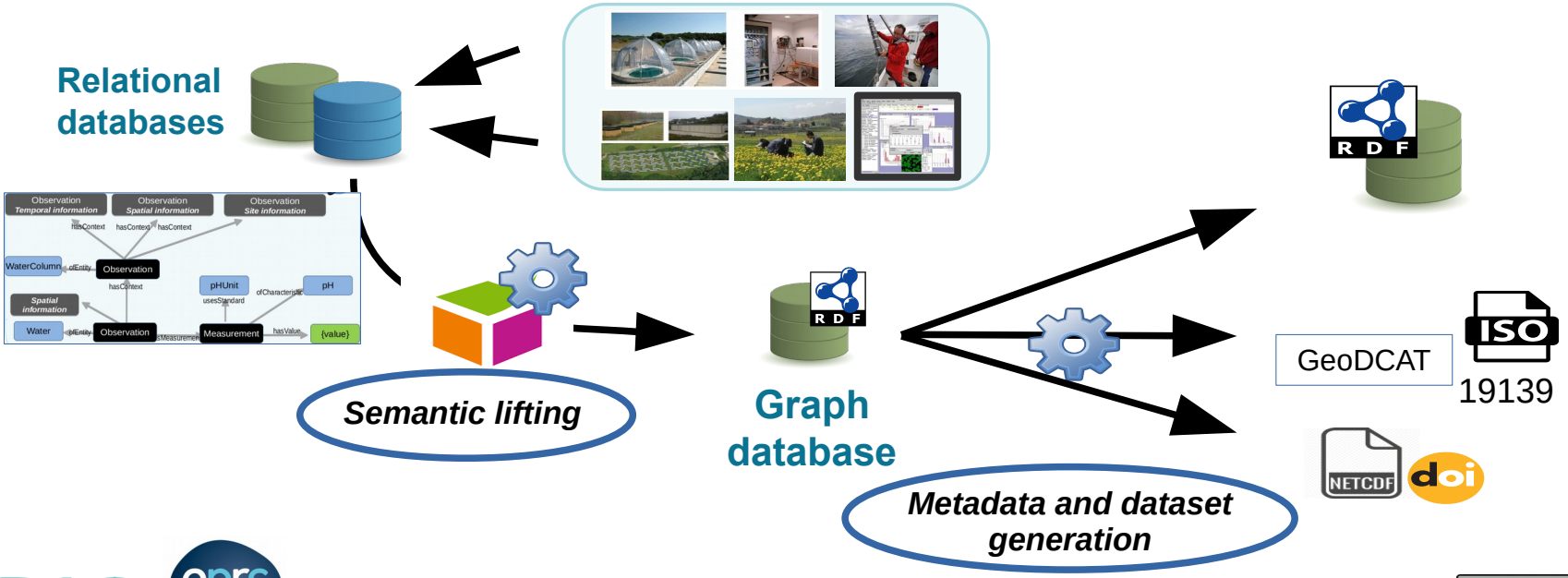
AnaEE standard	Category	Context	Entity	Characteristic	Unit
Phytoplankton	Biodiversity	Water	Phytoplankton	Volume Per Volume	MicroMeterCubed Per Millimeter
WaterPH	Physical Chemistry	Water Column	Water	pH	pHUnit
...	...	...	...	...	...

Semantic data graph (per category)



Semantic lifting and data exploitation

Graph patterns and variable semantic descriptions are processed by a pipeline for semantic lifting of the data before their exploitation



CC BY

Lessons from this work

- The OBOE generic ‘observation model’ allows for efficient atomic modeling of the components of the system and of their nested or crossed relationships.
- In addition to the provided OBOE extensions (characteristics, spatial, temporal, standards), the unique identification of the system components is ensured through new classes and individuals especially for Entity (e.g experimental entities) and Standards (e.g lists of variable names or of experimental facilities).
- A graph pattern approach for the modeling of the variables leads to a more efficient investment at greatly reduced cost
- When data are initially managed in structure databases, the data modeling is directly exploitable by pipelines for mass lifting to rdf graphs
- The whole process produces syntactically and semantically interoperable data, contributing to FAIR sharing and data reuse