The I-ADOPT RDA WG: Boosting the I in FAIR through the harmonization of observational data terminologies

Pier Luigi Buttigieg on behalf of Barbara Magagna and the i-ADOPT WG

GoFAIR Workshop: “Semantic Interoperability of Metadata for Cross-Domain Research of the Future”
Hamburg, November 11 2019
• **Core objective:** enable and perpetuate *testable*, machine-centric interoperability of existing terminologies across the semantic gradient

  ◦ Develop best practices and an interoperability framework for terminology resources pertinent to observable properties

  ◦ Test and ensure interoperation through annotation and multi-resource querying/mobilisation of research data

**Aims of the task group**
Analysing ecological phenomena across geographic, temporal, biological scales requires a variety of disparate observational data sets

Observational data are often represented in *tabular form* but differ in:
- The number of attributes
- the relationships implied or asserted between attributes
- the coding conventions used for representing information within data sets

**The challenge:**
**a case from ecology**

Semantics for biodiversity and ecosystem research, ICEI 2018, Jena
Collected observational data

<table>
<thead>
<tr>
<th>SF_ml</th>
<th>Tree-No</th>
<th>Species</th>
<th>TreeCircum_m</th>
<th>BHD_m</th>
<th>TreeHeight_m</th>
</tr>
</thead>
<tbody>
<tr>
<td>212</td>
<td>56</td>
<td>Birke</td>
<td>1.56</td>
<td>0.48</td>
<td>26.34</td>
</tr>
<tr>
<td>35</td>
<td>28</td>
<td>Eiche</td>
<td>1.19</td>
<td>0.35</td>
<td>24.86</td>
</tr>
<tr>
<td>62</td>
<td>34</td>
<td>Eiche</td>
<td>1.55</td>
<td>0.52</td>
<td>28.46</td>
</tr>
<tr>
<td>43</td>
<td>96</td>
<td>Buche</td>
<td>1.43</td>
<td>0.43</td>
<td>31.72</td>
</tr>
<tr>
<td>334</td>
<td>12</td>
<td>Buche</td>
<td>1.57</td>
<td>0.49</td>
<td>31.87</td>
</tr>
</tbody>
</table>

RDA P11 - March 2018 - VSSIG Harmonize the conceptualization of observation types
A number of different and incompatible semantic schemas for describing research data exist

**Key axes of differentiation**
- Degree of expressivity along the complexity of properties
- Domain-specificity
- Attempt to capture the value of attributes
- Specification of units

Most schemas conflate attributes along disciplinary conventions and thus are not suitable to describe complex properties to machine agents

**Shortcomings of existing schemas**

Semantics for biodiversity and ecosystem research, ICEI 2018, Jena
- Observations and Measurements
- Complex Property Model
- SOSA/SSN Ontologies
- SVO (Scientific Variable Ontology)
- CF conventions
- OBOÉ
- SERONTO
- EFO
- OBO Foundry conventions...

Some different models/schemas
Different terminologies used

- SDN-Voc
- ENVO
- BIPM
- IUPAC
- EnvThes
- CHEBI
- OM
- SWEET
- QUDT
- SDMX/DDI
- WORMS
- ITIS
- ..
Steps towards true interoperability: unity in diversity?

(PS: it’s not the user’s problem)
• Property/quality of an object of interest

• Any quantifiable or qualifiable characteristic of an object or subject of research or monitoring of a given "feature of interest"

• Biological, Chemical, Physical, Administrative

• "Observations", "Traits", "Variables", "Parameters", "Measurements", ...
• Continued....

• Observed directly or by proxy (modelling/calibration)
  ◦ Eg. chlorophyll-a fluorescence > chlorophyll-a concentration and productivity > biomass of photosynthetic material and primary production

• Field observations, Laboratory experiments, Remote sensing, Modelling

• Object of interests: Specimen, Populations, Samples, entire Environments

Define observable property...
Monthly mean dissolved lead (ppb) in water taken from the river Thames by autonomous sampling

Feature of Interest

Observable Property

Procedure

Unit

Complex observable properties

RDA P11 - March 2018 - VSSIG Harmonize the conceptualization of observation types
Monthly mean dissolved lead (ppb) in water taken from the river Thames by autonomous sampling.
Monthly mean dissolved lead (ppb) in water taken from the river Thames by autonomous sampling

Observable Property

Concentration of lead dissolved in water

Feature of Interest

River Thames

Procedure

Sampling & averaging...

Monitored Property

Atomisation of complex properties

RDA P11 - March 2018 - VSSIG Harmonize the conceptualization of observation types
Monitored Property

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

Observable Property

Concentration of lead dissolved in water

Feature of Interest

River Thames

Sampling & averaging...

ppb

Procedure

ATOMISATION

Concentration of lead dissolved in water

Feature of Interest

River Thames

Procedure

ATOMISATION

Concentration of lead dissolved in water

Feature of Interest

River Thames

RDA P11 - March 2018 - VSSIG Harmonize the conceptualization of observation types

Atomisation of complex properties
BODC PUV P01 Semantic model

Property
- concentration
- monthly mean
- lead
- water body
- dissolved
- ppb

Statistical property
- http://vocab.nerc.ac.uk/collection/S06/current/S060045/
- http://vocab.nerc.ac.uk/collection/S07/current/S070016/
- http://voca.b.nerc.ac.uk/collection/S27/current/CS002545/
- http://vocab.nerc.ac.uk/collection/S21/current/S21S027/
- http://vocab.nerc.ac.uk/collection/S06/current/UPPB/

Object of Interest
- http://purl.obolibrary.org/obo/CHEBI_27889

Matrix

Matrix phase

Units
Parameter model OBOE used by AquaDiva/AnaEE

RDA P11 - March 2018 - VSSIG Harmonize the conceptualization of observation types
VSSIG Harmonize the conceptualization of observation types

Parameter structure in PANGAEA

https://doi.org/10.1016/j.jbiotec.2017.07.016
(very rough) OBO Library patterns
A rough illustration of the semantic gradient

Ontologies

Taxonomies

Data models

Controlled vocabularies

Thesauri

Glossaries

Weaker semantics

Stronger semantics

Modified from McCreary D (2006) Patterns of Semantic Integration. CC 2.5
I-ADOPT: building and testing a common framework for persistent alignment across terminology resources
1. Collect user stories > Formalise into use cases
2. Collect observational data > annotation practices with used terminologies and representation strategies
3. Derive requirements from use cases
4. Check compliance of representation strategies with each requirement, analyse overlaps and gaps between them
5. Develop Interoperability Framework
6. Develop local Mapping Design Patterns

**Tasks of the I-ADOPT WG**
D1 Catalogue of domain-specific terminologies (2020-01-31)

D2 Synthesis report on expressiveness of representation strategies (2020-12-31)

D3 Interoperability Framework for observable properties in environmental research (2021-01-31)

D4 Guidelines on best practices on the implementation and use of the framework (2021-04-30)
• Officially endorsed RDA Working Group
• Kick-Off Oct 25 2019 in Helsinki
  ◦ November 2019 – April 2021 (18 months)
• Around 60 members: 20 RIs/Initiatives
  (eLTER, NERC, LifeWatch, PANGAEA, OBO
  Foundry, GoFAIR,..)
• Official I-ADOPT WG Site
• Github repo for I-ADOPT activities
The I-ADOPT RDA WG: Boosting the I in FAIR through the harmonization of observational data terminologies

Pier Luigi Buttigieg on behalf of Barbara Magagna and the i-ADOPT WG

GoFAIR Workshop: “Semantic Interoperability of Metadata for Cross-Domain Research of the Future”
Hamburg, November 11 2019