Semantic interoperability of operational parameter terminologies in marine sciences

Gwenaëlle Moncoiffé, BODC (UK), gmon@bodc.ac.uk
Alexandra Kokkinaki, BODC (UK), alexk@bodc.ac.uk
Alison Pamment, UKRI STFC (UK), alison.pamment@stfc.ac.uk
Neil Holdsworth, ICES (Denmark), NeilH@ices.dk
Hans Mose Jensen, ICES (Denmark), hans.jensen@ices.dk
Adam Shepherd, BCO-DMO (US), ashepherd@whoi.edu
Adam Leadbetter, Marine Institute (Ireland), adam.leadbetter@marine.ie
Rob Thomas, Marine Institute (Ireland), rob.thomas@marine.ie

In October 2019 a new Research Data Alliance's working group, InteroperAble Descriptions of Observable Property Terminology (I-ADOPT WG)[1], launched its 18-month work plan with the ultimate goal to deliver a semantic interoperability framework for observable property terminologies. The goal is ambitious but the group contains a broad range of scientific, data and semantic experts, a very committed core group, and a growing membership of users and developers of terminologies within a broad spectrum of scientific domains centered mainly but not exclusively on environmental sciences including terrestrial ecology, marine, atmospheric and earth sciences, as well as related human activity disciplines like agriculture, forestry. While the ultimate goal is semantic interoperability across domains, it is clear that this work will also benefit semantic interoperability within domains wherever there are overlaps between established resources.

In the case of marine sciences, interoperability between data sources has inadvertently been hindered by difficulties in aligning long established parameter naming schemes such as the Climate and Forecast (CF) Standard Names used by the Climate and Forecast community and users of the netCDF format, the BODC P01 Parameter Usage Vocabulary (PUV) adopted by the SeaDataNet (SDN) community and close partners, and the ICES parameter dictionary used by many monitoring government agencies and laboratories. Each of these terminology resources grows in response to requests from their user communities based on their own local semantic rules and constraints.

Table 1 shows examples of how the labels used in CF, PUV and ICES parameter vocabularies can be decomposed into useful atomic components if one knows their structure or "grammar". However these components are not always in perfect alignment and human interpretation is necessary in order to achieve an accurate mapping between the terminologies. For example, the quantity term "concentration" has to be inferred from the units in the ICES vocabulary; then "mole concentration" used in CF is broader than the quantity "concentration" used in PUV and ICES and an equivalence between the two would require access to the units; we also have important qualifier such as "dissolved" which is an attribute of the analyte in CF and ICES but a phase of the matrix in the PUV.

[1] www.rd-alliance.org/groups/interoperable-descriptions-observable-property-terminology-wg-i-adopt-wg

Community	Parameter vocabulary	Parameter label	Quantity term	Chemical substance term	Matrix/medium term
CF	Standard names	mole_concentration_of_dissolved_molecular_oxygen_in_sea_water[2]	mole concentration	dissolved molecular oxygen	sea water
SDN	PUV	Concentration of oxygen {O2 CAS 7782-44-7} per unit volume of the water body [dissolved plus reactive particulate phase][3]	concentration	oxygen	water body [dissolved plus reactive particulate phase]
ICES	<multiple>	Dissolved oxygen[4]	<inferred from units>	dissolved oxygen	water

Table 1: Examples of terminology used by the CF, PUV, and ICES vocabularies for dissolved oxygen concentrations and comparison of selected atomic component terms used for the quantity measured, for the chemical substance and for the matrix or medium

While manual matching of terms across vocabularies has been achieved for small vocabulary subsets, mainly driven by relevance to particular projects (such as the mapping of ICES and SDN vocabularies for contaminants for EMODnet Chemistry), this is not a scalable solution for complete vocabularies covering many types of observations. The increasing pressure to provide fast, Findable, Accessible, Interoperable, Reusable, machine readable and user friendly access to data from a growing number of sources and for a growing number of applications has made the need to improve semantic interoperability of scientific terminologies more acute.

The semantic interoperability framework that will be delivered as part of the I-ADOPT collaborative work will provide a common method to systematically express or represent observable properties. In the marine domain, example mappings from the BODC P01 PUV to the Complex Properties Model were prepared originally by Leadbetter and Vodden (2016). These alignments will need to be validated against the I-ADOPT agreed framework. They will then be operationalised and implemented using Linked Data principles to provide a trusted, reproducible, and harmonised framework to build semantic translation and brokering services. Such services will be needed in projects such as EnvriFAIR in order to access data across multiple marine observation networks and research infrastructures, and relate them to broader concepts like e.g. Essential Ocean and Climate Variables or environmental assessments indicators or Sustainable Development Goals criteria. Such a semantic framework will also help standardise the associations between fine granularity terminologies and higher level ontologies and semantic models such as Semantic Sensor Network Ontology, SensorML, schema.org.

This presentation aims to promote the work of the RDA I-ADOPT working group, highlight the benefits to the marine community, and stimulate collaboration and contributions.

References

Leadbetter, A. & Vodden, P. (2016) Semantic linking of complex properties, monitoring processes and facilities in web-based representations of the environment, International Journal of Digital Earth, 9:3, 300-324, DOI: 10.1080/17538947.2015.1033483

[2] vocab.nerc.ac.uk/collection/P07/current/CF14N29/

[3] vocab.nerc.ac.uk/collection/P01/current/DOXYZZXX/

[4] https://vocab.ices.dk/?CodeID=33506