The BODC Parameter Usage Vocabulary (PUV) semantic model exposed

Gwenaëlle Moncoiffé, *British Oceanographic Data Centre* (UK), gmon@bodc.ac.uk
Alexandra Kokkinaki, British Oceanographic Data Centre (UK), alexk@bodc.ac.uk

The BODC Parameter Usage Vocabulary started its existence in the 1980s as a simple list of terms. It has since evolved to constitute a fully structured controlled vocabulary schema, a meta-vocabulary whose building blocks are themselves controlled vocabularies. The logical semantic model that governs the construction of its text label enables us to express a complex array of scientific information related to the observation, in 8-byte codes. The BODB PUV has been published as a linked data SKOS vocabulary that makes it understood by both humans and machines but the underlying semantic model was not exposed, limiting its advantages to a closed community . In this presentation, we are attempting to introduce the different elements of the semantic model and publish it as linked data by describing it with the relevant ontology elements (properties and classes).

The three elements at the core of the semantic model are: the parameter entity, the object of interest and the matrix. The parameter entity can be modulated to be the property of interest itself or any statistical parameters derived from the statistical analysis of the property. The object of interest can be either a biological entity, a chemical entity or a physical object or phenomenon. Both these fields (the parameter entity and the object of interest) are mandatory for the semantic model The third element, which is the matrix, is the environment in which the object of interest is embedded or to which the property relates (e.g. a phase). This can be switched on and off depending on the type of parameter. It will be set to “not applicable” for example for many engineering parameters that do not relate to a measurement matrix. The matrix itself consists of four semantic elements, enabling us to define whether the measurement is applied on the whole matrix or on any of its sub-components or phases for example, analysis on a filtered sample or analysis on samples filtered through different filter sizes (size-fractionation). Additionally the model provides 3 optional fields to encompass key elements of the methodology at the sampling, analytical and data processing stage, as necessary.

Expressing the semantic model as linked data enables data aggregation but also integration of data across diverse disciplines and domains, where parts of these compound terminologies can be mapped to each other, boosting the semantic discovery.