

Moving controlled vocabularies into the semantic web

Rob Thomas, BODC (UK), room@bodc.ac.uk
Justin Buck, BODC (UK), juck@bodc.ac.uk
Alexandra Kokkinaki, BODC (UK), alexk@bodc.ac.uk
Roy Lowry, BODC (UK), rkl@bodc.ac.uk

One of the issues with legacy oceanographic data formats is that the only tool available for describing what a measurement is and how it was made is a single metadata tag known as the parameter code. The British Oceanographic Data Centre (BODC) has been supporting the international oceanographic community gain maximum benefit from this through a controlled vocabulary known as the BODC Parameter Usage Vocabulary (PUV). Over time this has grown to over 34,000 entries some of which have preferred labels with over 400 bytes of descriptive information detailing what was measured and how.

A decade ago the BODC pioneered making this information available in a more useful form with the implementation of a prototype vocabulary server (NERC Vocabulary Server - NVS) that referenced each 'parameter code' as a URL. This developed into the current server (NVS V2) in which the parameter URL resolves into a RDF document based on the SKOS data model which includes a list of resource URLs mapped to the 'parameter'.

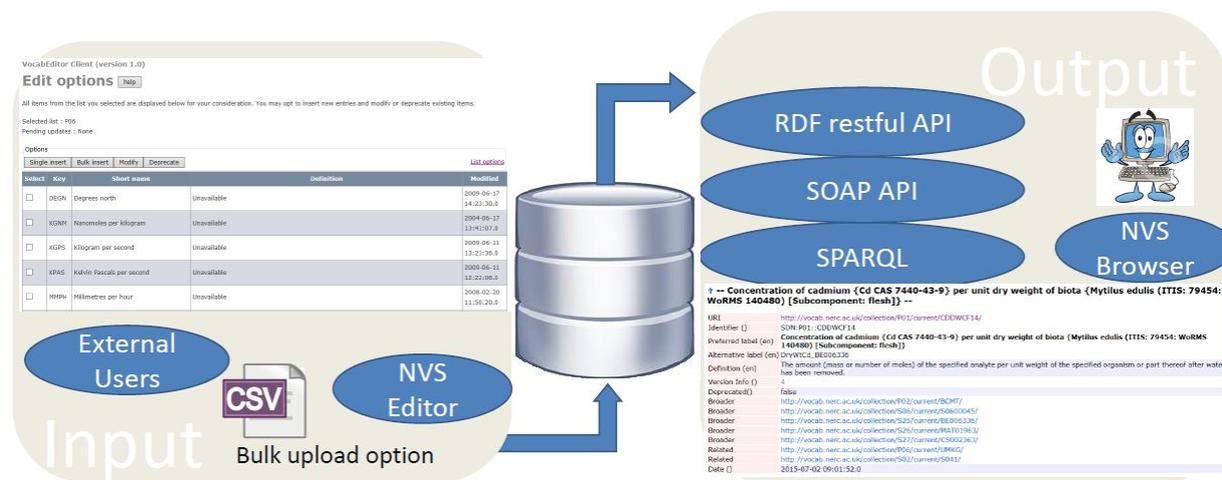


Figure 1: Overview of NVS user accessibility for inputs (via the Editor) and outputs as machine-to-machine or web browser content (NVS Browser)

The NVS contents are held within an Oracle RDBMS. A web based editor allows vocabulary content to be manipulated by authorised users outside BODC.

The NVS contents are then available via web services. Whilst this is ideal for technical users and those fluent in SPARQL, most users require something a little more user-friendly and so the NVS browser² was developed over the end point to allow more users to query the vocabularies and navigate the NVS ontology.

Exposing the semantic model that has underpinned the construction of PUV terms (green bubbles in Figure 2) using the Linked Data paradigm gives increased searchability by a method analogous to a fruit machine. The new Vocabulary Builder, released by BODC in summer 2016, allows the end user to build faceted style queries using subcomponent vocabularies through a user friendly web interface. Conceptually each wheel on the fruit machine representing a component of the semantic

model enhances the Vocab Builder's search functionality. Further this paradigm allows users to construct and submit new combinations for addition to a vocabulary by using the underlying semantic model that combines terms from other vocabularies.



Figure 2: Illustration of the key vocabulary linkage from the PUV, the underlying semantic model components (in green) and the links to external authoritative vocabularies (in red).

In addition by providing links via SKOS mappings into external ontologies (red bubbles in Figure 2) the information captured in a 1980s parameter code now conforms to the Linked Data paradigm of the Semantic Web, vastly increasing the descriptive information accessible to a user. For example the parameter code for a contaminant in biota, such as 'cadmium in *Mytilus edulis*', carries RDF triples leading to the entry for *Mytilus edulis* in the WoRMS and for cadmium in the ChEBI ontologies.

This presentation will demonstrate the new user functionality provided by tools created in the last 12 months:

- NVS Editor
- NVS Browser search tool
- External mappings
- NVS Vocabulary Builder