IMDIS-58: Moving Controlled Vocabularies into the Semantic Web

Robert Thomas (room@bodc.ac.uk), Alexandra Kokkinaki, Roy Lowry

British Oceanographic Data Centre, National Oceanography Centre, UK

Introduction

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 (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'.

NVS2 structure

The NVS contents are held within an Oracle RDBMS. A web based editor allows vocabulary content to be manipulated by authorised users outside BODC.
https://www.bodc.ac.uk/data/codes_and_formats/vocabulary_editor/
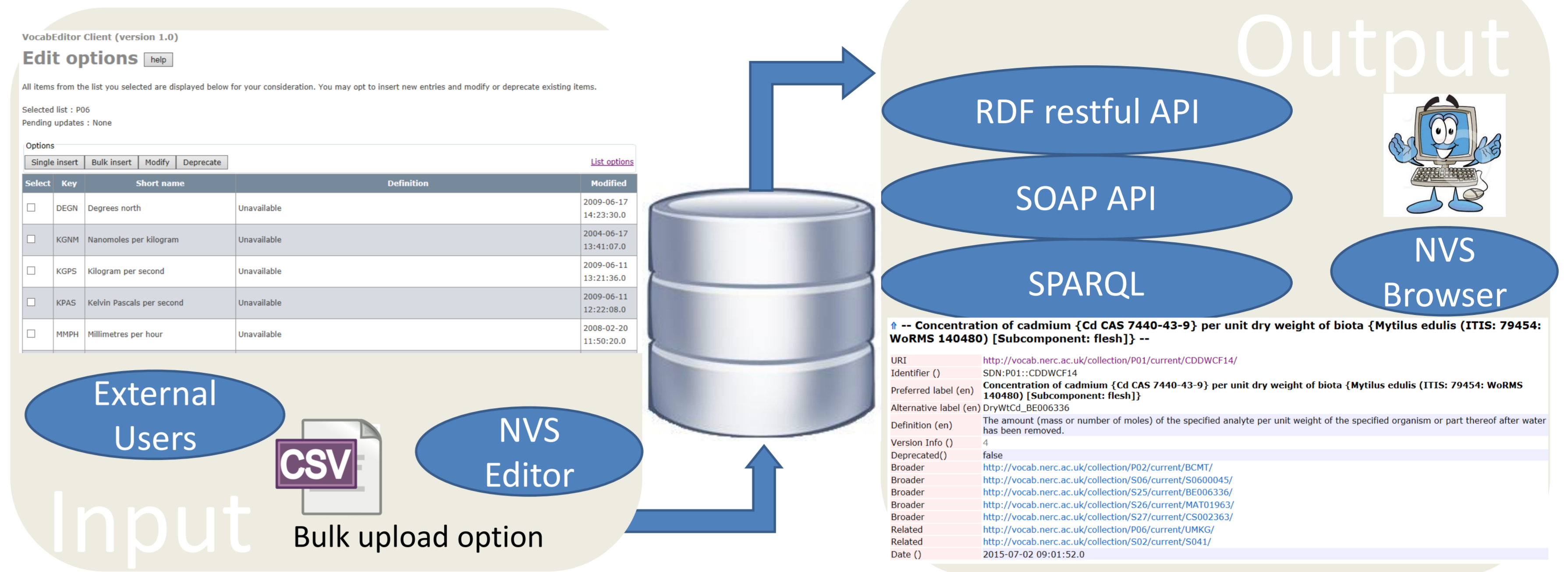
The NVS contents are then available via web services. Whilst this is ideal for technical users and those fluent in SPARQL, most users require something a little more user-friendly and so the NVS browser[2] was developed over the end point to allow more users to query the vocabularies and navigate the NVS ontology.

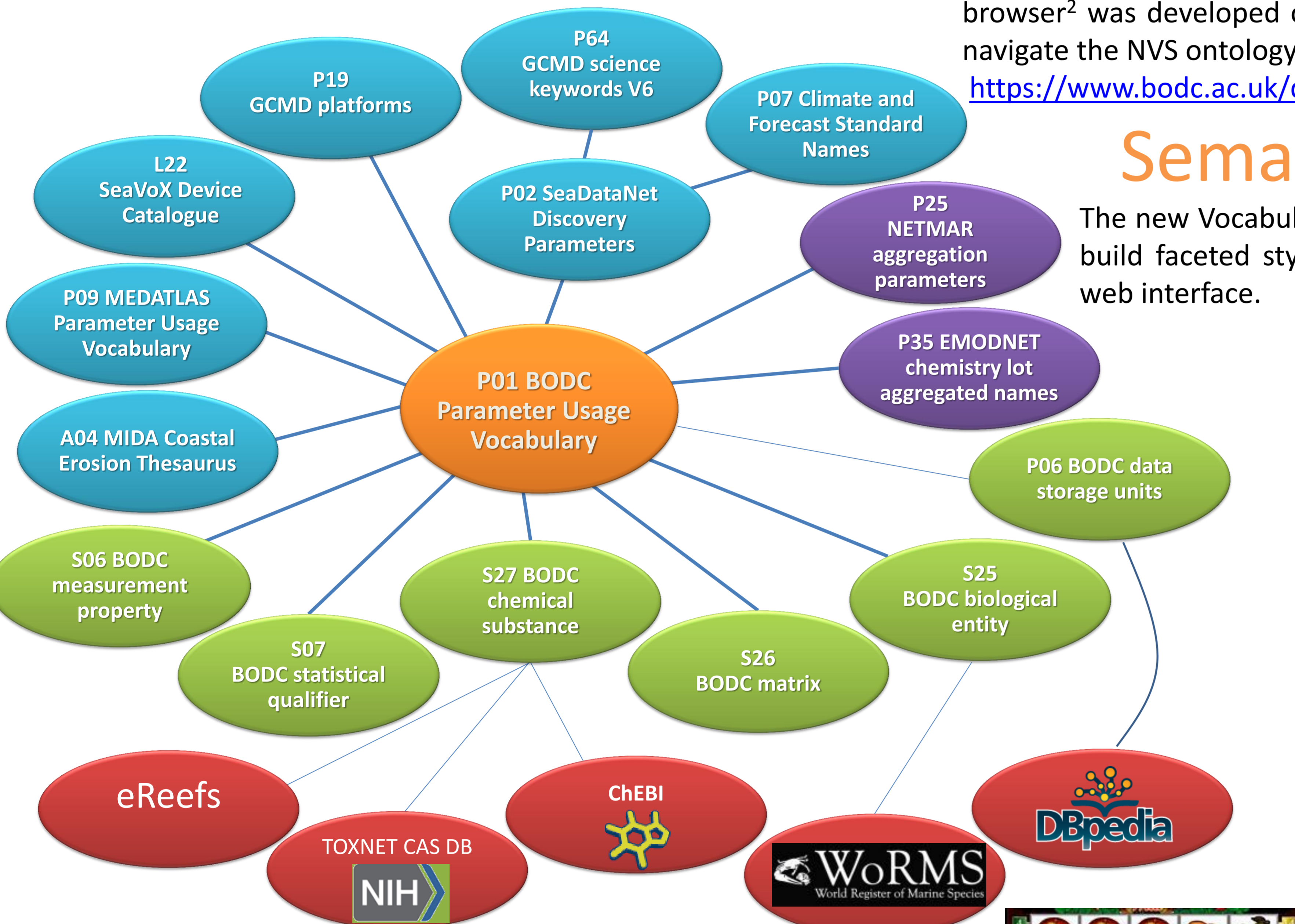https://www.bodc.ac.uk/data/codes_and_formats/vocabulary_search/

Semantic fruit machines

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.

https://www.bodc.ac.uk/data/codes_and_formats/vocabulary_builder/

Exposing the semantic model that has underpinned the construction of PUV terms (green bubbles) using the Linked Data paradigm gives increased searchability by a method analogous to a fruit machine. Each wheel on the machine being a component of the semantic model.

CHEBI:37249 - elemental cadmium

External linkages

By providing links via SKOS mappings into external ontologies (red bubbles above) 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.

NVS2 usage

Images for usage 30 days up to 4 Oct 2016. QR code links to most recent statistics.