

Managing Australian Ocean Data Network (AODN) Vocabularies

Kim Finney, Integrated Marine Observing System (Australia), kimfinney@gmail.com

Sebastien Mancini, Integrated Marine Observing System (Australia),

Sebastien.Mancini@utas.edu.au

Roger Proctor, Integrated Marine Observing System (Australia), Roger.Proctor@utas.edu.au

Natalia Atkins, Integrated Marine Observing System (Australia), Natalia.Atkins@utas.edu.au

Using Controlled Vocabularies In The AODN Data Infrastructure.

The Australian Ocean Data Network (AODN) comprises of data captured by Facilities of Australia's Integrated Marine Observing System (IMOS) and other organisations active in the coastal and marine realms. IMOS itself is a multi-institutional collaboration led by the University of Tasmania (UTAS). It deploys a wide range of observational equipment and has been capturing and publishing marine data since 2008.

These data are documented using an agreed community (ISO 19115 compliant) metadata schema and are served through a common infrastructure – the AODN Portal, established and hosted by IMOS. To enable the reliable discovery and consumption of AODN data via this infrastructure, dataset descriptions encompass controlled vocabularies. Portal data indexing and facet-based searching harness vocabulary terms to improve the precision and recall of search results. These same terms should also be used inline (e.g. within self describing netcdf files) and work is already underway within IMOS to establish automated procedures for checking Facility submitted NetCDF files for syntactic and semantic compliance.

Research Vocabularies Australia (RVA) - Semantic Tooling For Australia's Research Groups/Institutions.

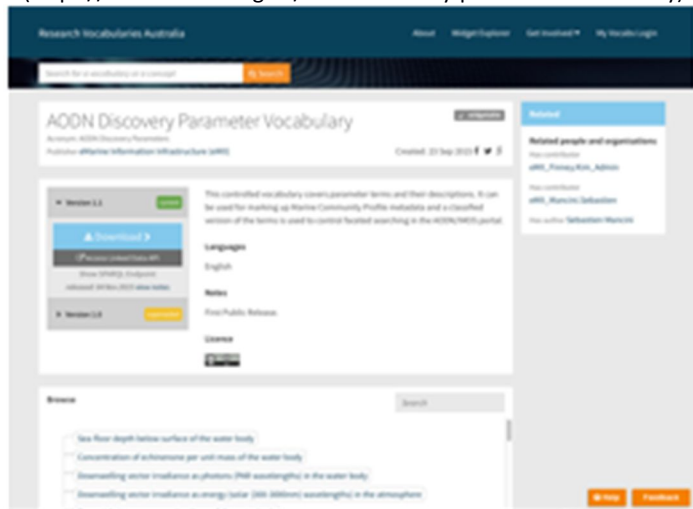
Until late 2015 the creation and management of AODN vocabularies was a labour intensive activity, involving relational databases, spreadsheets and custom scripting. Vocabulary publication was restricted to periodic posting of static ISO 19135:2005 registers and community participation in vocabulary development was restricted. Since then, a partnership between IMOS and the Australian National Data Service (ANDS), has delivered a national set of services that not only help the AODN with its vocabulary management, but any Australian agency wishing to create and publicly deploy vocabularies of any flavour.

These freely available ANDS hosted vocabulary services are bound together as a suite of interoperating tools, comprising of a skos editor (academic license of Pool Party Semantic Suite); a SPARQL-enabled semantic data store (Sesame); a REST-based API (SiSSVoc) and a Catalogue Portal (based on ANDS custom software). These tools have been deliberately loosely coupled so that inevitable advances in semantic technologies can be capitalised upon easily without breaking offered functionality. Vocabularies created using the Pool Party editor can be accessed directly from the ANDS Catalogue Portal, versioned, uploaded to the sesame store and then published. For human-centric use-cases published vocabularies can be searched and browsed in the Catalogue and downloaded via a user interface (see Fig. 1). For use-cases requiring machine-to-machine transactions there is a SPARQL end-point and an API for filtering a vocabulary to access individual terms. This API is also augmented by a user interface (UI).

During testing of the tool suite a number of limitations were encountered with the Pool Party technology, despite it being one of the best commercial editors available. To address some of these short-comings ANDS, with input from IMOS, has developed some stand-alone utilities that execute against the Pool Party data store, by-passing the Pool Party UI. For more intrinsic deficiencies ANDS and IMOS have been influencing the Semantic Web Company (Pool Party vendors) to improve Pool

Party functionality for scientific and educational users. To date the company has been positive and responsive to suggestions.

Fig. 1. A vocabulary entry in the ANDS RVA Catalogue Portal (<https://vocabs.ands.org.au/aodn-discovery-parameter-vocabulary>)



Managing published vocabularies is performed through a 'MyVocabs' administration console which is part of the ANDS Catalogue. User accounts are provisioned using Australian Access Federation (AAF) credentials. AAF is a combination of technology and policy that provides a trust framework. Subscribers, which are largely the university and education sector, get a national single sign-on that allows individuals across many different organisations to collaborate and access online resources, within a trusted environment.

AODN Vocabularies In Action.

Now that basic tooling is in place to create, manage and publish controlled vocabularies experimenting has begun to develop a shared governance model for vocabulary management amongst AODN partners. Greater participation by partners in curating and moderating community vocabularies should engender stronger ownership, expand the range and number of vocabularies and their terms and make it easier for data providers to deliver conformant metadata mark-up. How to best implement updates to metadata records submitted to the AODN central GeoNetwork-based metadata catalogue, when vocabularies are re-versioned is also under trial. The over-riding aim in these endeavours is to ensure that controlled vocabulary usage does not become a burden, or barrier to participation, for AODN data providers.

Importantly, the basic tooling required to create and use vocabularies, particularly in Linked Data scenarios is now in place. This means more time can be spent on developing innovative ways in which to use the vocabularies, rather than an excessive focus on their curation. To this end IMOS is currently working with a UTAS research team who are developing a national shelf rocky reef habitat classification scheme and associated spatial products. The intent is to develop the classification scheme using RDF and the ANDS tooling in a manner that adds value to the research and makes it easier for future reef data providers to embellish the national products. By using an encoding such as RDF it will be possible to align this classification scheme with past schemes no longer in vogue opening the path for an intelligent re-use of existing data.