**Potential Use Cases For A Vocabulary Management Tool**

**(Basic Functional Requirements)**

The overall infrastructure context in which a vocabulary management tool would operate is provided generally in the attached diagram (Figure 1). This tool is considered an intrinsic component in the IMOS/AODN infrastructure related specifically to vocabulary management.

The following categories of high-level use-cases apply directly to this component:

**1.** **Community Metadata/Dataset Creator\_Submitter**

I am an IMOS/AODN facility manager and want to contribute metadata and data to the IMOS/AODN data discovery network. I understand that my submissions should conform to community agreed schema which themselves include controlled vocabularies of specific terms. To ensure my controlled vocabulary needs are met I wish to be able to do the following:

1. Discover which vocabularies are required within schema that have been deemed relevant (mandatory ?) for publishing my specific data.
2. Peruse a list of terms within a required vocabulary to see if the terms available are relevant for my needs.
3. If there are missing terms within a vocabulary I would like to be able to suggest additional terms that meet my needs.
4. Suggest new terms that can be sourced from existing published vocabularies.
5. Register as a vocabulary contributor and update my details when my details change.
6. Be presented with the naming rules for the term I am about to suggest.
7. Be given the option of uploading more than one term at a time provided my upload information conforms to a given format and file-type.
8. Clone an existing term and any of its necessary ancillary information to create a slightly modified new term.
9. If I suggest a term I would like to be notified when my additional term has been accepted so that I can use it in marking up my data/metadata.
10. If I suggest a term I would like to be notified about who will be moderating my term and their contact details.
11. If I choose to use my own ‘tools’ for marking up and formatting my metadata/data I would like to download versioned copies of specific existing vocabularies in a number of formats (e.g. Skos, rdfs, XML).
12. View all terms submitted by me, organized by vocabulary.
13. **Vocabulary Management Tool Term Moderator**

I am an IMOS/AODN Term Moderator and want to moderate terms in a vocabulary, which are used in the IMOS/AODN data discovery network. I wish to be able to do the following:

1. Receive notification when a community member submits a new term for moderation for a vocabulary where I am registered as the ‘moderator’.
2. Be able to see all ‘pending’, ‘accepted’ and ‘rejected’ term moderations relating to the vocabulary I am managing.
3. Be able to see all existing terms already registered in the vocabulary.
4. Be able to see ‘similar’ terms already registered in the vocabulary.
5. Be reminded of the business naming rules for creating terms in this vocabulary.
6. View the registered details of the term submitter including email and phone contact details.
7. If the term is well formed, does not duplicate an existing term and passes moderation I would like to ‘accept’ the term without further moderation and in parallel notify the term submitter by email that the term they have submitted has been added to the vocabulary and when it will appear in related online servers.
8. If the term requires minor modification to suit the term naming rules of the vocabulary I would like to ‘accept with minor edits’ the term in question and in parallel notify the term submitter of my actions, that the term has been added to the vocabulary and when it will appear in related online servers.
9. Modify the term name.
10. Add any relevant associations to the term definition (e.g. links to synonyms or close matches to other existing terms).
11. If the term is a duplicate or is unacceptable for any other reason I want to be able to contact the term submitter to discuss term moderation.
12. Be able to ‘reject’ submitted terms and log the reason for rejection (after discussing the issue with a term submitter).
13. Confirm an auto-generated permanent address for the ‘accepted’ term.

**3. Vocabulary Management Tool System Administrator**

I am an IMOS/AODN System Administrator and want to administer aspects of the Vocabulary Management Tool, which is used in the IMOS/AODN data discovery network. I wish to be able to do the following:

1. View and edit the status (e.g. deprecated, current, pending moderation) of any (existing or submitted) terms across all vocabularies managed by the tool.
2. Search for specific terms in specific vocabularies.
3. View all stored details pertaining to specific terms.
4. Create new vocabularies with which terms will be associated.
5. Enter new terms for moderation.
6. Edit any register information associated with existing terms and log any changes made (schema should ideally be based on 19135 Register model).
7. View and edit the details of all people who are registered to submit terms.
8. View and edit the details of all people who are registered as moderators.
9. Manage and control the versioning of vocabularies and their included terms (e.g. be able to deprecate terms, change version details of registered vocabularies and update included/deprecated terms).
10. Control the generation of export files, or export links that will be used by associated vocabulary servers (e.g. GeoNetwork MEST Thesaurus, GeoNetwork MEST Registers, IMOS Sissvoc instance, eREEFs Sissvoc instance. NERC Vocab Server).
11. Manually create links from a term in the ‘Parameter Name’ vocabulary to relevant terms in the eREEFs Observable Properties vocabulary and/or suggest additions to the terms in the Observable Properties vocabulary so that appropriate links can be made.
12. Add synonym or close match links (or other legitimate associations) to existing terms in the repository.
13. Access a bulk upload facility to import terms and their metadata (e.g. having identified schema for the import files).

**Assumptions Made Relating To IMOS/AODN Vocabulary Use**

**(Further explanation of Figure 1)**

1. It is assumed that for any given dataset metadata, or dataset schema (that will be published via the 123 Portal infrastructure) all included vocabularies will be:
2. considered a common resource, governed, updated, managed or alternatively adopted from elsewhere by the AODN community (all coordinated by a community-sanctioned entity e.g. IMOS),
3. available for download in specified formats so that any potential metadata/data submitter can incorporate required vocabulary terms in the metadata, or the data they submit to IMOS/AODN.
4. moderated in a distributed fashion by designated domain experts,
5. extended with new terms through the use of a public web client that can be used by any registered term submitter (or moderator),
6. given permanent http URIs,
7. defined and able to be queried via one or more services using a URI where the definition may also indicate term associations,
8. Metadata contributors using the GeoNetwork MEST should be able to use the GeoNetwork Thesaurus widget to load and select any controlled vocabularies mandated in any given metadata schema. The terms selected will be identified by a URI (as a minimum) but may include a label and text description. MCP 2.0 is the metadata schema that contributors must conform (or map) to in order to publish data through the IMOS/AODN data network.
9. Metadata and data documents marked up with vocabulary terms will list the term URI as a minimum and this term should be resolvable at the given address.
10. Metadata and dataset schema should not be brought into common operation by the community unless the requisite vocabularies and their terms are available to the community. It is recognized that no vocabulary list can ever be considered complete, but the vocabulary must have a number of core terms and be subject to the actions identified in (1) above (before being mandated for use).
11. The introduction of SOS service endpoints (for already filtered and subsetted data) as legitimate data outputs from the 123 Portal will necessitate the tagging of ingested data with agreed vocabularies (NOT JUST THE DATASET METADATA). This tagging should ideally be done at time of creation and included in the data file submitted to IMOS but could be added later at the time of ingestion into the IMOS/AODN portal database. In reality all data should be tagged inline with necessary controlled vocabularies whether they become part of a SOS endpoint or not. Any resultant SOS data output documents must carry the URIs for terms in these controlled vocabularies (inclusive of any linked SensorML documents).
12. Figure 1 separates a vocabulary management component from a Sissvoc server component, however, it is feasible to have both the services and functions provided by Sissvoc software and a vocab manager tool respectively (as described in the use-cases) combined into one product/platform (e.g. see the GBIF vocabulary portal ViBRANT which is based on semantic media wiki <http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_D4.2b%20—%20Ontology%20Tools-%20Status%20report%202012.pdf> Note it is possibly not ISO 19135 compliant).

**Necessary Tasks To Introduce Controlled Vocabulary Use**

1. Establish a Vocabulary Management Tool with functionality as defined earlier in this paper, which can leverage some of the capability of Sissvoc (or similar tool designed for the publication and machine querying of terms). A new IMOS Sissvoc instance and the Vocabulary Management tool should be tightly coupled.
2. In concert with task (1) above, design the schema for various content files (e.g. in Skos, RDFS) which can represent the vocabularies in an appropriate triple store (e.g. sesame) to permit addressable term publication, linkages to other terms in the same or other repository and to support thesaurus publication in GeoNetwork.
3. Establish an IMOS Sissvoc instance (or similar tool) that can respond to queries on:
4. Parameter Names (couched as n-ary constructs from the Observable Properties vocabulary – possibly also mapped to NERC vocabularies and others),
5. Platforms
6. Instruments
7. Analytical Methods
8. Organisations (may submit these instead to EDMOD for management)
9. Determination Methods (which probably need to align with an O&M and OP concept of ‘Procedure’ – and can also be an n-ary construct of platform, instrument and ‘analytical method’ vocabs)
10. Geographic Region (unless we push this off to someone else as a gazetteer task)
11. Feature Type (but we should liaise with eREEFs to make sure our definitions of ‘Feature Type’ are semantically clear and do not contradict, or over-lap with terms in a Geographic Region vocabulary).
12. Liaise with eREEFs to ensure that the OP vocabulary includes the supporting terms that we need to express our ‘Parameter Names’.
13. Port the IMOS/AODN MEST to MCP version 2.0 and GeoNetwork 10.X so that the Thesaurus vocabulary picker can be used to mark-up dataset metadata. The HTML5 (in 10.X) interface also improves the look and feel of the tool and usability.
14. Refine (customise) each GeoNetwork vocabulary picker to ensure that the desired elements of individual vocabularies are exposed at selection time to make selection easy for the user. Ensure that a picker is available to provide content for any term in a metadata schema – NOT just those defined as ‘Keywords’.
15. Review how the GeoNetwork metadata editor permits completion of potentially repetitive metadata schema information blocks (e.g. where the same Parameter type is collected by several different instruments/platforms within a described dataset).
16. Mark-up existing IMOS (data) database records with appropriate controlled vocabularies. Some of this work is underway and has been approached in the manner shown in the workflow in Figure 2. Ideally, in the future, the workflow to ingest any new data into the database should more closely resemble that depicted in Figure 3.
17. Review the MCP 2.0 ‘Data Parameters’ encoding to ensure an ability to include Observable Properties compatible vocabularies.
18. Develop conventions/guides/examples for community on how to mark-up typical datasets including advice on metadata granularity and the use of taxonomy elements.
19. Coordinate the community mapping of existing MCP1.4 vocabularies to vocabularies identified for use in MCP 2.0.
20. Update the AODN Cookbook.

Figure 2 – Workflow For Existing Mark-Up And Ingestion Of Data Into IMOS Database Tables

publishes

Various data outputs

Portal 123

delivers WMS, WFS & annotated CSV data

GeoServer

uses

Controlled Vocabulary DB

Manual Vocabulary Mapping Process

(maps informal variable names to agreed vocabulary terms)

data upload

PostGres DB

data available to

(Public) THREADS

Server

harvests data after

data subjected to

netCDF data migration to

Talend Harvester

(1 per type of dataset collection)

AccessDB data subjected to

Manual Vocabulary Mapping Process

(for creating within-datafile consistency in term usage)

data upload

OpenDAP Server

IMOS Private Fileshare

(staging place for beginning ingestion process)

**IMOS Facility**

(currently creating variable names that are not standardized but do follow IMOS instruction manual)

Raw Data

NetCDF

Access Db

Excel Files

Excel data subjected to

Figure 3 – Suggested Workflow For Mark-Up And Ingestion Of Data Into IMOS Database Tables

publishes

Various data outputs

Portal 123

delivers WMS, WFS & annotated CSV data

Registers service in

Reference vocabulary URIs

data & services subjected to

OGC Service

netCDF data migration to

uses

data available to

Controlled Vocabulary DB/

SissVoc Server

harvests Excel data after validation

harvests AccessDB data after validation

(Public) THREADS

Server

sent back to

Rejected datasets

uses

uses

Dataset Validation Tool

(Talend ?)

(for validating (& correcting ?) the format and vocabularies used in submitted files)

OpenDAP Server

GeoServer

data upload

PostGres DB

harvests data

Talend Harvester

(1 per type of dataset collection)

data upload

IMOS Private Fileshare

(staging place for beginning ingestion process)

**IMOS Facility**

(create conformant datasets according to agreed schema and vocabularies)

Raw Data

NetCDF

Access Db

Excel Files