Managing and sharing common vocabularies through NVS services

...OPTION FOR MANAGING ARGO VOCABS?

Presented by Robin McCandliss - using slides from Gwen Moncoiffe (Leader of NVS service at BODC)
Why use vocabularies?

Divided by a common language
What is the NERC Vocabulary Server (NVS)?

Essentially a server that provides persistent URIs for:

• vocabulary term (often referred to as “concept”)
• vocabulary lists (collections of related concepts)
• vocabulary thesaurus (aggregation of concepts from one or more collections)
What is the NERC Vocabulary Server (NVS)?

By extension, it also refers to a set of web services and applications designed to manage and share vocabularies served by the NERC Vocab Server;

They can be shared as a trusted resource because:
• Each vocabulary is referenced as a unique resource identifier (URI);
  e.g. http://vocab.nerc.ac.uk/collection/L22/current/TOOL0668/
• The identifier provides access to machine and human readable information;
• The information is structured and managed according to W3C standards;
• The NVS supports mapping to internally and externally hosted vocabularies;
• Once published the URIs are permanent;
• They can be deprecated but never removed or redefined;
Structure based on SKOS standard

Simple Knowledge Organisation System (SKOS)

A concept has:

• Identifier e.g. “31”
• Preferred Label e.g. “research vessel”
• Alternative Label -
• Definition e.g. “A platform of any size operating on the surface of the water column in unpredictable locations that is specifically equipped, manned and operated for scientific, usually oceanographic, research.”

http://vocab.nerc.ac.uk/collection/L06/current/31/
An example from the SeaVox platform categories vocabulary (L06)
An example from the SeaVox Device Catalogue (L22)

**-- Sea-Bird SBE 41 CTD --**

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>URI</td>
<td><a href="http://vocab.nerc.ac.uk/collection/L22/current/TOOL0668/">http://vocab.nerc.ac.uk/collection/L22/current/TOOL0668/</a></td>
</tr>
<tr>
<td>Identifier (SDN)</td>
<td>L22::TOOL0668</td>
</tr>
<tr>
<td>Preferred label</td>
<td>Sea-Bird SBE 41 CTD</td>
</tr>
<tr>
<td>Alternative label</td>
<td>SBE 41 CTD</td>
</tr>
<tr>
<td>Definition</td>
<td>A self-contained unit comprising the MicroCAT temperature, conductivity and pressure sensors and a pump that is designed specifically for deployment on profiling floats, particularly Argo. The unit is designed to provide stable salinity data accurate to 0.005 PSU for periods in excess of three years without any form of maintenance. Temperature is within 0.002 °C (stability 0.0002 °C/year) and pressure within 2 dbar (stability 0.8 dbar/year). During float ascent spot samples are taken and transmitted to the float controller. More information is given in <a href="http://www.seabird.com/products/spec_sheets/41data.htm">http://www.seabird.com/products/spec_sheets/41data.htm</a>.</td>
</tr>
<tr>
<td>Version Info</td>
<td>1</td>
</tr>
<tr>
<td>Deprecated</td>
<td>false</td>
</tr>
<tr>
<td>Broader (L05)</td>
<td><a href="http://vocab.nerc.ac.uk/collection/L05/current/134/">http://vocab.nerc.ac.uk/collection/L05/current/134/</a></td>
</tr>
<tr>
<td>Broader (L05)</td>
<td><a href="http://vocab.nerc.ac.uk/collection/L05/current/130/">http://vocab.nerc.ac.uk/collection/L05/current/130/</a></td>
</tr>
<tr>
<td>Broader (L05)</td>
<td><a href="http://vocab.nerc.ac.uk/collection/L05/current/350/">http://vocab.nerc.ac.uk/collection/L05/current/350/</a></td>
</tr>
<tr>
<td>Date</td>
<td>2014-03-11 15:55:41.0</td>
</tr>
</tbody>
</table>

Is broadly related with “Temperature sensor”
Is broadly related with “CTD”
Is broadly related with “Salinity sensor”
Management tools

Externally governed vocabularies are increasingly managed via the NVS VocabEditor –
Allow external editors to submit and edit new terms and mappings to vocabularies they are
authorised to access (access control)
https://www.bodc.ac.uk/resources/vocabularies/vocabulary_editor/
Content and technical governance

- **Collections and governance**
  - 231 vocabulary collections accessible through the NVS
  - 70 owned and governed by BODC
  - 48 are managed by BODC on behalf of SeaDataCloud, EMODnet, SeaVox, SWE content governance authorities
  - Remainder (113) are owned by 25 different governing authorities

- **Technical governance and support**
  - assured by BODC
  - On-going work improving access to governance and version information, improving user interfaces (part of the European SeaDataCloud project)
Steps to hosting vocabularies on the NVS

Agree a governance model e.g.
• an ARGO vocabulary governance committee?
• one or multiple authorised vocabulary managers including a member of the BODC Vocab Management team

Identify collections and make them SKOS compliant by providing:
• Title e.g. Climate and Forecast Standard Names
• Short name e.g. CF Standard Names
• Definition e.g. Terms used for definitive but not necessarily comprehensive descriptions of measured phenomena in the CF conventions (a content standard for data stored in NetCDF)

Ensure concepts are unambiguous by providing:
• A preferred label
• An optional alternative label
• A definition description of the term

See for example: http://vocab.nerc.ac.uk/collection/W03/current/
Steps to hosting vocabularies on the NVS

- BODC Vocabs Team provides infrastructure and expert advice
- Community develops the vocabularies (terms, definitions etc.)

- Community to identify Argo vocabs for management
- Community to ensure Argo vocabs are clean, well-defined and SKOS compliant

- Could pilot taking an example of a good, well-defined Argo vocab into NVS?
- Get Argo vocabs into good shape beforehand \(\rightarrow\) reduces effort to get them into NVS

- Note: If significant BODC effort required to get Argo vocabs into NVS, then this would need funding
- Will require time and effort over the longer term