

Controlled Vocabulary Practical Exercise with Answers

SeaDataNet Vocabulary Identification

Objective of this exercise is to get people looking through the list of vocabularies in the Maris client, opening some of them up to see what's inside. It also introduces the issue of parameter and instrument vocabulary granularity for CDI and data files. Finest available granularity is used for usage metadata (i.e. data files) with coarser granularity for discovery metadata. It also introduces the idea of using 3-byte labels to refer to vocabularies.

Which SeaDataNet vocabulary would be used to populate the following metadata model or data file field?

- Cruise Summary Report (CSR) cruise start and end ports **C38**
- The country in which a port is located **C32**
- Sea areas in an EDMED record (clue: it's not C16) **C19**
- Spatial feature type in CDI record **L02**
- Common Data Index (CDI) instrument type **L05**
- ODV or NetCDF data file instrument semantic linkage **L22**
- EMODNET product parameter **P35**
- CDI parameter **P02**
- ODV or NetCDF data file parameter semantic linkage **P01**

REST Time

The purpose of this exercise is to get people used to the RESTful API syntax and reading RDF XML documents.

- Obtain the RDF XML catalogue of the controlled vocabularies held in the NERC Vocabulary Server <http://vocab.nerc.ac.uk/collection/> *Worth noting that the returned RDF XML includes information on the vocabulary governance and also documents the vocabularies to which concepts in the featured vocabulary are mapped.*
- Obtain the RDF XML document listing the deprecated concepts in the P02 vocabulary <http://vocab.nerc.ac.uk/collection/P02/current/deprecated/> *Worth looking through to get an idea as to what has been deprecated.*
- What accepted P02 code should be used instead of the deprecated P02 code MTSD? *Search out MTSD in the P02 deprecated term listing and then the dc:isReplacedby element which holds the URL for the replacement concept (RMIN). Worth following this up to see what it is (Inorganic chemical composition of sediment or rocks) and noting the fact that semantically this is a much broader concept, which shows that MTSD was deprecated as part of an attempt to keep the size of P02 under control.*

Instrument Mappings

There are two objectives to this exercise. One is to introduce searching the vocabularies - in this case L05 - using the SeaDataNet client to find a given concept within a vocabulary. The second - not covered in my presentation - is to introduce the idea of concept URIs and how to convert the ConceptID for the concept into a URN (used in data files) and a URL (RESTful API call). It also brings some familiarity to the instrument vocabularies.

- *What are the L05 URIs (URN and URL) for a dissolved oxygen sensor? The concept ID is easily found (search for dissolved or oxygen) to be '351'. To get a URN concatenate 'SDN:' to vocabulary label to ':' to concept ID - i.e. SDN:L05::351. To get a URL concatenate 'http://vocab.nerc.ac.uk/collection/' to vocabulary label to '/current/' to conceptID - i.e. <http://vocab.nerc.ac.uk/collection/L05/current/351>*
- *What dissolved oxygen sensors are included in L22? Two ways of doing this. One is to retrieve the RDF XML document from the dissolved gas sensor URL and then follow up each of the L22 mappings. There's 23 of them, which is somewhat boring and repetitive. An easier way is to open up Chrome or Firefox with the URL <http://vocab.nerc.ac.uk/scheme/SDNDEV> and drill down through the result that appears after about a minute (IE doesn't handle the style sheet properly). Delay is because there's a large XML document to render locally.*
- *What are the L22 and L05 URNs for a Sea-Bird SBE9-11plus CTD? Easiest way to do this is to use the SeaDataNet client to get the L22 conceptID(TOOL0058) which becomes SDN:L22::TOOL0058 as a URN and <http://vocab.nerc.ac.uk/collection/L22/current/TOOL0058> as a URL. Resolve the URL and the RDF XML document shows the three L05 concepts (SDN:L05::130, SDN:L05::350 and SDN:L05::134).*

Parameter mappings

This is largely practice in searching the SeaDataNet client, but includes some examples of the difficulties resulting from parameter semantics.

- Find an appropriate P01 URN to match to each of the following local parameters:
 - *Concentration of suspended particulate material (GF/C filtered) Pretty straightforward. Using the search string 'conc%susp%part%mat%GF/C' brings back six hits. Only watch point is to avoid 'organic' and 'inorganic'. The required URN is SDN:P01::TSEDGVP3. It includes method information, but it's pretty superfluous - can't think of any other way - and I want to gently encourage a move away from depending on methodology information in parameter codes.*
 - *PCB101 in sediment fines (<63um) Again straightforward - search on 'PCB101%sed%63'. In this case there is a method-independent code. Worth noting the three different ways the chemical is described (IUPAC name, PCB and CAS numbers) and promoting CAS numbers as a reliable search criterion for chemicals in P01. Required URN is SDN:P01::CB101SAD.*
 - *Fluorometric chlorophyll-a (acetone extraction GF/F filtered) Again straightforward except that the acetone and fluorometric need to come at the end of the search string not the beginning - chl%-a%GF/F%acetone%fluor - works. There's a few*

possibilities quoting different protocols, but my suggestion would be SDN:P01::CPHLFLP1.

- Temperature (C) *The point I'm making here is temperature of what (water body, atmosphere, conductivity measurement). Without more information it's impossible to give an answer.*
- Phosphate (uMol/l) in an unfiltered sample *There are a couple of points in this example. First, is that the units indicate we are looking for a concentration per unit volume. The second is to introduce people to the way filtered and unfiltered samples are handled in P01. Try the search term 'concentration of phosphate%vol%water body' which gives 18 hits. Scan down these looking at the bit in square brackets:*
 - *[dissolved plus reactive particulate phase] = unfiltered*
 - *[dissolved plus reactive <unknown phase] = filtered but filter type is unknown*
 - *[unknown phase] = don't know whether it was filtered*
 - *[dissolved plus reactive <xxx phase] = filtered using filter type xxx*

For nutrients this isn't that important because the particulate phase has very little influence on dissolved nutrient concentrations, but for metals the presence and nature of particles in the analysed sample can make a huge difference. The URN I'd choose is SDN:P01::PHOSYYDZ.

- Dissolved zinc (uMol/kg). Dissolved defined as 0.2um filtered. *This is reinforcing the same points as the previous example. In this case the units are per kg, indicating that we want a concentration per unit mass. The search string 'concentration of%zinc%mass%water body%0.2' return two hits. There's no method-independent code so for the purpose of the exercise either is OK.*
- Particulate cadmium (ppb) *Again the units provide the clue to what is going on - it's a mass/mass unit so what we almost certainly have is cadmium concentration in SPM. A search for 'cadmium%susp%mat%' gives a seven hits. Best answers are SDN:P01::MSPMM005 and SDN:P01::MSPMM017.*
- Concentration of cadmium in mussels (ug/kg) *Another impossible one. There's two species of mussel - M edulis and M galloprovincialis - so more information is needed. Also not clear whether the cadmium is in the flesh or shells of the mussels. Many questions need to be asked when mapping parameters.*
- Concentration of mercaptothion in filtered water (ug/l) *This is a bit of a stinker that is designed to show the value of CAS numbers when searching for organics. A search for 'mercaptothion' finds absolutely nothing. However mercaptothion is one of the many names for malathion. Strategy is to use Google to find the CAS number for mercaptothion (search term mercaptothion CAS) which gives us CAS 121-75-5. Searching P01 for '121-75-5%water body' gives four hits, one of which is appropriate for filtered water unknown filter type (SDN:p01::MALAFX1).*
- Optical backscatter at 470 nm at 117 degrees (per m) *This is another example that requires considerable knowledge of optics. The issue is 'What is optical backscatter?' Is it the proportion of transmitted light received by a transceiver? In fact it turns out that it is loose talk for 'attenuance attributed to optical backscatter'. Knowing that a simple search for '470%117%' finds SDN:P01::BB117B01. The whole point of this is*

that a single thing can be described in many different ways and it is essential to take this into account when mapping.

Green Dogs



Which of the following parameter descriptions are the six 'Green Dogs' and why?

- Practical salinity standard deviation of the water body by CTD and computation using UNESCO 1983 algorithm and calibration against independent measurements
- Practical salinity per unit dry weight of sediment <63um
 - *Fairly obvious that the concept of salinity is inapplicable to a dry powder*
- Concentration of *Thalassiosira gravida* (ITIS: 2490; WoRMS 149102) [Size: 25um] per unit volume of incubated water sample by optical microscopy
 - *Concept of concentration doesn't apply cleanly to a taxon*
- Concentration of 2,3',4,4',5-pentachlorobiphenyl {PCB118 CAS 31508-00-6} per unit volume of the water body [dissolved plus reactive particulate <GF/F phase]
- Concentration of 2,3',4,4',5-pentachlorobiphenyl {PCB118 CAS 31508-00-6} per unit volume of the water body [dissolved plus reactive particulate >GF/F phase]
 - *Impossible for a filtrate to have particles above a given threshold - they'll be in the residue*
- Concentration of 2,3',4,4',5-pentachlorobiphenyl {PCB118 CAS 31508-00-6} in the water body [dissolved plus reactive particulate <GF/F phase]
- Abundance of *Thalassiosira gravida* (ITIS: 2490; WoRMS 149102) [Size: 25um] per unit volume of incubated water sample by optical microscopy
- Count of zinc per unit dry weight of sediment <63 um
 - *Concept of count inappropriate for a chemical species*
- Count of *Demospongiae* (ITIS: 47528; WoRMS 164811) [Size: height ~10cm, diameter 30cm Morphology: mass of branches Colour: brown] on the bed by image analysis
- Temperature
 - *Temperature is meaningless without a matrix (e.g. water body or atmosphere)*