

ODIP Common Vocabularies

Roy Lowry – BODC

Kim Finney – IMOS

Adam Leadbetter – BODC

Karen Stocks - SDSC

Presentation Objectives

- To focus on the usage of vocabularies, rather than their creation and management
- To avoid repeating what has been said before
- Presentation Facets
 - Semantic Aggregation for EMODNET
 - The Australian/US Perspectives
 - Sharing vocabularies with ‘Sparkling Interoperability’

Semantic Aggregation

- SeaDataNet parameter mark-up uses the BODC Parameter Usage Vocabulary (Po1)
 - Designed for marking up measurement values
 - Very fine-grained
 - Many concepts carry methodology information
 - Significant historical baggage
 - Duplicate concepts
 - Granularity too fine
 - Common physical parameters like water temperature and salinity are some of the worst affected

Semantic Aggregation

- The primary aggregation tool in SeaDataNet is ODV
- ODV uses the SeaDataNet semantic mark-up and creates a data channel in an aggregated collection for every Po1 code in the input files
- This can lead to problems, particularly when different codes have been used unnecessarily for the same thing

Semantic Aggregation

- Aggregation is an essential component of EMODNET product generation
- During the pilot, two parameter semantic questions were asked, but not necessarily answered
 - What Po1 codes should be included in a given product parameter?
 - What label should be applied to that product parameter?

Semantic Aggregation

- Controlled vocabularies could provide a powerful and elegant solution
- This has been suggested several times in meetings, but never presented in detail
- Until now.....

Semantic Aggregation

- The scientific experts responsible for the aggregation need to provide definitive answers to the parameter semantic questions
 - Agree on a name for the aggregated parameter
 - Agree on the Po1 concepts that make it up
- The resulting knowledge needs to be made universally available in an encoding that software agents can understand

Semantic Aggregation

- A similar use case was encountered in NETMAR with semantic validation of service chains
 - Service outputs are marked up with Po1
 - Valid inputs for a service are defined as a group of Po1 codes
 - This Po1 grouping is given a name from a new controlled vocabulary (P25)
- <http://vocab.nerc.ac.uk/collection/P25/current/NO3/>
- This generates a SKOS document that specifies 'Nitrate concentration' in terms of concepts from Po1 **and the CF Standard Names**
- It also includes support mappings for faceted discovery, but that's a different story

Subset of the SKOS document returned by the URL

- Semantics of label presented as a plaintext definition
- Component concepts taken from SeaDataNet AND CF Vocabularies

```
<skos:Concept rdf:about="http://vocab.nerc.ac.uk/collection/P25/current/NO3/">
<skos:prefLabel xml:lang="en">Nitrate concentration</skos:prefLabel>
<skos:altLabel/><skos:definition xml:lang="en">The amount (mass or quantity of matter) of nitrate
per amount (volume or mass) of a body of water</skos:definition>
<dc:identifier>SDN:P25::NO3</dc:identifier>
<skos:notation>SDN:P25::NO3</skos:notation>
<owl:versionInfo>2</owl:versionInfo>
<dc:date>2011-11-24 11:34:12.0</dc:date>
<skos:note xml:lang="en">accepted</skos:note>
<owl:deprecated>>false</owl:deprecated>
<skos:narrower
rdf:resource="http://vocab.nerc.ac.uk/collection/P01/current/MDMAP005/"><skos:narrower
rdf:resource="http://vocab.nerc.ac.uk/collection/P01/current/NTRAYYDZ/"><skos:narrower
rdf:resource="http://vocab.nerc.ac.uk/collection/P01/current/CHEMM012/"><skos:narrower
rdf:resource="http://vocab.nerc.ac.uk/collection/P01/current/NO3UFLIC/"><skos:narrower
rdf:resource="http://vocab.nerc.ac.uk/collection/P01/current/NTRAZZXX/"><skos:narrower
rdf:resource="http://vocab.nerc.ac.uk/collection/P01/current/NTRZLWTX/"><skos:narrower
rdf:resource="http://vocab.nerc.ac.uk/collection/P07/current/CFSN0515/">
<skos:Concept>
```

Semantic Aggregation

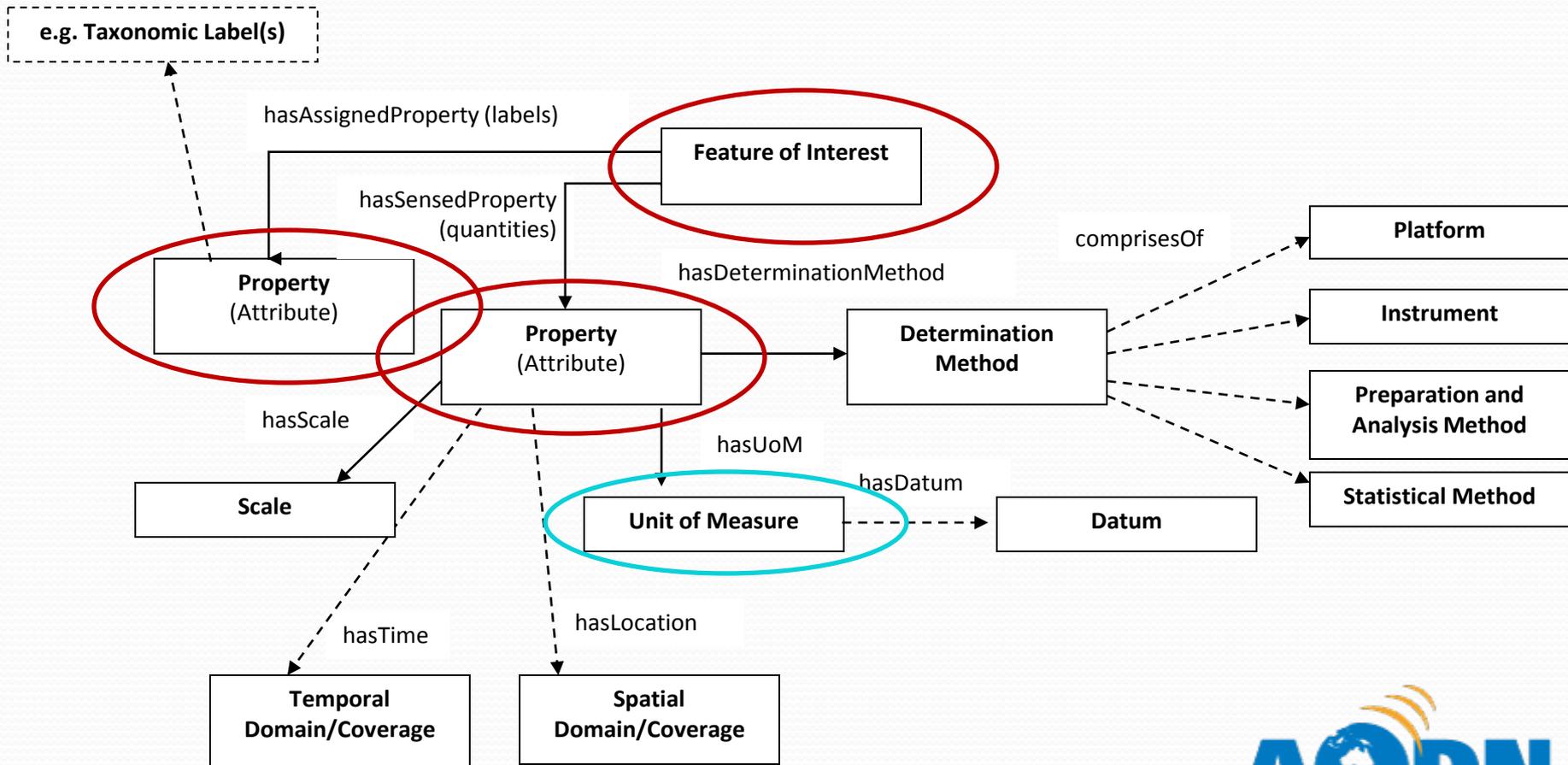
- Domain expert content governance specifies the product labels and composition in terms of mark-up concepts
- Interoperability is delivered because the components come from ANY parameter vocabulary
- Technical governance delivers machine-readable documents that:
 - Are universally available at the end of a URL
 - Can be used to drive data aggregation tools
 - Can be used to populate usage metadata

Common Vocabularies

- IMOS currently working on vocabularies for use in the AODN infrastructure. Three types of significant interest are:
 - *Parameter Names* (and associated *Determination Methods* – Instrument, Platform, Analysis Method, Statistical Method, plus *UoM*)
 - *Taxonomic Names*
 - *Geographic Names* (marine and terrestrial)
- AODN Portal-based data exchange infrastructure hinges on the availability of descriptive metadata:
 - in ISO 19115 compliant metadata - MCP Profile (analogous to CDI Profile);
 - in Web Map Service capability documents.
- AODN infrastructure doesn't directly use any 'agreed' inline data (file) metadata mark-up (except in IMOS-derived netCDF files).
- Metadata records and data files are created and provided by data generators. Metadata is harvested to a central repository to drive the infrastructure. Data (services) generally reside at point of origin and are accessed via reference from a central AODN Portal. Discovered data not conformant to any mandated standard formats or vocabularies (except a number of IMOS-centric data services).



Parameter Names



Taxonomic Names

Current prototype Taxonomic Name element as proposed by Simon Pigot (uses Taxon Concept Schema – but suggestion (By Tony Rees) is to move to Darwin Core (DwC).

```
<mcp:taxonomicElement>
  <mcp:EX_TaxonomicCoverage>
    <mcp:taxonConcepts xlink:href="http://biodiversity.org.au/apni.name/38090.xml"/>
  </mcp:EX_TaxonomicCoverage>
```

.....

XML with XLink resolved (what everyone sees):

```
<mcp:taxonomicElement>
  <mcp:EX_TaxonomicCoverage>
    <mcp:taxonConcepts>
      <app:documents ....>
        <TaxonName....">
          <dc:title>Eucryphia lucida (Labill.) Baill. var. lucida</dc:title>
          <dc:date>20101013T17:15:09.763004+11:00</dc:date>
          <dcterms:modified>20090302</dcterms:modified>
          <ibis:NameComplete>Eucryphia lucida (Labill.) Baill. var.
lucida</ibis:NameComplete>
          <ibis:NameString>Eucryphia lucida var. lucida</ibis:NameString>
          <NomenclaturalCode>Botanical</NomenclaturalCode>
```

.....

Geographic Extent Names

- MCP has an inadequate codelist for geographic extent names.
- Should be co-ordinating nationally through efforts with the Australian National Gazetteer (parties with active naming services are AAD, GA and the Navy).
- These national efforts should be dovetailing with SEAVOX and

<http://www.marineregions.org/about.php>

Term Mapping

- Need to map an 'updated' MCP schema to CDI schema, particularly codelists of significance. Don't anticipate too many 'mapping' problems at this level. Probably not hard to engineer data discovery using vocabs that will be common across AODN and say, SEADATANET.
- Once we get to the point, however, of compiling datasets into common deliverable formats for delivery through the various infrastructures, issues become more complex (e.g. SEADATANET uses ODV harnessing PUV, AODN has no mandated format).
- SKOS mapping documents could be used to concatenate AODN parameter -related elements and map them to matching PUV terms.

R2R Vocabulary Perspective



- R2R relies on a suite of vocabularies
 - Some are local: e.g. meteorological parameters, cruise, device Type, Device Model, person
 - Some are community: e.g. IHO Sea Areas, ICES Vessels, UNOLS ports
 - See <http://www.rvdata.us/voc> for details
- Vary from informal, to SKOS thesauri, to OWL ontologies, depending on history and use



R2R Proposed ODIP Vocab. Work

Moving toward community vocabularies

- Mapping to SDN (and suggesting additions/expansions as needed.) Including mapping to
 - Device Categories (L05)
 - BODC Instrument List (L22)
 - Parameter Groupings (P02)
 - Units (P06)
 - Port Gazetteer (C38)
 - SDN/ODS QC flags (for SAMOS near-real-time meteorological QC flags)
- Publishing mappings in SKOS
- Collaborating on community device “action” vocabulary for Event Logger

Federated SPARQL

- What?
 - A SPARQL endpoint is an on-line query interface that allows the content of a semantic resource to be queried using the W₃C SPARQL query language.
 - Federation directs a portion of a query at a remote SPARQL endpoint
 - Results are returned and combined with results from the rest of the query
 - Many endpoints may be included
 - Experience says: need a (very) specific result set for performance reasons

Federated SPARQL

- Why?
 - Search multiple data sources at once
 - Visualise semantic interoperability
 - E.g. BCO-DMO and R2R, linked through mappings to NERC Vocabulary Server
 - Combine mappings implicitly
 - R2R maps to NVS, not the other way round

Federated SPARQL

- Technicalities
 - Federated query uses the SERVICE keyword
 - SERVICE is an extension to SPARQL 1.1
- Example
 - R2R map instruments to SeaDataNet device categories
 - Get instrument categories from NVS
 - Get instrument names from R2R

Federated SPARQL

```
PREFIX skos:<http://www.w3.org/2004/02/skos/core#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?nvsp ?rvl WHERE
{
    SERVICE <http://linked.rvdata.us/sparql>
    {
        ?rvs rdf:type ?nvs
        FILTER CONTAINS (str(?nvs),'vocab.nerc.ac.uk').
        ?rvs rdfs:label ?rvl
    }
    ?nvs skos:prefLabel ?nvsp
} ORDER BY lcase(str(?nvsp))
```

Execute

SPARQLer Query Results

nvsp	rvi
"anemometers" @en	"Gill WindObserver II anemometer on Oceanus (bow)"
"anemometers" @en	"RM Young 5103 anemometer on Cape Hatteras (port)"
"anemometers" @en	"RM Young 5103 anemometer on Cape Hatteras (starboard)"
"anemometers" @en	"RM Young 5103 anemometer on Marcus G. Langseth"
"anemometers" @en	"RM Young 5103 anemometer on Wecoma (mainmastport)"
"anemometers" @en	"RM Young 5103 anemometer on Wecoma (mainmaststarboard)"
"bathythermographs" @en	"Sippican MK12 expendableprobe on Atlantis"
"bathythermographs" @en	"Sippican MK12 expendableprobe on Cape Hatteras"
"bathythermographs" @en	"Sippican MK12 expendableprobe on New Horizon"
"bathythermographs" @en	"Sippican MK12 expendableprobe on Point Sur"
"bathythermographs" @en	"Sippican MK12 expendableprobe on Robert Gordon Sproul"
"bathythermographs" @en	"Sippican MK21 expendableprobe on Endeavor"
"bathythermographs" @en	"Sippican MK21 expendableprobe on Healy"
"bathythermographs" @en	"Sippican MK21 expendableprobe on Hugh R. Sharp"
"bathythermographs" @en	"Sippican MK21 expendableprobe on Kilo Moana"
"bathythermographs" @en	"Sippican MK21 expendableprobe on Knorr"
"bathythermographs" @en	"Sippican MK21 expendableprobe on Marcus G. Langseth"
"bathythermographs" @en	"Sippican MK21 expendableprobe on Melville"
"bathythermographs" @en	"Sippican MK21 expendableprobe on Pelican"