

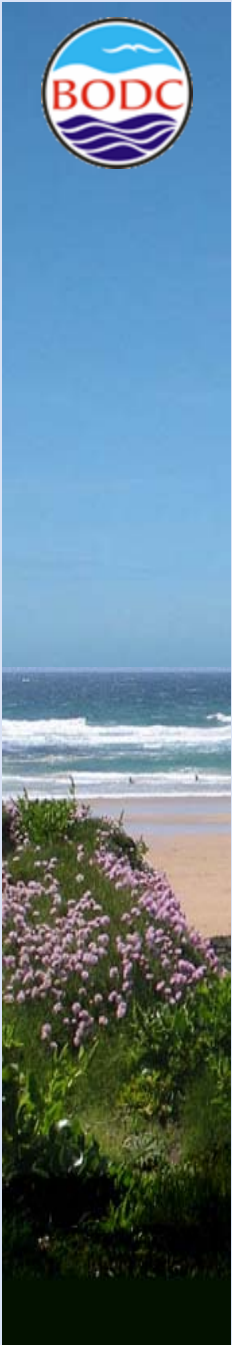


IMDIS 2008 Conference, Athens, 31 March – 2 April 2008

Putting Meaning into SeaDataNet

Roy Lowry & Geoff Williams

British Oceanographic Data Centre





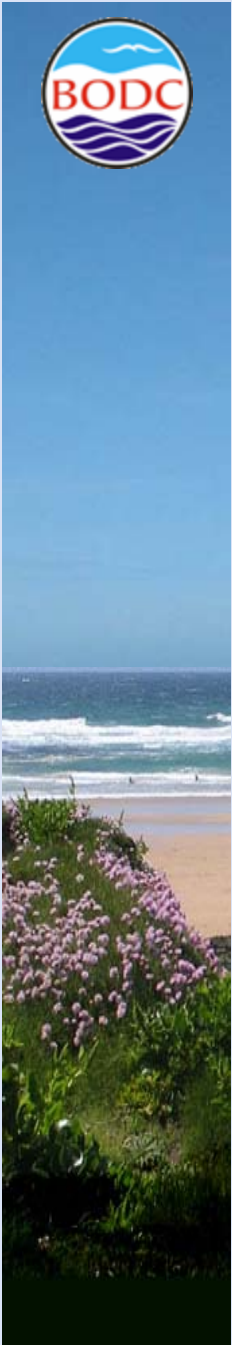
Presentation Summary

- **Weaknesses of plaintext and poorly managed controlled vocabularies**
- **Semantic markup in SeaDataNet data and metadata**
- **URN to URL resolution**
- **NERC DataGrid Vocabulary Server**
- **Semantic cross-walk Use Case**
- **Conclusions**



The Bad Old Days

- **Plaintext used extensively to populate metadata**
 - Works for humans but not intended software agents
- **Some fields were populated from controlled vocabularies**
- **However, the vocabularies were:**
 - Poorly managed
 - Lacking in definitions
 - Abused to facilitate shoe-horning





SDN Semantic Markup

- **SeaDataNet now has governed vocabularies in a managed semantic technical infrastructure (NDG Vocabulary Server)**
- **SeaDataNet semantics in data and metadata are represented by permanent machine-readable labels (URNs)**





SDN Semantic Markup

- **An example URN is SDN:P021:21:TEMP**
 - **Where:**
 - **SDN = namespace identifier**
 - **P021 = resource identifier**
 - **21 = resource version number**
 - **TEMP = resource component identifier**



From URN to URL

- **Some simple string substitution converts this URN into something that may be found on the internet (URL)**
 - **http://vocab.ndg.nerc.ac.uk/ = namespace**
 - **term/P021 = resource**
 - **/21 = version**
 - **/TEMP = resource component identifier**
- **This gives**
 - **<http://vocab.ndg.nerc.ac.uk/term/P021/21/TEMP>**
- **Which is an NDG Vocabulary Server term identifier that returns a SKOS document describing the concept and its mappings**
- **A SeaDataNet URN to URL translation service is planned**





The NDG Vocabulary Server

- This is a Semantic Web resource for the technical governance of controlled vocabularies
- Built by BODC as part of the NERC DataGrid project
- Adopted by, and being further developed for, SeaDataNet
- Fundamental entity is a concept that is represented by an identifier, a term, an abbreviation and a definition
- Concepts are organised into lists, designed to populate particular metadata elements



The NDG Vocabulary Server

- Each concept has a web presence (i.e. a URL) corresponding to a dynamically generated SKOS (RDF-based) XML document
- This document delivers the concept identifiers and labels plus its mappings to other concepts
- For example, the SeaDataNet platform class category concept 'amphibious vehicle' is represented by the URL:
 - <http://vocab.ndg.nerc.ac.uk/term/L062/current/95>
- This delivers an XML document thus....





The NDG Vocabulary Server

```
<?xml version="1.0" ?>
```

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:skos="http://www.w3.org/2004/02/skos/core#" xmlns:dc="http://purl.org/dc/elements/1.1/">
```

```
<skos:Concept rdf:about="http://vocab.ndg.nerc.ac.uk/term/L062/2/95">
```

```
<skos:externalID>SDN:L062:2:95</skos:externalID>
```

```
<skos:prefLabel>amphibious vehicle</skos:prefLabel>
```

```
<skos:altLabel />
```

```
<skos:definition>A self-propelled platform capable of operating on land and within or on the surface of a water body.</skos:definition>
```

```
<dc:date>2008-02-26T14:25:51.822+0000</dc:date>
```

```
<skos:narrowMatch rdf:resource="http://vocab.ndg.nerc.ac.uk/term/L061/6/9A" /> !DUKW
```

```
<skos:narrowMatch rdf:resource="http://vocab.ndg.nerc.ac.uk/term/L061/6/9B" /> !Hovercraft
```

```
</skos:Concept>
```

```
</rdf:RDF>
```



The NDG Vocabulary Server

- **The server has other more sophisticated access mechanisms designed to facilitate bulk access**
 - **URL access to SKOS lists (no mappings)**
 - **HTTP-POX API (see http://www.bodc.ac.uk/products/web_services/vocab/methods.html)**
 - **SOAP API and method interface clients (see <http://vocab.ndg.nerc.ac.uk/>)**





The NDG Vocabulary Server

- **The server currently (2008-02-25) holds**
 - 105 public lists
 - 120878 concepts
 - 75766 mappings (RDF triples)

- **New content added daily**

- **Semantic content heavily (>200,000 hits per month) mined by robots**

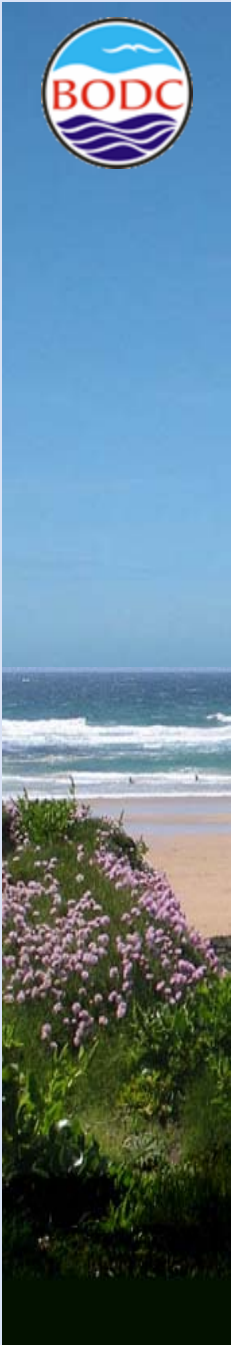
- **Typical 'human' activity currently running at approximately 400 catalogue and 3000 list accesses per month**





Semantic Crosswalk Use Case

- **BODC wishes to produce a GCMD DIF document from an EDMED V1.2 document**
- **The “parameter” sections of the two documents are populated using different vocabularies (BODC PDV and GCMD Science Keywords)**
- **This situation was usually addressed by having no parameter section in the output document**
- **We can now do better.....**

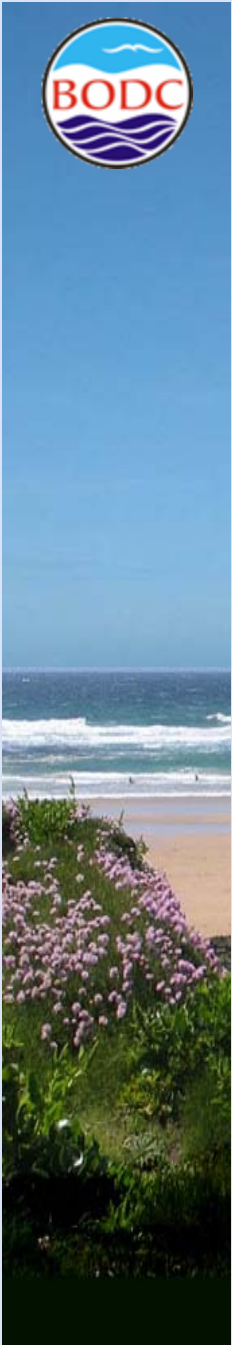




Semantic Crosswalk Use Case

- **A list of BODC PDV terms as parameter URNs is obtained from the EDMED document, for example:**
 - * SDN:P021:24:TEMP, SDN:P021:24:PSAL , SDN:P021:24:CPWC

- **This may then translated into a list of URLs**
 - * <http://vocab.ndg.nerc.ac.uk/term/24/TEMP>
 - * <http://vocab.ndg.nerc.ac.uk/term/24/PSAL>
 - * <http://vocab.ndg.nerc.ac.uk/term/24/CPWC>



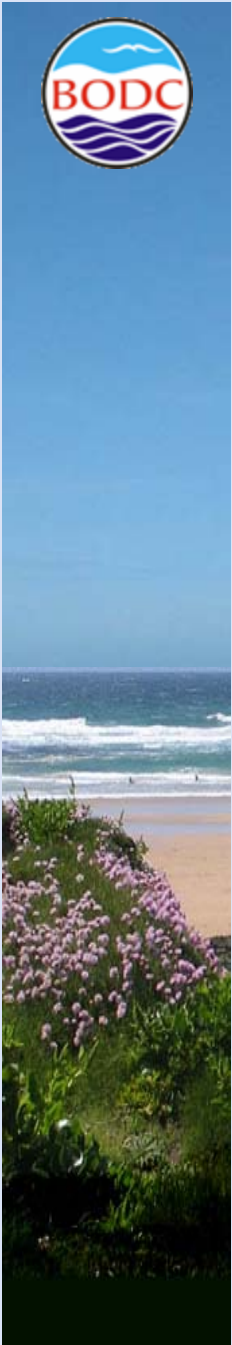


Semantic Crosswalk Use Case

This list may be rolled into an HTTP get request thus:

- `http://vocab.ndg.nerc.ac.uk/axis2/services/vocab/getRelatedRecordByTerm?subjectTerm=http://vocab.ndg.nerc.ac.uk/term/P021/current/TEMP&subjectTerm=http://vocab.ndg.nerc.ac.uk/term/P021/current/PSAL&subjectTerm=http://vocab.ndg.nerc.ac.uk/term/P021/current/CPWC&objectList=http://vocab.ndg.nerc.ac.uk/list/P041/current&predicate=255&inferences=true`

- **An XML document is returned containing the GCMD Science Keywords that map to the three BODC terms as both text strings and URLs**
- **The document may be reformatted using XSLT or XQuery to generate the “parameters” section for the DIF**





Conclusions

- **Fields populated from controlled vocabularies are replacing plaintext in SeaDataNet**
- **SeaDataNet vocabularies have formalised content and technical governance**
- **Semantic Web technology (URNs, URLs and RDF triples) is being used to integrate semantics into SeaDataNet data and metadata**





Conclusions

- **Thank you for your attention**
- **Any questions?**

