

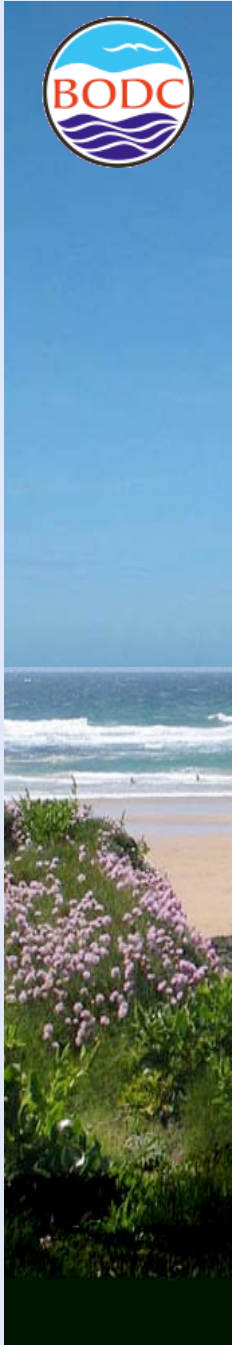


ICAN Workshop, Copenhagen, 07-08 July 2008

Semantic Interoperability and SeaDataNet

Roy Lowry

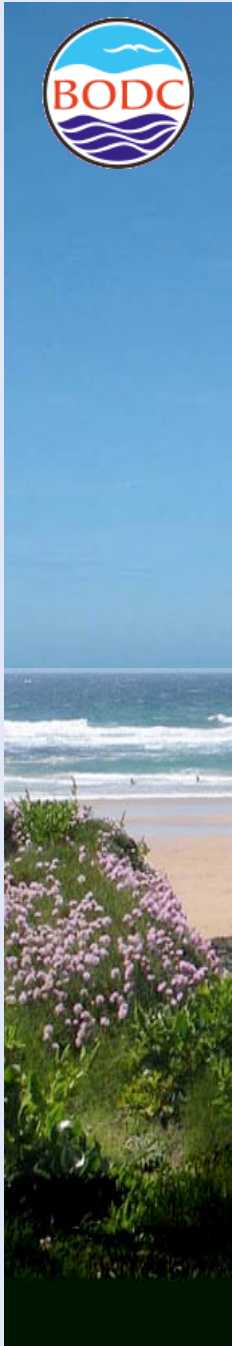
British Oceanographic Data Centre





SeaSearch: SeaDataNet precursor

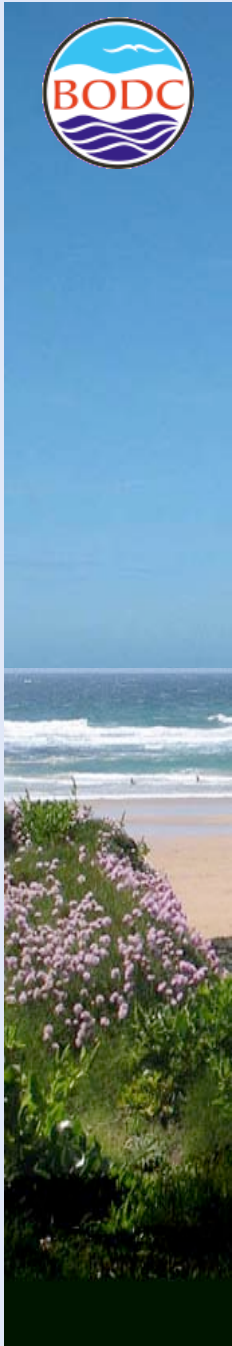
- **Plaintext used extensively to populate metadata**
 - Works for humans but not intended software agents
- **Some fields were populated from controlled vocabularies**
- **However, the vocabularies were:**
 - Poorly managed
 - Lacking in definitions
 - Abused to facilitate shoe-horning





SeaDataNet Semantic Markup

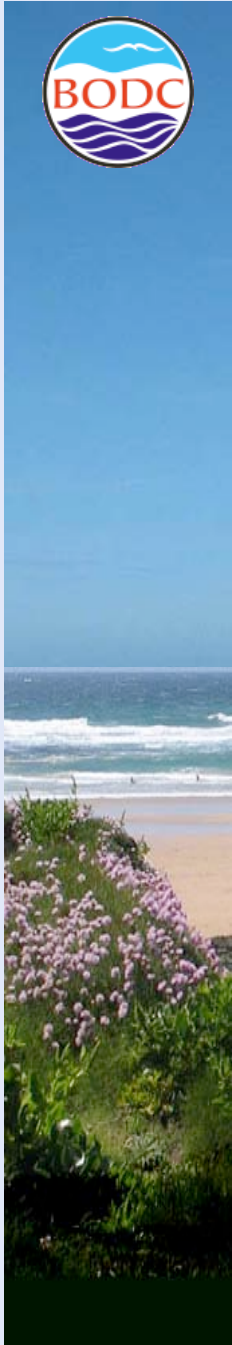
- **SeaDataNet now has governed vocabularies in a managed semantic technical infrastructure (NDG Vocabulary Server)**
- **SeaDataNet semantics in data and metadata are represented by permanent machine-readable labels (URNs)**
- **SeaDataNet metadata schemas incorporate Schematron facilitating CONTENT validation against master vocabularies.**





SDN Semantic Markup

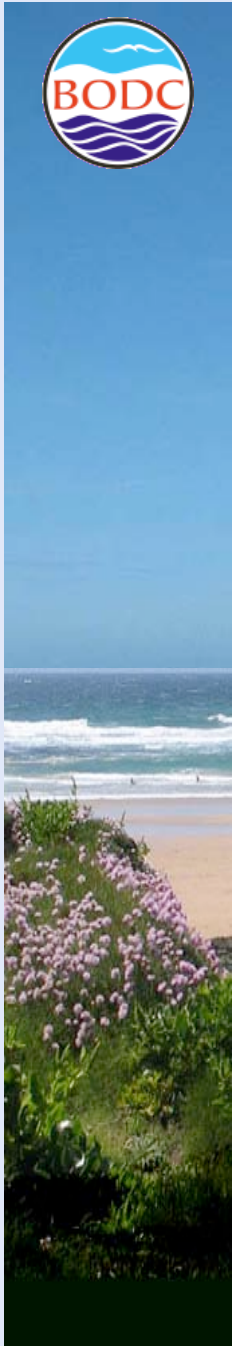
- **An example URN is SDN:P021:21:TEMP**
 - **Where:**
 - **SDN = namespace identifier**
 - **P021 = resource identifier**
 - **21 = resource version number**
 - **TEMP = resource component identifier**





From URN to URL

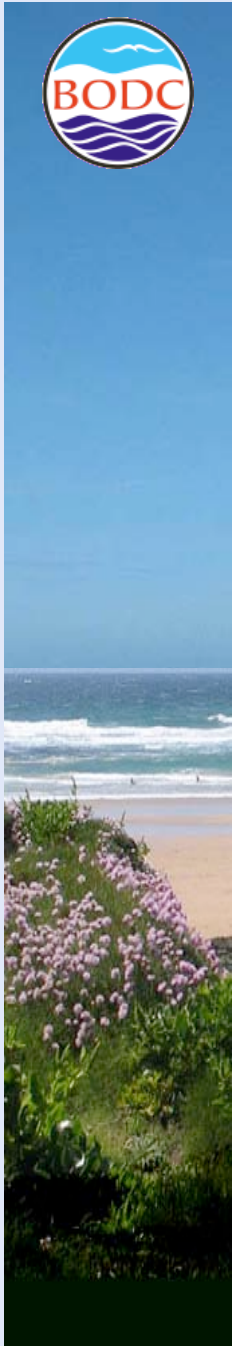
- **Some simple string substitution converts this URN into something that may be found on the internet (URL)**
 - **http://vocab.ndg.nerc.ac.uk/ = namespace**
 - **term/P021 = resource**
 - **/21 = version**
 - **/TEMP = resource component identifier**
- **This gives**
 - **<http://vocab.ndg.nerc.ac.uk/term/P021/21/TEMP>**
- **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

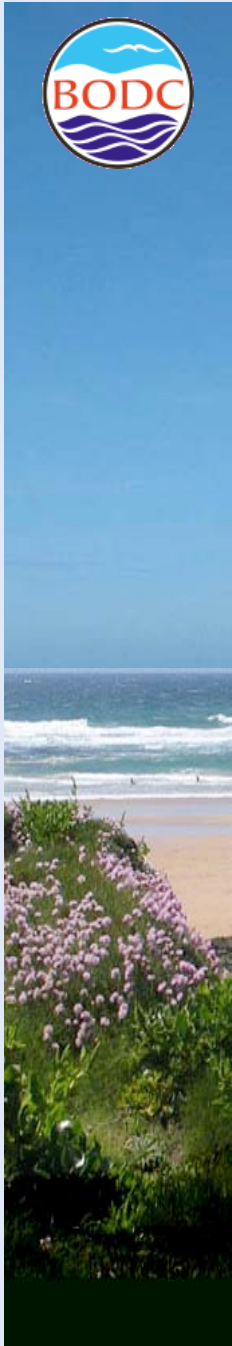
- This is a Semantic Web resource for the technical governance of controlled vocabularies
- Built by BODC as part of the NERC DataGrid project
- Adopted by, and being further developed for, SeaDataNet
- Fundamental entity is a concept that is represented by an identifier, a term, an abbreviation and 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>
```

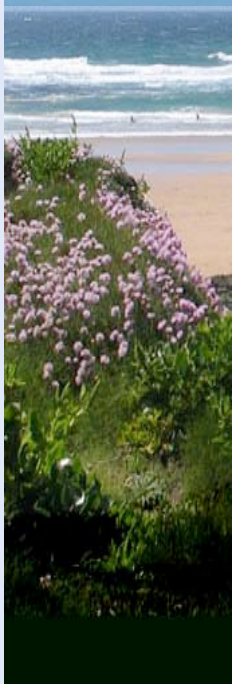
```
<dc:date>2008-02-26T14:25:51.822+0000</dc:date>
```

```
<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>
```

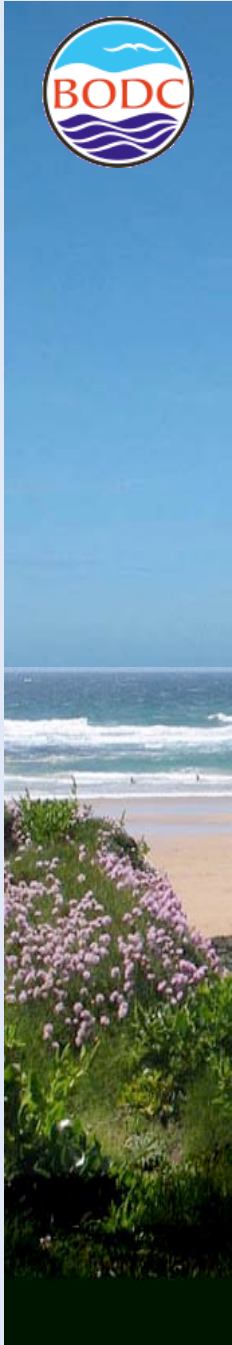
```
</rdf:RDF>
```





The NDG Vocabulary Server

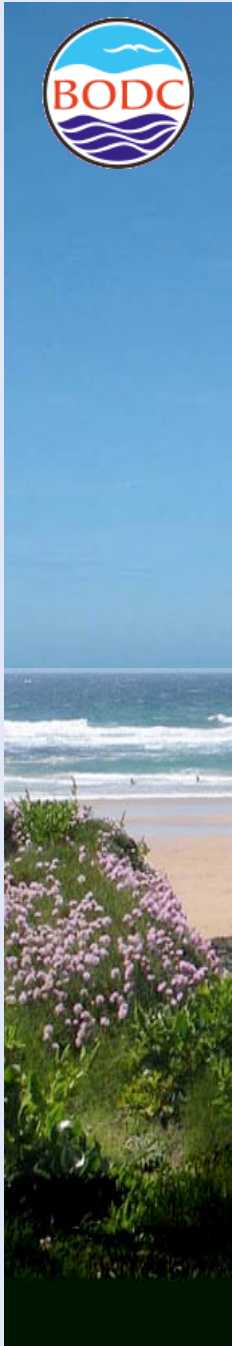
- **The server has other more sophisticated access mechanisms designed to facilitate bulk access**
 - **URL access to SKOS lists (no mappings)**
 - **HTTP-POX API (see http://www.bodc.ac.uk/products/web_services/vocab/methods.html)**
 - **SOAP API and method interface clients (see <http://vocab.ndg.nerc.ac.uk/>)**





The NDG Vocabulary Server

- **The server currently (2008-06-25) holds**
 - 112 public lists
 - 122166 concepts
 - 77347 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**

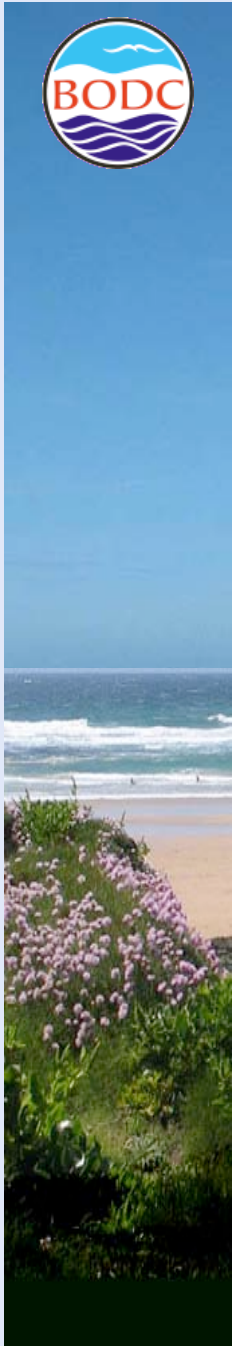




The NDG Vocabulary Server

➤ Limitations

- **Planned to be addressed by V1.2 (funded: scheduled end 2007)**
 - * RDF predicates restricted to SKOS
 - Lift restriction
 - Serve OWL for list and term URLs
 - * Implement versioned list serving (V1.1 always serves current list)
 - * Label lists with governance
- **Planned to be addressed by V2 (currently unfunded)**
 - * Upgrade to 'ontology server'
 - V1.2 OWL functionality
 - Allow mappings to concepts from other servers (currently must hold a list to map to it)

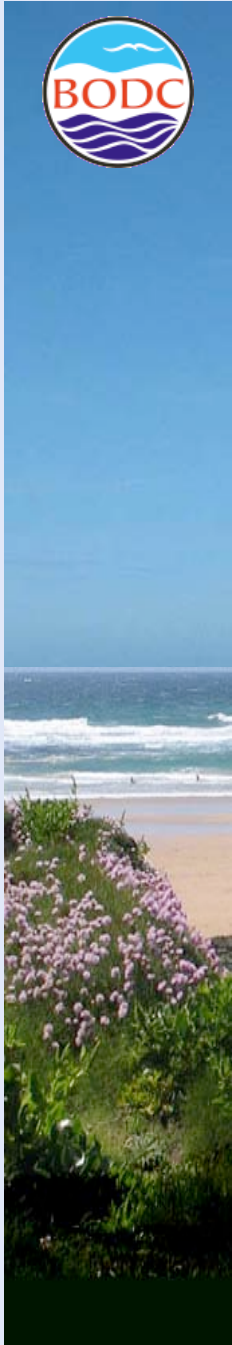




Relevance to ICAN

➤ Vocabulary Server

- **Operational server (Alternative to MMI)**
 - * My view (not necessarily shared by John Graybeal)
 - MMI strengths
 - » Metadata reference library
 - » Guides to best practice
 - » Solution developers
 - » Harnessing 'bleeding edge' technology
 - MMI weaknesses
 - » Operational aspects (e.g. synchronisation breakdowns)
 - » URLs to concepts deliver whole ontology
 - » Dependence on soft money
- **Rich semantic vein to be mined by ICAN**
 - * Developing an established list is much easier than starting from scratch!
 - * SeaDataNet content governance infrastructure





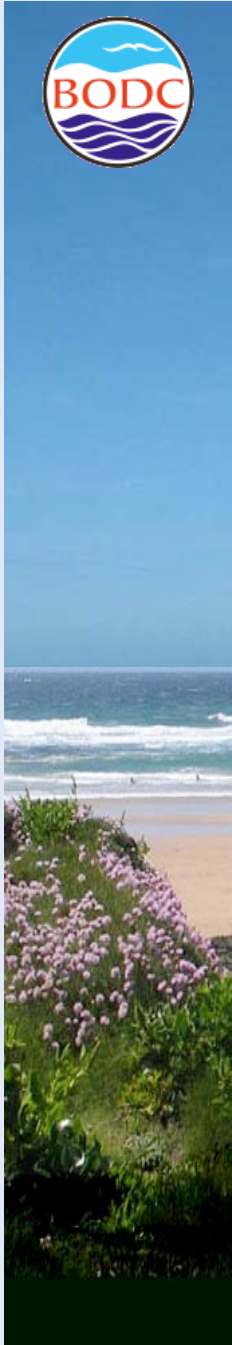
Relevance to ICAN

➤ Metadata validation technology

- Ensures only terms from master vocabulary/ontology are used in metadata documents
- Picks up typos and 'quick fix' local extensions (the bain of SeaSearch)
- Already arisen as an issue in ICAN (I recall an e-mail requesting only ontology terms be used in metadata documents)

➤ SeaDataNet Experience

- We're bound to make mistakes in SeaDataNet
- Watch and learn from our experience!





That's All Folks

- **Thank you for your attention**

- **Any questions?**

