

LINKED DATA: AN OCEANOGRAPHIC PERSPECTIVE

Adam Leadbetter
British Oceanographic Data Centre
alead@bodc.ac.uk



OVERVIEW

1. Where we came from
2. Where we are now
3. Where we are going to

1. WHERE WE CAME FROM

- 25 years of history!
- 1987 - GF3 vocabularies as published book
 - In English, French, Spanish & Russian
 - But books are expensive to reproduce
 - Hence lack of version and control...
 - ... leading to vocabulary abuse

1. WHERE WE CAME FROM

- 2000 - SeaSearch
 - Digital library of ten vocabularies
 - But content governance delegated to "volunteers" ...
 - ... no term definitions, no consensus
 - Weak technical governance...
 - ... local / master clashes

1. WHERE WE CAME FROM

- 2006 - SeaDataNet
 - Designed around semantic interoperability
 - Content governance groups established
 - Technical governance used NERC Vocab Server
 - 100+ vocabularies
 - Some technical issues

2. WHERE WE ARE NOW

- Version 2 of the NERC Vocabulary Server
 - <http://vocab.nerc.ac.uk/>
- Conforms to latest SKOS specification
- Has SPARQL endpoint
 - <http://vocab.nerc.ac.uk/sparql>
- 140,000 concepts
- 210,000 "internal" relationships
- 10,000 "external" relationships
- Published on datahub.io - submitted to LOD-Cloud

2. WHERE WE ARE NOW

- NVS2.0 content used by:
 - NERC Data Catalogue Service
 - MEDIN Data Discovery Portal
 - EU FP 7 SeaDataNet
 - EU FP 7 NETMAR
 - NSF Rolling Deck to Repository
 - NSF BCO-DMO



2. WHERE WE ARE NOW

- NVS2.0 content used by:
 - NERC Data Catalogue Service (3 / 4*)
 - MEDIN Data Discovery Portal (3 / 4*)
 - EU FP 7 SeaDataNet (3 / 4*)
 - EU FP 7 NETMAR (3 / 4*)
 - NSF Rolling Deck to Repository (5*)
 - NSF BCO-DMO (5*)

2. WHERE WE ARE NOW

- NVS2.0 content used by:
 - BODC's data publication landing pages
 - HTML with RDFa
 - Structured content accessible through SPARQL



British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Home Contact us Glossary Site map Site styles SHARE

My account Register Log in

About us Data Projects Partners Products Help and hints

Where to find data
Online delivery>
Online request>
Published Data Library>
Information and inventories>
Code and format definitions>
Submitting data to BODC
Portals and links>

Published Data Library (PDL)
Western Channel Observatory (L4) nutrient concentration profiles (2000 - 2012)

Dataset title	Nutrient concentration profiles from long term time series at Station L4 in the Western English Channel from 2000 to 2012
Dataset creators	Dr. E. Malcolm S. Woodward , Mrs. Carolyn Harris , Mrs. Lisa Al-Moosawi
Dataset subject	Environment
Dataset	Results are reported for the analysis of water samples taken from the L4 long-term sampling station (50°15.0'N; 4°53.0'W) situated off the south-west coast of England, United Kingdom. The analyses are reported for the period from 2000 to 2012.

<http://purl.org/dc/elements/1.1/creator> " Dr. E. Malcolm S. Woodward "
<http://purl.org/dc/elements/1.1/contributor> " Mrs. Carolyn Harris "
<http://purl.org/dc/elements/1.1/contributor> " Mrs. Lisa Al-Moosawi "
<http://purl.org/dc/elements/1.1/subject>
<http://www.isotc211.org/2005/resources/Codelist/gmxCodeLists.xml#MD_TopicCategoryCode_environment>



3. WHERE WE ARE GOING TO

- Proof of concept for NERC SIS
 - "Source-to-sea" contaminant "story"
 - Moves NERC discovery service to 5*
- Follow up on interest from ODP etc... in NVS2.0
- Linked Observation (Meta)Data
 - Linked SemSOS (?)
- Deploy a "data" namespace endpoint

ACKNOWLEDGMENTS

- Roy Lowry, British Oceanographic Data Centre
- Robert Arko, Lamont-Doherty Earth Observatory
- Cynthia Chandler and Adam Shepherd, Woods Hole Oceanographic Institution
- John Watkins and Dan Wright, Centre for Ecology and Hydrology, Lancaster

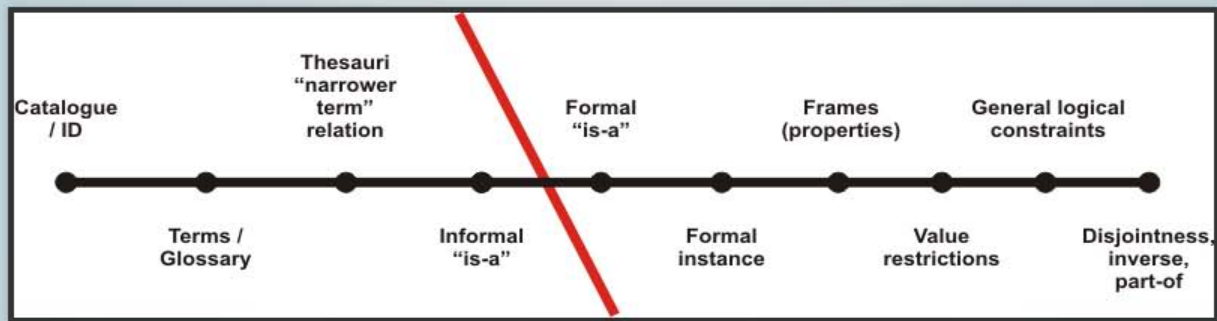


THANK YOU

Any questions?



ASIDE 1: THE SEMANTIC SPECTRUM



ASIDE 2: FEDERATED SPARQL

- Requires a SPARQL 1.1 endpoint for "orchestration"
 - SERVICE keyword is SPARQL 1.1 extension
 - We use [Apache Jena Fuseki](#)
- Tested with up to three endpoints
- Tightly constrained remote responses required...
... otherwise performance drops markedly

ASIDE 2: 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 ?rvl ?nvsp 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
}
```

Results [here](#). More queries [here](#).

ASIDE 3: LINKED DATA SCALE

- * - Data are online ...
- ** - ... in a machine readable format ...
- *** - ... which is not proprietary ...
- **** - ... using URIs, RDF, SPARQL ...
- ***** - ... and links to other people's URIs.