

AGU2017 vocabulary server and SPARQL end point demo

British Oceanographic Data Centre, National Oceanography Centre, UK

Contacts for help and questions:

- Presenter; Justin Buck juck@bodc.ac.uk
- Vocabulary server lead; Alexandra Kokkinaki alexk@bodc.ac.uk
- SPARQL endpoints lead; Chris Wood chwood@bodc.ac.uk

The handout includes key linkages and basic queries to introduce the BODC vocabulary server and hosted SPARQL endpoints. It starts with the vocabulary server then moves onto finding data (including a simple federated query) before finishing with the Celtic Seas Portal that is a proof of concept showing web portals built on SPARQL endpoints.

NERC Vocabulary server

- Search: https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/
- Builder: https://www.bodc.ac.uk/resources/vocabularies/vocabulary_builder/
- SOAP service described at <http://vocab.nerc.ac.uk/>
- SPARQL endpoint: <http://vocab.nerc.ac.uk/sparql/>

NERC Vocabulary server SPARQL endpoint

NVS is simple: just describes SKOS collections, concepts, and schemes:

```
select distinct ?b where { ?a a ?b . }
```

A list of vocabularies is then just:

```
select distinct ?a where { ?a a <http://www.w3.org/2004/02/skos/core#Collection> . }
```

And then, the preferred labels of each of the concepts of a given collection (e.g. P08, because it's easy to think about):

```
select * where { <http://vocab.nerc.ac.uk/collection/P08/current/> a  
<http://www.w3.org/2004/02/skos/core#Collection> ;  
<http://www.w3.org/2004/02/skos/core#member> ?c .  
?c <http://www.w3.org/2004/02/skos/core#prefLabel> ?d }
```

ConceptSchemes are for specific metadata catalogues held within NVS:

```
select distinct ?c where { ?c a <http://www.w3.org/2004/02/skos/core#ConceptScheme> . }
```

...and show how they link to vocabs (e.g. picking DCAT themes for EDMED):

```
select * where { <http://vocab.nerc.ac.uk/scheme/EDMED\_DCAT\_THEMES/current/> ?a ?b .  
}
```

shows that this uses five D01 terms

Good practical example: shows that *Discovery* is narrower than *Research Vessel*, and the name and alternative name of Discovery

- The subject of a triple is almost always a URI (so that it can be referred to uniquely)
- The object of a triple can be either be a URI or a literal property

Subject	Predicate	Object
http://vocab.nerc.ac.uk/collection/L06/current/31	skos:narrowMatch	http://vocab.nerc.ac.uk/collection/C17/current/74EQ
http://vocab.nerc.ac.uk/collection/C17/current/74EQ	skos:prefLabel	"Discovery"
http://vocab.nerc.ac.uk/collection/C17/current/74EQ	skos:altLabel	"DY"

Series schema SAPRQL endpoint (SeSSE)

- Series schema endpoint (SPARQL endpoint on top of BODC held datasets): <http://linked.bodc.ac.uk/>
- Link to basic documentation: <http://linked.bodc.ac.uk/documentation/>
- each data series is a <http://www.w3.org/ns/dcat#Dataset>, so we can count the number of datasets:

```
select count(?b) where {?b a <http://www.w3.org/ns/dcat#Dataset> . }
```

All relevant metadata described, e.g.

```
select * where {<http://linked.bodc.ac.uk/series/65462/> ?b ?c . }
```

and especially interesting is probably the distribution (direct link to download), both ODV...

```
select * where {<http://linked.bodc.ac.uk/series/65462#odv> ?b ?c . }
```

...and netCDF...

```
select * where {<http://linked.bodc.ac.uk/series/65462#netCDF> ?b ?c . }
```

Since we use NVS to describe the *themes* (i.e. P01 description) of datasets (using dcat:theme), then we can use federated queries to get the info about the vocabs used.

This shows that 65462 has 3 *themes* (P01 terms) associated with it

```
select * where {<http://linked.bodc.ac.uk/series/65462/>
<http://www.w3.org/ns/dcat#theme> ?c . }
```

This following query was written specifically for the Ellett line repeat hydrographic section in the North Atlantic to identify reference data within BODC holdings near to an ocean glider deployment, it uses a geospatial range to search available datasets:

```
select * where {
?a <http://www.w3.org/2003/01/geo/wgs84\_pos#lat> ?lat .
?a <http://www.w3.org/2003/01/geo/wgs84\_pos#lon> ?lon .
?a <http://www.w3.org/ns/dcat#distribution> ?dist .
?dist <http://www.w3.org/ns/dcat#downloadURL> ?url .
filter(?lon > -21 && ?lon < -5 && ?lat > 56 && ?lat < 64)
}
limit 10
```

Celtic Seas

- This uses a different software stack to NVS & SeSSE (D2RQ instead of Jena) – D2RQ allows dynamic triplestores for data stored in a database rather than needing a separate triplestore.
 - Advantages: quicker proof of concept development; no duplication of data
 - Disadvantages: slower / less robust than dedicated triplestore
- We have 2 triplestores:
 - the metadata catalogue: <http://linked.bodc.ac.uk/celticseaspartnership/metadata/>
 - the library catalogue (for relevant docs): <http://linked.bodc.ac.uk/celticseaspartnership/library/>
- We have used this as a proof of concept that a whole portal can be built on top of triplestores (i.e. triplestores & SPARQL as an API): <http://resources.celticseaspartnership.eu/>
- again, datasets described with dcat:Dataset:

```
SELECT * WHERE {
  CSPDatasets:4 ?o ?p
}
```

http://linked.bodc.ac.uk/celticseaspartnership/metadata/snorql/?query=SELECT+*+WHERE+%7B%0D%0A++CSPDatasets%3A4+%3Fo+%3Fp%0D%0A%7D%0D%0A

- C45, C46, and C47 (MSFD Descriptor, Criteria, and Indicator) related to a particular dataset can be easily found:

```
SELECT distinct ?w WHERE {
  CSPDatasets:4 dcat:theme ?p. ?p dc:description ?w
}
```

(Note: we publish the vocabs via D2RQ, when technically we don't need to because they're part of NVS triplestore anyway...)