

Demonstration of adding content to an ICAN Semantic Resource

Roy Lowry, Adam Leadbetter, Olly Clements
(NETMAR - BODC)

Tanya Haddad (ICAN - OCA)



What is a Semantic Resource?

- A semantic resource for federated smart discovery requires the following components
 - Markup terms for component datasets (local vocabularies)
 - Terms that populate a global semantic discovery hierarchy (global thesaurus)
 - Semantic mappings between the local vocabularies and the global thesaurus



What is a Semantic Resource?

- If a semantic resource is to be of use to a computer system, it needs to be encoded in a so that semantically aware software can understand it, but most humans can't
- These encodings are called Knowledge Organisation Systems
- This presentation describes how go from human understandable information to an encoding in Simple Knowledge Organisation System using the NETMAR-developed V2 of the NERC Vocabulary Server



Simple Knowledge Organisation System

- Simple Knowledge Organisation System (SKOS)
 - Lightweight W3C standard for knowledge organisation
 - Fundamental SKOS element is the ‘concept’
 - Simply a term in a local vocabulary or discovery keyword hierarchy
 - Concepts may be organised into collections
 - Groups of concepts that have something in common – such as being valid labels for Oregon Coastal Atlas datasets
 - Concepts may also be organised into schemes
 - Semantically related groups of concepts



The ICAN Semantic Resource

- For the first prototype ICAN prepared a Global Ontology comprising 12 OWL classes
- Connecting a new local vocabulary to this required the following steps
 - Conversion of the local vocabulary to OWL using tools like Protégé or TopBraid Composer
 - Upload into the MMI Ontology Registry and Repository
 - Mapping ontology classes using the MMI VINE tool



The ICAN Semantic Resource

- Not many people have connected to the semantic resource
- One possible reason is that OWL and its associated tooling is heavy engineering for what is a thesaurus rather than an ontology
- So, I thought I'd try a simpler approach....



ICAN Semantic Resource

- The Global Ontology comprises
 - Three top concepts
 - Each of these maps the three narrower concepts leaf nodes)
- How can we link this to something like the Oregon Coastal Atlas (OCA) Erosion Ontology?



ICAN Semantic Resource

- The OCA Erosion Ontology contains both dataset markup terms like ‘Bathymetric grid’ overlain by a hierarchy of broader discovery terms like ‘Goal19’
- First job was to tease out the markup terms
- Assumed (oversimplification!) that these were the leaf nodes in the ontology



ICAN Semantic Resource

- OCA coastal erosion ontology opened in MMI ORR followed by copy/paste into Excel
- Result split by cut/paste into two [Excel worksheets](#)
 - Markup terms (leaf nodes)
 - Discovery terms (everything else)
- Loaded as two concept collections (A02, A03) into the NERC Vocabulary Server back office database
- Mappings between them were then uploaded from the subclass relationships in the MMI document



ICAN Semantic Resource

- The OCA coastal ontology markup terms were mapped one by one to the 9 leaf nodes of the ICAN Global Coastal Erosion Ontology (Excel copy/paste)
- ICAN Global Ontology (markup + discovery) was uploaded as a concept collection (A01) into NVS
- Mappings between ICAN and OCA were uploaded into NVS



ICAN Semantic Resource

- Getting to this stage only took 1.5 days, most of which was the derivation of the 164 mappings
- The NVS is now able to serve the following quite useful SKOS documents
 - [ICAN Coastal Erosion Global Thesaurus](#) as a concept collection can power a smart discovery interface
 - [OCA Coastal Erosion Thesaurus markup terms](#) can power a metadata or data labelling tool at OCA



ICAN Semantic Resource

- However, we could do with a little more
 - The A01 collection has no entry points (SKOS top concepts)
 - It also cannot power an ontology browser without multiple server calls (sedated slug syndrome)
- So, I converted A01 to the [ICANDIS concept scheme](#)
- I also converted the entire Global Ontology plus OCA ontology into the [ICANCOERO concept scheme](#)
- The conversion required my typing in 5 lines of SQL, which took about 10 minutes.



ICAN Semantic Resource

- Some may regard this build as the work of Mr W Heath Robinson
- However, it works and could form the basis for ICAN extension over the next 12 months
- An Excel template would be provided by BODC to be populated by the ICAN content provider then loaded into the NVS by BODC



ICAN Semantic Resource

- I like to work like this, but some others will not (I've met them)
- So, we need alternatives
 - BODC already has a Web Form vocabulary editor that allows secure external update of concept collections but not mappings or schemes
 - Plan is to provide this with mapping support as part of NETMAR



ICAN Semantic Resource

- There are also those who prefer other resources than the NVS such as MMI ORR or tother vocabulary servers
- Providing these resources address concepts using URLs, they can be linked to concepts inside the NVS and then be served by the NVS
- However, this comes at a price



ICAN Semantic Resource

- The client will receive a SKOS document from the NVS with a concept element like:

```
</skos:Concept>
<skos:Concept rdf:about="http://vocab.nerc.ac.uk/collection/A01/current/Habitat_Alteration">
<dc:identifier>SDN:A01:1:Habitat_Alteration</dc:identifier>
<skos:prefLabel>Habitat Alteration</skos:prefLabel>
<skos:altLabel/>
<skos:prefLabel>Processes, activities and events that change the nature of the environment provided for living organisms.</skos:prefLabel>
<skos:comment>accepted</skos:comment>
<dc:date>2011-08-15 11:25:11.0</dc:date>
-- Mapping to concept in MMI ORR
<skos:exactMatch rdf:resource="http://mmisw.org/orr/#http://mmisw.org/ont/ican/thesaurus/Habitat_Alteration"/>
-- Mappings within the NERC Vocabulary Server
<skos:narrowMatch rdf:resource="http://vocab.nerc.ac.uk/collection/A02/current/OilSpill"/>
<skos:narrowMatch rdf:resource="http://vocab.nerc.ac.uk/collection/A02/current/DuneRestoration"/>
<skos:narrowMatch rdf:resource="http://vocab.nerc.ac.uk/collection/A02/current/DuneGrading"/>
-- Mapping to the British Geological Survey Vocabulary Server
<skos:narrowMatch rdf:resource="http://webservices.bgs.ac.uk/data/services/vocabulary/1.0/vocabularies/DIC_BUILDING_DAMAGE/terms/3"/>
<skos:broadMatch rdf:resource="http://vocab.nerc.ac.uk/collection/A01/current/Effects_of_Coastal_Change"/>
<skos:broadTransitive rdf:resource="http://vocab.nerc.ac.uk/collection/A01/current/Effects_of_Coastal_Change"/>
</skos:Concept>
```



ICAN Semantic Resource

- The BGS URL returns XML but different XML to the NVS
- The MMI URL opens up the MMI ORR interface
- Consequently, the client needs to be able to parse every XML dialect served up by the components of the extended ontology
- The situation cries out for standardisation
- W3C set up an incubator group to look at this, but no reports of progress to date

