Vocabulary and Semantics in the Virtual Observatory

Norman Gray
VO-TECH / AstroGrid / Uni. Leicester / Uni. Glasgow, UK

VOEvent BoF, ADASS, London, 2007 September 24

rdf

- Resource Description Framework (W3C Rec in 1999)
- . . . *not* Format
- You can write it down (verbosely) but that's not really the point
- It helps separate syntax from semantics.
- (a bit like assembler for the web)
- Oooooold technology

rdf intro

- All the world is *triples*, consisting of *resources* named by URIs (`http:...` or `ivo:...` or `urn:example#Norman`)
- . . . which have *properties* whose *values* are resources or literals.
- RDF/RDFS/OWL describe these using `rdf:type`, `rdfs:subClassOf`, `owl:symmetricProperty`, and so on.

There is an analogy with XML Schemas, *but it is a loose one* -- they're not addressing the same problem. Same for O-O.

rdf/owl/semweb/sql/xml — respective strengths

RDF/OWL/reasoning now largely stable (though The Semantic Web will forever be Vision). Now engineering rather than CS.

Using the architectural principles which let HTML take over the internet. Very open and flexible; has existing powerful query language. Did I mention standards?

RDB to XML to RDF — spectrum of strengths. XML is more natural than RDF where the information density is high, and the information regular or highly constrained; RDF/SW is natural for incomplete or ragged data.

rdf schemas give you reasoning

```
<http://x> a ns:SecondaryEducationContentLevel.
ns:SecondaryEducationContentLevel
    rdfs:subClassOf ns:SchoolContentLevel.
```

Thus `http://x` is School Content Level, too.

Or. . .

```
<http://x> ns:emailAddress <mailto:foo@example.org>.
<http://y> ns:emailAddress <mailto:foo@example.org>.
ns:emailAddress a owl:InverseFunctionalProperty.
```

implies `http://x` and `http://y` are the same entity.

semantic applications

- *Semantic search* : ‘doing better than just string matching’. Thus search for ‘plough’ and get astronomy, not agriculture; search for ‘compact object’ and get black holes.

- *Browsing* : ‘you asked for compact objects, but is it black holes or quasars you're interested in?’

- *Classification* : ‘this object has properties A, B and C, so it must be a Wolf-Rayet star’

These need different reasoning.

spectrum/ladder

- *ontologies* : logic (RDFS to OWL)
- *thesaurus* : synonyms, antonyms, see-also
- *taxonomy* : broader/narrower, subclass/superclass
- *vocabulary* : controlled list of words
- *folksonomy* : free keywords/tags with counts

ontology

A formal specification of a shared conceptualisation.

Eh?

- conceptualisation = a set of things/concepts/types
- shared = . . . which at least one other person agrees with
- specification = . . . and which you've written down
- formal = . . . in a machine-readable way.

Not in front of the users!

‘Simple Knowledge Organisation System’

- W3C standardisation: SKOS Core is close;
- . . . but SKOS Mappings is early (but trimming)
- Includes relations ‘broader’, ‘narrower’, ‘related term’, and other apparatus.
- Comes from long experience in the library community.

costs

- Processing costs : heavy reasoning is expensive (but not always necessary, and you might be able to do it off-line)
- Acquisition costs : the people who pay have to get the benefit (but if they get benefit they'll pay willingly)
- Development costs : scary (so don't do it)
- Opportunity costs : do your own thing (and you're on your own)

for example…

```
<ivo://VOcabulary/AOIM#startypevariablenova>
    a skos:Concept;
    skos:prefLabel "Star: Type: Variable: Nova";
    skos:broader <ivo://VOcabulary/AOIM#star>;
    rdfs:subClassOf <http://www.ivoa.net/ns/ucd#star.binary.CV.nova> .
<ivo://ivoa/VOcabulary/AAkeys#starsnovaecataclysmicvariables>
    a skos:Concept;
    skos:prefLabel "(Stars:) novae, cataclysmic variables";
    skos:broader <ivo://ivoa/VOcabulary/AAkeys#stars>;
    rdfs:subClassOf <http://www.ivoa.net/ns/ucd#star.binary.CV.nova> .
```

a sparql query

```
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix AOIM: <ivo://VOcabulary/AOIM#>

select ?r ?b
where {
  AOIM:startypevariablenova rdfs:subClassOf ?r.
  ?r skos:broader ?b.
}
```

a sparql result!

```
<results>
 <result>
  <binding name="r">
   <uri>ivo://VOcabulary/AOIM#startypevariablenova</uri>
  </binding>
  <binding name="b"><uri>ivo://VOcabulary/AOIM#star</uri></binding>
 </result>
 <result>
  <binding name="r">
  <uri>ivo://ivoa/VOcabulary/AAkeys#starsnovaecataclysmicvariables</uri>
  </binding>
 <binding name="b"><uri>ivo://ivoa/VOcabulary/AAkeys#stars</uri></binding>
 </result>
</results>
```

What this has shown is:

- You can go from a term in one (deployed) vocabulary to a broader term in another, without a new consensus vocabulary.
- The linkage was indirect (via UCDs in this case), but could be direct, or use another intermediate.
- Existing vocabularies aren't very structured, because structure was hard to use (separately from being hard to acquire).
- Browsing support makes it feasible to get the extra payoff of structure.