OWL: Path to Massive Deployment

Dean Allemang
Chief Scientist, TopQuadrant Inc.
dallemang@topquadrant.com
Web-Scale Deployment

- Number of pages

- Amount of Data

- Awareness
  - “I’m a Web Developer” “Have you heard of X?”
  - For X = Java, Ruby, HTML5, Semantic Web, RDF, SPARQL, Jena, OWL, Linked data, ...
Facebook Open Graph Protocol
Facebook and OGP

Example: OGP use growing quickly
Facebook incentivizing use of RDFa like buttons

Tetherless World Constellation

15,178 sites of top 1,000,000 as of 3/3/11

Oct 2010: FB reports RDFa is ~ 10-15% of > 3,000,000 likes per day!

Facebook is encouraging developers to use the RDFaversion

Source: Jim Hendler
OGP and Simplicity

<meta property="og:latitude" content="37.416343"/>
<meta property="og:longitude" content="-122.153013"/>
<meta property="og:title" content="The Rock"/>
<meta property="og:type" content="movie"/>
<meta property="og:url" content="http://www.imdb.com/title/tt0117500/"/>
<meta property="og:image" content="http://ia.media-imdb.com/rock.jpg"/>
<meta property="og:site_name" content="IMDb"/>
<meta property="fb:admins" content="USER_ID"/>
<meta property="og:description" content="A group of U.S. Marines, under command of a renegade general, take over Alcatraz and threaten San Francisco Bay with biological weapons."/>
KISS
keep it simple, stupid

KISS
simplify.
h.d. thoreau

KISS
keep it simple...

KEEP IT SIMPLE STUPID
Good Relations

- Vocabulary for eCommerce
- Lets vendors describe their products, services, prices, etc.
- Makes this data mergeable, dynamic, queryable, etc.

Source: Martin Hepp at http://purl.org/goodrelations
### OWL Success Stories: Good Relations

<table>
<thead>
<tr>
<th>site</th>
<th>pages</th>
<th>triples</th>
</tr>
</thead>
<tbody>
<tr>
<td>overstock.com</td>
<td>1,000,000</td>
<td>100,000,000</td>
</tr>
<tr>
<td>CSNstores*</td>
<td>2,000,000</td>
<td>20,000,000</td>
</tr>
<tr>
<td>BestBuy*</td>
<td>500,000</td>
<td>5,000,000</td>
</tr>
<tr>
<td>Amazon*</td>
<td>20,000,000</td>
<td>4,000,000,000</td>
</tr>
<tr>
<td>Eurobau*</td>
<td></td>
<td>60,000,000</td>
</tr>
<tr>
<td>O’Reilly*</td>
<td>25,000</td>
<td>2,000,000</td>
</tr>
<tr>
<td>Bitmunk*</td>
<td></td>
<td>11,000,000</td>
</tr>
</tbody>
</table>

*Source: Good Relations wiki*
SKOS (Simple Knowledge Organization System)

- System for managing controlled vocabularies
- 28 Vocabularies on the W3C page, including:
  - Dewey decimal system
  - Library of Congress
  - United Nations Agrovoc
  - Many more not on this page . . .

- Each vocabulary is referenced from sometimes thousands of sites
Open Biological and Biomedical Ontologies (OBO)

- Over a hundred curated models
- Some are quite large
  - CHEBI > 1 million triples
  - Gene Ontology ~ 2.5 million triples
- Over 100 million triples – that’s getting to scale!
## OWL Utilization in common resources

<table>
<thead>
<tr>
<th></th>
<th>Good Relations</th>
<th>SKOS</th>
<th>OBO (CHEBI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AnnotationProperty</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Class</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>DatatypeProperty</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeprecatedProperty</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FunctionalProperty</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>ObjectProperty</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Ontology</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restriction</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>SymmetricProperty</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TransitiveProperty</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>disjointWith</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>inverseOf</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>onProperty</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>someValuesFrom</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>unionOf</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>versionInfo</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Optional Axioms in Good Relations – in SPARQL

Expand gr:includes

The gr:includes property is a shortcut for the many cases in which the product includes just one item. By that you spare the effort for an additional gr:TypeAndQuantityNode or even the gr:ProductOrServiceSomeInstancesPlaceholder node. However, when querying, you should search for the full pattern, or both. In order to expand all usages of gr:includes to the full pattern, activate the following TWO (!) SPARQL CONSTRUCT rules.

Note: In the 2010-09-16 service update of GoodRelations, the usage of gr:includes is being expanded to links between gr:Offering and gr:ProductOrServiceModels, which can drastically reduce the markup effort.

SPARQL CONSTRUCT Rule #1:

```sparql
PREFIX gr: <http://purl.org/goodrelations/v1#>
CONSTRUCT { 
  ?ggr:includesObject _in .
  _in rdf:type gr:TypeAndQuantityNode,
  _in gr:quantity gr:uns00ThisGood "1.0"^^xsd:float,
  _in gr:hasUnitOfMeasurement "C42"^^xsd:string,
  _in gr:type02Good ?p.}
WHERE 
  ?g rdf:type gr:Offering.
```
Namespaces used in Overstock.com product pages

- The actual overstock.com product pages include commerce data about their products. This data uses:
  - OGP (image etc.)
  - GR (Good Relations)
  - FOAF (depiction)
  - RDFS (comment and label)
  - RDF review vocabulary (http://vocab.org/review/terms.html)
  - RDF data vocabulary (http://rdf.data-vocabulary.org/)
  - Facebook markup (http://www.facebook.com/2008/fbml#)

- OWL is missing – and that’s a good thing!
Using Good Relations doesn’t require learning OWL
How about OBO?

How do I say “glyphosate is made of phosphorus” in CHEBI?

“glyphosate is a subclass of the set of things that have some part that is phosphorus”
How to Make $$$ in OWL Today

- Make OWL look really hard ("Drinking from a firehose" "Very Complicated")
  - Sell a training course to help people out
  - Write a book about OWL
  - Sell OWL consulting ("don’t try this at home!")
How to Make $$$ in OWL Tomorrow

OWL must not be mysterious
Who is interested in the Semantic Web?

- Database architect
- Java programmer
- Library scientist
- Data modeler
- IT Specialist
- Enterprise Architect
- Knowledge Manager
- Scientist (geographer, chemist, geologist, linguist)
- Entrepreneur
- Vocabulary Specialist
- Product manager
- Systems Engineer
- Consultant
- Software Engineer
- Taxonomist
What is hard about OWL?

- **Open World Assumption**
  - Data applications have all been closed world
  - Usually I want to use that data again anyway

- **Domain and Range**

- **Inferencing vs. Processing**
  - Declarative vs. procedural
  - Programmers want control
  - Data folks want queries

- **Logic**
  - allValuesFrom can mean “none”
“all” can mean “none”
Lotico Topics of Interest

From Lotico meeting titles and abstracts

© Copyright 2007-2010 TopQuadrant Inc.
You may be surprised by what other people don’t know – and whether they are motivated to learn it!

One of the most successful OWL deployments is one where end deployers don’t need to know OWL – It can be done!
Read more

- Semantic Web for the Working Ontologist by Dean Allemang and Jim Hendler

- Second Edition available now!

- Features
  - Good Relations
  - QUdT
  - Lots more!