Semantic Web: Core Concepts and Mechanisms

MMI ORR – Ontology Registry and Repository
What’s all this about?
• It’s all about formally capturing knowledge about the world
• so computers can be more useful
• so we can tackle pressing problems more effectively and efficiently
Capturing knowledge

- Knowledge expressed as statements
- Statements modeled as *triples* of the form:

```
Subject   Predicate   Object
```
Some knowledge

Hobbes

Calvin

friends

likes

classmates

Susie

Miss Wormwood

has teacher
Capturing semantics with triples

Calvin has friend Hobbes

Hobbes likes Susie

Calvin has classmate Susie

Calvin has teacher Wormwood
RDF: Resource Description Framework

- W3C standard to express information about resources
- Anything can be a resource, including physical things, documents, abstract concepts, numbers and strings
- The triple components denote resources
Resources

- Resources are denoted by IRIs and literals
- IRI = Internationalized Resource Identifier
- To identify resources, and to link to them
- Literals denote values according to known datatypes (numbers, strings, dates, ..)
IRIs or URIs?

• URIs used in RDF 1.0

• IRIs now used in RDF 1.1
  IRI: Generalization of URI allowing non-ASCII characters to be used in the IRI character string

• Every URI is an IRI

• URIs still prevalent, with mapping needed from IRIs to URIs when retrieval over the HTTP protocol
Rules for inference

- Example

If

X has classmate Y

and

X has teacher T

then

Y has teacher T
Inference

• So, given these facts:

- Calvin has classmate Susie
- Calvin has teacher Wormwood

• one can infer the following:

- Susie has teacher Wormwood
Graph-based data model

- Calvin
  - friend: Hobbes
  - teacher: Wormwood

- Hobbes
  - likes: Susie

- Susie
  - teacher: Wormwood

- Wormwood
  - classmate: Calvin
Reification
Capturing RDF triple data

<table>
<thead>
<tr>
<th>subject</th>
<th>predicate</th>
<th>object</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>p</td>
<td>j</td>
</tr>
<tr>
<td>a</td>
<td>q</td>
<td>k</td>
</tr>
<tr>
<td>b</td>
<td>r</td>
<td>m</td>
</tr>
<tr>
<td>c</td>
<td>p</td>
<td>j</td>
</tr>
<tr>
<td>c</td>
<td>p</td>
<td>w</td>
</tr>
<tr>
<td>d</td>
<td>t</td>
<td>a</td>
</tr>
</tbody>
</table>

\[(a, p, j) \]
\[(a, q, k) \]
\[(b, r, m) \]
\[(c, p, j) \]
\[(c, p, w) \]
\[(d, t, a) \]
Capturing RDF triple data

<table>
<thead>
<tr>
<th>predicates</th>
<th>p1</th>
<th>p2</th>
<th>p3</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>y</td>
<td>D</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>w</td>
<td></td>
<td></td>
<td>K</td>
</tr>
</tbody>
</table>

(x, p1, A)
(x, p2, B)
(x, p3, C)
(y, p1, D)
(y, p2, E)
(w, p3, K)
Capturing RDF triple data

<table>
<thead>
<tr>
<th>friend</th>
<th>likes</th>
<th>classmate</th>
<th>teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calvin</td>
<td>Hobbes</td>
<td>Susie</td>
<td>Wormwood</td>
</tr>
<tr>
<td>Hobbes</td>
<td>Calvin</td>
<td>Susie</td>
<td></td>
</tr>
<tr>
<td>Susie</td>
<td></td>
<td>Calvin</td>
<td>Wormwood</td>
</tr>
</tbody>
</table>
Capturing RDF triple data

- Ontology Editors
  - Protégé / WebProtégé (Stanford)
  - TopBraid Composer (TopQuadrant)
- Libraries
  - Apache Jena; OWL API; RDFLib;
Vocabularies

- Referring to particular subjects, properties and objects in triples means we are dealing with vocabularies.

- That is, naming things and using names introduced by others.

- “This ‘SST’ dataset was produced by organization ‘Acme’”
What about ontologies?

- Vocabularies are ontologies

- A way to think of a possible (loose) differentiation:
  - Tend to use “ontology” when the resources in your triples and the relationships among those resources are increasingly more elaborate in terms of intended semantics

- Let’s use “vocabulary” and “ontology” interchangeably here
Vocabularies

• Should be controlled vocabularies:

  • with names (and associated definitions/attributes) agreed by the community
  • to reduce discrepancies
  • to facilitate data discovery, reuse, and integration
  • to enable crosswalks/mappings
  • is short, to promote and facilitate interoperability
Isn't it weird how scientists can imagine all the matter of the universe exploding out of a dot smaller than the head of a pin, but they can't come up with a more evocative name for it than "The Big Bang"?
Isn't it weird how scientists...

What would you call the creation of the universe?

"The horrendous space kabloolie!"
"Verbing weirds language"

I take nouns and adjectives and use them as verbs. Remember when "access" was a thing? Now it's something you do. It got verbed.

Verbing weirds language. Maybe we can eventually make language a complete impediment to understanding.
Controlled vocabulary example:

**CF Standard names**

- [http://cfconventions.org/standard-names.html](http://cfconventions.org/standard-names.html)
- Precise description of 2,700+ physical quantities
  - name
  - description
  - canonical units
Vocabularies to use in your vocabularies

- **RDF**: *(Resource Description Framework)*
  - type, Property, Statement, …
  - subject, predicate, object, …

- **RDFS**: *(RDF Schema)*
  - Resource, Class, subClassOf, subPropertyOf, …
  - comment, label, seeAlso, isDefinedBy, …
Vocabularies to use in your vocabularies

- **SKOS**: *(Simple Knowledge Organization System)*
  - definition, note, ...
  - exactMatch, closeMatch, relatedMatch, ...

- **OWL**: *(Web Ontology Language)*
  - Ontology, inverseOf, ReflexiveProperty, ...
  - sameAs, versionInfo, ...
Vocabularies to use in your vocabularies

- **DCT**: *(Dublin Core Terms)*
  - title, description, creator, contributor…
  - rights, license, …

- **OMV**: *(Ontology Metadata Vocabulary)*
  - name, description, hasCreator, keywords,…
  - sameAs, …
Does semantic interoperability need an overarching vocabulary?

• No! … and such a goal is overly unrealistic in general

• But it’s fine to

  • Define what makes sense to your case

  • Map your names to names is other vocabularies as convenient/needed for interoperability

  • Propose additions to common vocabularies
Vocabularies: Summary

• Use standard vocabularies
  • in your data/metadata
  • in your own vocabularies, too!

• Participate in community vocabulary development activities
All of the above in practice:

ORR – Ontology Registry and Repository
ORR Origins

• *MMI* – Marine Metadata Interoperability project
  • [https://marinemetadata.org/](https://marinemetadata.org/)
  
  "Promoting the exchange, integration and use of marine data through enhanced data publishing, discovery, documentation and accessibility."

• ORR born as part of MMI’s vision for a *Semantic Framework*
ORR Origins

• Born as part of MMI’s vision for a Semantic Framework
## MMI ORR (v.2)

### Marine Metadata Interoperability

**Ontology Registry and Repository**

<table>
<thead>
<tr>
<th>URI</th>
<th>Name</th>
<th>Author</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><a href="http://www.w3.org/ns/ssn">http://www.w3.org/ns/ssn</a></td>
<td>Semantic Sensor Network Ontology</td>
<td>W3C</td>
</tr>
<tr>
<td>2</td>
<td><a href="http://vocab.nerc.ac.uk/standard_name">http://vocab.nerc.ac.uk/standard_name</a></td>
<td>BODC Climate and Forecast Standard Names (v.34)</td>
<td>BODC</td>
</tr>
<tr>
<td>3</td>
<td><a href="http://mmsie.org/ont/mmi/cformap">http://mmsie.org/ont/mmi/cformap</a></td>
<td>ORR-NVS CF standard name mapping (v.34)</td>
<td>MMI</td>
</tr>
<tr>
<td>4</td>
<td><a href="http://mmsie.org/ont/cf/parameter">http://mmsie.org/ont/cf/parameter</a></td>
<td>Climate and Forecast (CF) Standard Names (v.34)</td>
<td>CF Metadata</td>
</tr>
<tr>
<td>5</td>
<td><a href="http://mmsie.org/ont/icos/marine_bibiogeography">http://mmsie.org/ont/icos/marine_bibiogeography</a></td>
<td>US IOOS and OBIS-USA Marine Biogeography Terms v2.1</td>
<td>Philip Goldstein</td>
</tr>
<tr>
<td>6</td>
<td><a href="http://mmsie.org/ont/wwrMSS/wwrMSS">http://mmsie.org/ont/wwrMSS/wwrMSS</a></td>
<td>World Register of Marine Species</td>
<td>WWRMS Editorial Board</td>
</tr>
<tr>
<td>7</td>
<td><a href="http://mmsie.org/ont/mvcc/MVCCtoCFswell">http://mmsie.org/ont/mvcc/MVCCtoCFswell</a></td>
<td>MVCtoCFswell</td>
<td>Janet Fredericks</td>
</tr>
<tr>
<td>8</td>
<td><a href="http://mmsie.org/ont/mmi/tool">http://mmsie.org/ont/mmi/tool</a></td>
<td>MMI tools</td>
<td>Carlos Rueda</td>
</tr>
<tr>
<td>9</td>
<td><a href="http://omv.ontoware.org/2005/05/ontology">http://omv.ontoware.org/2005/05/ontology</a></td>
<td>Ontology Metadata Vocabulary - OMV</td>
<td>Raul Palma, Jens Hartmann, Peter Haase</td>
</tr>
<tr>
<td>11</td>
<td><a href="http://mmsie.org/ont/ooi_epe/science">http://mmsie.org/ont/ooi_epe/science</a></td>
<td>OOI EPE Science Vocabulary</td>
<td>OOI Education &amp; Public Engagement</td>
</tr>
<tr>
<td>12</td>
<td><a href="http://mmsie.org/ont/ooi_epe/platforms">http://mmsie.org/ont/ooi_epe/platforms</a></td>
<td>OOI EPE Platform Vocabulary</td>
<td>OOI Education &amp; Public Engagement</td>
</tr>
<tr>
<td>13</td>
<td><a href="http://mmsie.org/ont/ooi_epe/data_products">http://mmsie.org/ont/ooi_epe/data_products</a></td>
<td>OOI EPE Data Product Vocabulary</td>
<td>OOI Education &amp; Public Engagement</td>
</tr>
<tr>
<td>14</td>
<td><a href="http://mmsie.org/ont/ooi_epe/sites">http://mmsie.org/ont/ooi_epe/sites</a></td>
<td>OOI EPE Site Vocabulary</td>
<td>OOI Education &amp; Public Engagement</td>
</tr>
</tbody>
</table>
## MMI ORR (v.3)

[Image of MMI ORR (v.3) interface]

<table>
<thead>
<tr>
<th>URI</th>
<th>Author</th>
<th>Owner</th>
<th>Version</th>
<th>Submitter</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://cor.esipfed.org/ont/testorg/st">http://cor.esipfed.org/ont/testorg/st</a></td>
<td>Janet Fredericks</td>
<td>testorg</td>
<td>20160718T043440</td>
<td>carueda</td>
</tr>
<tr>
<td><a href="http://cor.esipfed.org/ont/testorg/op">http://cor.esipfed.org/ont/testorg/op</a></td>
<td>Janet Fredericks</td>
<td>testorg</td>
<td>20160718T043223</td>
<td>carueda</td>
</tr>
<tr>
<td><a href="http://cor.esipfed.org/ont/testorg/fish">http://cor.esipfed.org/ont/testorg/fish</a></td>
<td>Carlos Rueda</td>
<td>testorg</td>
<td>20160717T053050</td>
<td>carueda</td>
</tr>
<tr>
<td><a href="http://cor.esipfed.org/ont3/~esipdemo/testmapping">http://cor.esipfed.org/ont3/~esipdemo/testmapping</a></td>
<td>ESIP Demonstrator</td>
<td>testorg</td>
<td>20160714T055038</td>
<td>graybeal</td>
</tr>
<tr>
<td><a href="http://cor.esipfed.org/ont/testorg/avoc">http://cor.esipfed.org/ont/testorg/avoc</a></td>
<td>Carlos Rueda</td>
<td>testorg</td>
<td>20160713T185049</td>
<td>carueda</td>
</tr>
<tr>
<td><a href="http://cor.esipfed.org/ont3/~graybeal/authority">http://cor.esipfed.org/ont3/~graybeal/authority</a></td>
<td>John Graybeal</td>
<td>testorg</td>
<td>20160707T062716</td>
<td>carueda</td>
</tr>
<tr>
<td><a href="https://www.w3.org/ns/ssn">https://www.w3.org/ns/ssn</a></td>
<td>The World Wide Web Consortium</td>
<td>w3c</td>
<td>20160705T050854</td>
<td>carueda</td>
</tr>
<tr>
<td><a href="http://cor.esipfed.org/ont3/~esipdemo/bodyofwater">http://cor.esipfed.org/ont3/~esipdemo/bodyofwater</a></td>
<td>ESIP Demonstrator</td>
<td>testorg</td>
<td>20160705T033321</td>
<td>esipdemo</td>
</tr>
<tr>
<td><a href="http://sweet.jpl.nasa.gov/2.3/propFraction.owl">http://sweet.jpl.nasa.gov/2.3/propFraction.owl</a></td>
<td>Rob Raskin</td>
<td>~carueda</td>
<td>20160704T233258</td>
<td>graybeal</td>
</tr>
</tbody>
</table>
MMI ORR (v.3)

- Enhanced user/organization/permission model
- Overhauled authentication mechanism
- Enhanced performance
- RESTful backend endpoint
- MongoDB; AllegroGraph
- Backend: Scala; comprehensive tests; Travis CI; good coverage
- Front-end: AngularJS
- Docker images for streamlined installation of integrated system
  https://hub.docker.com/r/mmisw/orr-ont/tags/
MMI ORR (v.3)

• Status

  • Recently transitioned to *beta*
    …mostly according to internal testing

• So, please help us as we make progress toward a stable version. Your feedback is most welcome!
ORR

• Registry
  • ORR is a catalog of pointers to ontologies and associated metadata

• Repository
  • ORR hosts the registered ontologies
ORR Capabilities

• Repository of controlled vocabularies and term mappings
• Web resolvable identifiers for ontologies and terms
• Enable added-value applications with semantic and inference
• Ontology metadata
• Versioning
Key requirements

- Community driven, collaborative creation
- Easy-to-use tools
Client applications–ORR interactions

- **Data Portals** create/use ontologies that capture **categories** to be exposed

- **Data providers** create/use ontologies:
  - For the **terms** (concepts) used in their data products and services
  - With mappings between **Data Provider**’s **terms** and **Data Portal** **categories**

- **Data Portal and client applications**
  - Access; Resolve; Query; Aggregate; Archive; ...
ORR instances

- mmisw.org – MMI ORR
- cor.esipfed.org – ESIP COR
- sensorml.com – SensorML ORR
...LET'S GO EXPLORING!
Is your data 5 Star?

🌟 Available on the web (whatever format), but with an open license
🌟🌟 Available as machine-readable structured data (e.g., excel instead of an image scan)
🌟🌟🌟 As before, but using a non-proprietary format (e.g., CSV instead of excel)
🌟🌟🌟🌟 All the above, plus use open standards (RDF & Co.) to identify things, so that people could point at your stuff
🌟🌟🌟🌟🌟 All the above, plus link your data to other people’s data to provide context