The MMI Ontology Registry and Repository: A Portal for Marine Metadata Interoperability

Carlos Rueda – Luis Bermudez – Janet Fredericks

Monterey Bay Aquarium Research Institute
Southeastern Universities Research Association
Woods Hole Oceanographic Institution

Marine Metadata Interoperability

Oceans’09 MTS/IEEE Biloxi – October 27, 2009
Outline

• The Challenge and the Approach
• MMI Ontology Registry and Repository
• Enhancing Standards with Semantics
• Ongoing and Future Work
Got data?

• **Doubling each year** in the physical, Earth, and life sciences*
  • Sloan Digital Sky Survey, GDB Human Genome Database, Global Biodiversity Information Facility, ...

• **Earth sciences**
  – Dozens of terabytes per day (NASA, NOAA)
  – 2002–2017 NOAA’s data holdings expected to grow by a factor of 100 to 74 petabytes**

** NOAA The Nation's Environmental Data: Treasures at Risk (2001)
But, are we exploiting the data as well?

• Sharing data effectively?
• Discovering data easily?
• Easily assimilating and integrating data?

• Taking action to facilitate all of the above?
Interoperability

• Standards efforts
  – International Organization for Standardization (ISO)
  – Open Geospatial Consortium (OGC)
  – World Wide Web Consortium (W3C)

• But there still remains a gap for the effective exploitation of data: semantic heterogeneity
Semantic Web
Linked data*

• Make the content on the web interconnected in a meaningful way for both humans and machines

• Key technologies:
  – RDF and Ontologies
    • Formalized resource descriptions and knowledge representation
  – URI
    • Uniform identification of resources
  – HTTP
    • Hypertext Transfer Protocol

• “Hurricanes,” “Tsunamis,” and “Harmful algal blooms” are “aquatic phenomena events”
• “Katrina” is a particular “Hurricane” event
Ontologies

• “Hurricanes,” “Tsunamis,” and “Harmful algal blooms” are “aquatic phenomena events”
• “Katrina” is a particular “Hurricane” event

However, different communities may use different terminologies
  – Eg., use the term “red tide” instead of “harmful algal bloom” to designate the same phenomenon
• Need mechanisms to link terms across multiple vocabularies with a variety of possible relationships
MMI Ontology Registry and Repository
MMI Ontology Registry and Repository

• Registry
  – MMI ORR is a catalog of (pointers to) ontologies and associated metadata
MMI Ontology Registry and Repository

• Registry
  – MMI ORR is a catalog of (pointers to) ontologies and associated metadata

• Repository
  – MMI ORR contains the registered ontologies
MMI Ontology Registry and Repository

• Registry
  – MMI ORR is a catalog of (pointers to) ontologies and associated metadata

• Repository
  – MMI ORR contains the registered ontologies
MMI Ontology Registry and Repository

• Registry
  – MMI ORR is a catalog of (pointers to) ontologies and associated metadata

• Repository
  – MMI ORR contains the registered ontologies

• But wait, there’s more ...
MMI Ontology Registry and Repository

**MMI ORR**
- SPARQL
- Voc2RDF
- Reasoner
- VINE
- Portal
- Ontologies
- URI Resolver

**Semantic Provider**
- Create and register vocabularies, mappings, ontologies

**Data Provider**
- Sensor; Model; Observing System; Data Processor; Data Aggregator; Person

**Data User/Software Agent**
- Search, query, and resolve ontologies and terms
- Portal; Visualizer; Data Processor; Data Aggregator
MMI Ontology Registry and Repository

Some key requirements

- Controlled vocabulary creation
- Registration/storage of vocabularies
- Metadata association
- Versioning
- Semantic query support
- Term mapping
- Mapping repository
- Vocabulary and term URI resolution

– **Easy-to-use** tools for creation of vocabularies and mappings

– **Semantic mediation services** to diverse data portals and tools
## MMI ORR Portal

![Ontology Registry and Repository](image)

<table>
<thead>
<tr>
<th>URI</th>
<th>Name</th>
<th>Author</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://mmisw.org/ont/agu/indexterm">http://mmisw.org/ont/agu/indexterm</a></td>
<td>AGU Index Terms</td>
<td>AGU</td>
<td>20090604T210443</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/CUAHSI/AquaBiologicalCompound">http://mmisw.org/ont/CUAHSI/AquaBiologicalCompound</a></td>
<td>Aqua Biological Compound</td>
<td>mattr</td>
<td>20090914T213519</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/CUAHSI/AquaBioCore">http://mmisw.org/ont/CUAHSI/AquaBioCore</a></td>
<td>Aqua Biological Core</td>
<td>mattr</td>
<td>20090914T213352</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/argo/instrument">http://mmisw.org/ont/argo/instrument</a></td>
<td>ARGO Instruments</td>
<td>ARGO</td>
<td>20090605T194103</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/argo/parameter">http://mmisw.org/ont/argo/parameter</a></td>
<td>ARGO Parameters</td>
<td>ARGO</td>
<td>20090605T194513</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/argo/qualityFlag">http://mmisw.org/ont/argo/qualityFlag</a></td>
<td>Argo QA/QC Flags</td>
<td>Stephanie Watson</td>
<td>20081116T040146</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/CUAHSI/AtmosphericHydrologicCompound">http://mmisw.org/ont/CUAHSI/AtmosphericHydrologicCompound</a></td>
<td>Atmospheric Hydrologic Compound</td>
<td>mattr</td>
<td>20090914T213157</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/CUAHSI/AtmosphericHydrologicCore">http://mmisw.org/ont/CUAHSI/AtmosphericHydrologicCore</a></td>
<td>Atmospheric Hydrologic Core</td>
<td>mattr</td>
<td>20090914T212958</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/CUAHSI/AtmosphericHydrologicSynthesis">http://mmisw.org/ont/CUAHSI/AtmosphericHydrologicSynthesis</a></td>
<td>Atmospheric Hydrologic Synthesis</td>
<td>mattr</td>
<td>20090914T212842</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/mmi/authority">http://mmisw.org/ont/mmi/authority</a></td>
<td>Authority Vocabulary</td>
<td>MMI</td>
<td>20090317T183911</td>
</tr>
</tbody>
</table>
Vocabulary Editor (Voc2RDF)

Metadata

Contents

<table>
<thead>
<tr>
<th>Class name:*</th>
<th>id</th>
<th>short_name</th>
<th>definition</th>
<th>long_name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>1</td>
<td>D000</td>
<td>Determinations of the abundance of diatom remains in sediment linked to taxonomic identifications that may be mapped to entities in the ITIS taxonomy. These may be at any taxonomic level from sub-species upwards. Diatom taxonomy-related abundance per unit mass of sediment.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>OHWC</td>
<td>‘Other’ (not alkanes, alkenes, alkynes or PAHs) hydrocarbon concentration parameters (including saturation of gaseous species) in all phases of the water column. Does not include parameters expressed per unit weight of SPM.</td>
<td>Concentration of ‘other’ hydrocarbons in the water column.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>ZCTC</td>
<td>Zooplankton (mesozooplankton plus larger pelagic animals excluding fish, reptiles and mammals) carbon biomass parameters presented at the level of taxa that may be mapped to entities in the ITIS taxonomy. These may be at any taxonomic level from sub-species upwards.</td>
<td>Zooplankton biomass expressed as carbon per unit volume of the water column.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>ASLV</td>
<td>Measurements of the displacement of the water column surface from a fixed, stable reference.</td>
<td>Sea level</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>PCHW</td>
<td>PAH concentrations in all phases of the water column. This group includes concentrations per unit volume of the water column in the particulate phase, but not concentrations per unit weight of SPM.</td>
<td>Concentration of polycyclic aromatic hydrocarbons (PAHs) in the water column.</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>OPBS</td>
<td>Urea concentration parameters (including statistical parameters such as standard deviation) in the water column.</td>
<td>Urea concentration parameters in the water column.</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>UREA</td>
<td>Urea concentration parameters (including statistical parameters such as standard deviation) in the water column.</td>
<td>Urea concentration parameters in the water column.</td>
</tr>
</tbody>
</table>
Term Mapping Editor (VINE)
Search

• SPARQL: Query Language for RDF
• Output formats: RDF/XML, N3, JSON, CSV
Search

- SPARQL: Query Language for RDF
- Output formats: RDF/XML, N3, JSON, CSV

<table>
<thead>
<tr>
<th>subject</th>
<th>predicate</th>
<th>object</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://mmisw.org/ont/bodc/instrument/Acoustic_doppler_current_profiler_ADCP">http://mmisw.org/ont/bodc/instrument/Acoustic_doppler_current_profiler_ADCP</a></td>
<td><a href="http://www.w3.org/2000/01/rdf-schema#label">http://www.w3.org/2000/01/rdf-schema#label</a></td>
<td>Acoustic doppler current profiler (ADCP)</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/bodc/instrument/Lowered_ADCP">http://mmisw.org/ont/bodc/instrument/Lowered_ADCP</a></td>
<td><a href="http://mmisw.org/ont/bodc/instrument/name">http://mmisw.org/ont/bodc/instrument/name</a></td>
<td>Lowered ADCP</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/bodc/instrument/Lowered_ADCP">http://mmisw.org/ont/bodc/instrument/Lowered_ADCP</a></td>
<td><a href="http://www.w3.org/2000/01/rdf-schema#label">http://www.w3.org/2000/01/rdf-schema#label</a></td>
<td>Lowered ADCP</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/bodc/parametergroupcode/LERR">http://mmisw.org/ont/bodc/parametergroupcode/LERR</a></td>
<td><a href="http://mmisw.org/ont/bodc/parametergroupcode/definition">http://mmisw.org/ont/bodc/parametergroupcode/definition</a></td>
<td>Parameters as logged by measuring devices, including raw counts/voltages and ADCP relative velocities</td>
</tr>
</tbody>
</table>
Search

- SPARQL: Query Language for RDF
- Output formats: RDF/XML, N3, JSON, CSV

SPARQL query:
```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT ?resource ?creator
WHERE { ?resource dc:creator ?creator. } LIMIT 12
```

<table>
<thead>
<tr>
<th>resource</th>
<th>creator</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://mmisw.org/ont/igoss/20081113T203318/qualityFlag">http://mmisw.org/ont/igoss/20081113T203318/qualityFlag</a></td>
<td>Stephanie Watson</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/gcoos/20081118T044314/sensor">http://mmisw.org/ont/gcoos/20081118T044314/sensor</a></td>
<td>YIXIN Luo</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/mtmi/20081119T135714/usjgofs_hydro">http://mmisw.org/ont/mtmi/20081119T135714/usjgofs_hydro</a></td>
<td>Cyndy Chandler</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/q2o/20081117T133042/qualityFlag">http://mmisw.org/ont/q2o/20081117T133042/qualityFlag</a></td>
<td>Sara Haines</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/qartod/20081113T212733/qualityFlag">http://mmisw.org/ont/qartod/20081113T212733/qualityFlag</a></td>
<td>Stephanie Watson</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/qf/20081116T050933/parameter">http://mmisw.org/ont/qf/20081116T050933/parameter</a></td>
<td>Luis Bermudez</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/ecs/20081120T164852/device_map">http://mmisw.org/ont/ecs/20081120T164852/device_map</a></td>
<td>Amilan</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/cencoos/20081116T071031/parameter">http://mmisw.org/ont/cencoos/20081116T071031/parameter</a></td>
<td>Dale Robinson</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/ndbc/20081114T205115/qualityFlag_h">http://mmisw.org/ont/ndbc/20081114T205115/qualityFlag_h</a></td>
<td>Stephanie Watson</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/seadatapert/20081116T060423/qualityFlag">http://mmisw.org/ont/seadatapert/20081116T060423/qualityFlag</a></td>
<td>Stephanie Watson</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/mmtest/20081117T202337/qualityFlag">http://mmisw.org/ont/mmtest/20081117T202337/qualityFlag</a></td>
<td>Carlos Rueda</td>
</tr>
<tr>
<td><a href="http://mmisw.org/ont/cencoos/20081119T231517/parameter">http://mmisw.org/ont/cencoos/20081119T231517/parameter</a></td>
<td>Dale Robinson</td>
</tr>
</tbody>
</table>
Open Geospatial Consortium
Sensor Web Enablement (OGC SWE)

• “Sensor Web”
discover, access, command sensors of all types

• SensorML
  – to describe sensors and sensor platforms

• O&M
  – to encode observations and measurements

• SOS: Sensor Observation Service
Sensor Observation Service

[Diagram showing the interaction between Client, SOS, and Backend with messages like GetCapabilities, SOS Capabilities, DescribeSensor, SML doc, GetObservation, O&M observation, Open stream, and stream.]
Enhancing SOS with Semantics
Linked (meta)data

• Example with a SOS observation offering

```xml
<sos:ObservationOffering gml:id="observationOffering_1455">
  <gml:description/>
  <gml:boundedBy>
    <gml:Envelope srsName="urn:ogc:def:crs:EPSG:6.5:4326">
      <gml:lowerCorner>36.69623 -122.39965</gml:lowerCorner>
      <gml:upperCorner>36.69623 -122.39965</gml:upperCorner>
    </gml:Envelope>
  </gml:boundedBy>
  <gml:time>
    <gml:TimePeriod gml:id="timePeriod3">
      <gml:beginPosition>2008-06-09T09:36:19Z</gml:beginPosition>
      <gml:endPosition>2008-06-10T02:06:21Z</gml:endPosition>
    </gml:TimePeriod>
  </gml:time>
  <sos:observedProperty xlink:href="http://mmisw.org/ont/cf/parameter/sea_water_temperature"/>
  <sos:observedProperty xlink:href="http://mmisw.org/ont/cf/parameter/conductivity"/>
  <sos:observedProperty xlink:href="http://mmisw.org/ont/cf/parameter/pressure"/>
  <sos:observedProperty xlink:href="http://mmisw.org/ont/cf/parameter/sea_water_salinity"/>
  <sos:featureOfInterest xlink:href="http://mmisw.org/mmi/20080516/system#EarthRealm"/>
  <sos:responseFormat>text/xml; subtype="om/1.0.0"</sos:responseFormat>
  <sos:responseMode>inline</sos:responseMode>
</sos:ObservationOffering>
```
OGC Ocean Science Interoperability Experiment (OSIE)

Initiative to advance standards for interoperability of ocean observing systems

OGC: Open Geospatial Consortium, Inc.
SURA: Southeastern Universities Research Association

Supporting organizations:
OGC Ocean Science Interoperability Experiment (OSIE)

• Use case: Find sensors/systems/data given:
  • sensor/measurement types, region of interest, time period, ...
  • free text or keywords
    – Include results for related terms (broader, narrower)

• Categorization of sensors and observations
  – supports the discovery functionality
  – based on ontologies
The Overall Process

• Portal
  – Creates/uses ontology representing the portal categories

• Data providers
  – Create/uses ontologies for the concepts used in their services
  – Create/uses ontology with mappings between service provider’s terms and portal categories

• Portal and clients in general
  – Query the MMI ORR
Conclusions

• MMI ORR allows data providers and users to include, use, and exploit semantic information in real world applications

• OGC Sensor Web Enablement services can be enriched with semantic references that are resolvable against the MMI ORR

• Marine science data interoperability can be realized with semantic web technologies
Ongoing and future work

• End-to-end semantic solutions for the Marine and Earth science communities
  • Semantically enabled tools
  • Device and observations ontologies

• MMI ORR
  – Collaborative features
  – Enhance mappings support
  – Integration with external ontology tools
• http://mmisw.org/orr – MMI ORR

• http://marinemetadata.org – Marine Metadata Interoperability Project

Thank you!

Carlos Rueda – carueda@mbari.org
Luis Bermudez – bermudez@sura.org
Janet Fredericks – jfredericks@whoi.edu
Registration Options

• Fully hosted
  – mmisw.org-based URIs
  – Direct resolution of URIs
  – Versioning

• Re-hosted
  – Original namespace preserved
  – Indirect resolution of URIs

• Indexed
  – Ontology just incorporated in knowledge base