Using the Semantic Web

M. Benno Blumenthal
International Research Institute for Climate and Society
Columbia University

31 July 2012
CU Metadata Group
Semantic Web

- “a web of data that can be processed directly and indirectly by machines”
- Web 3.0

really it is about Explicit Semantics

- Universal Resource Identifier (URI)
- Resource Description Framework (RDF)
- Web Ontology Language (OWL)
Semantic Web Stack

User interface and applications

Trust

Proof

Unifying logic

Querying: SPARQL

Ontologies: OWL

Rules: RIF/SWRL

Taxonomies: RDFS

Data interchange: RDF

Syntax: XML

Identifiers: URI

Character set: UNICODE

Cryptography

T. Berners-Lee
Why use RDF?

Make implicit semantics explicit

Web-based system for interoperating semantics

RDF/OWL is an emerging technology, so tools are being built that help solve the semantic problems in handling data
Standard Metadata
Many Data Communities

- Semantic walls
- Exchange walls
Super Schema

Standard metadata schema
Super Schema: direct

Standard metadata schema/data service
Flaws

• A lot of work
• Super Schema/Service is the Lowest-Common-Denominator
• Science keeps evolving, so that standards either fall behind or constantly change
RDF Standard Data Model
Exchange

Standard metadata schema
Why is this better?

- Maps the original dataset metadata into a standard format that can be transported and manipulated.
- Still the same impedance mismatch when mapped to the least-common-denominator standard metadata, but...
- When a better standard comes along, the original complete-but-nonstandard metadata is already there to be remapped, and “late semantic binding” means everyone can use the new semantic mapping.
- Can use enhanced mappings between models that have common concepts beyond the least-common-denominator standard.
- EASIER – tools to enhance the mapping process, mappings build on other mappings.
RDF Architecture

Virtual (derived) RDF
Example: Search Interface

- Additional Semantics
- Dataset Ontology
- Users
- Datasets
- Search Ontology
- Search Interface
Sample Tool: Faceted Search

http://iridl.ldeo.columbia.edu/ontologies/query2.pl?

IRI Data Library & Map Room Search Interface

Click on search terms in the categories on the left side of this page to display a list of data sets or maps that are associated with the terms you have selected. Each of the matching data sets or maps includes a title, a short description, and an icon. Click on a title or icon to see the data set or map page you want. Use the check boxes in the list of "Search Terms You Have Selected" to remove individual terms from the search, or the "New Search" link to remove all selected terms at once.

New Search

Search Terms You Have Selected:
- Person: S. Levitus

Search Results:

- DASILVA SMD94
  - DASILVA Atlas of Surface Marine Data 1994: a five-volume atlas series depicting the seasonal and yearly variations of the surface marine atmosphere over the global oceans. Resolution: 1x1; Longitude: global; Latitude: global; Time: [Jan 1945, Dec 1993]; monthly

- NOAA NODC WOA01
  - NOAA NODC WOA01: World Ocean Atlas 2001, an atlas of objectively analyzed fields of major ocean parameters at monthly, seasonal, and annual time scales. Resolution: 1x1; Longitude: global; Latitude: global; Depth: [0 m, 5500 m]; Time: [Jan, Dec]; monthly
Distinctive Features of the search

- Search terms are interrelated
- terms that describe the set of returns are displayed (spanning and not)
- Returned items also have structure (sub-items and superseded items are not shown)
Architectural Features of the search

- Multiple search structures possible
- Multiple languages possible
- Search structure is kept in the database, not in the code

http://iridl.ldeo.columbia.edu/ontologies/query2.pl
RDF: framework for writing connections

Triplets of
• Subject
• Property (or Predicate)
• Object

URI’s identify things, i.e. most of the above Namespaces are used as a convenient shorthand for the URI’s
Datatype Properties

{WOA} dc:title "NOAA NODC WOA01"
{WOA} dc:description "NOAA NODC WOA01: World Ocean Atlas 2001, an atlas of objectively analyzed fields of major ocean parameters at monthly, seasonal, and annual time scales. Resolution: 1x1; Longitude: global; Latitude: global; Depth: [0 m, 5500 m]; Time: [Jan, Dec]; monthly"
Object Properties

{WOA} iridl:isContainerOf {Grid-1x1},  
{Grid-1x1} iridl:isContainerOf {Monthly}
# WOA01 diagram

<table>
<thead>
<tr>
<th>iridl:isContainerOf</th>
<th>NOAANODCWOA01Grid-1x1</th>
<th>NOAANODCWOA01Grid-5x5</th>
<th>NOAANODCWOA01Masks</th>
</tr>
</thead>
<tbody>
<tr>
<td>iridl:last_modified</td>
<td>%@dateTime2006-05-26T14:40:58Z</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iridl:hasDocumentation</td>
<td>p213:dataset_documentation.html</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>iriref:Conkright_etal2002</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>iriref:Levitus1982</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dc:title</td>
<td>NOAANODCWOA01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dc:description</td>
<td>NOAANODCWOA01: World Ocean Atlas 2001...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>term:isDescribedBy</td>
<td>iridl:Ocean</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Diagram of NOAANODCWOA01 and its components]
Standard Properties

{WOA} dcterms:hasPart {Grid-1x1},
{Grid-1x1} dcterms:hasPart {MONTHLY}

Alternatively

{WOA} iridl:isContainerOf {Grid-1x1},
{iridl:isContainerOf} rdfs:subPropertyOf
{dcterms:hasPart}
Object properties provide a framework for explicitly writing down relationships between data objects/components, e.g. vague meaning of nesting is made explicit.

Properties also can be related, since they are objects too.
Search Interface Term

Virtual Triples

Use Conventions to connect concepts to established sets of concepts
Generate additional “virtual” triples from the original set and semantics
RDFS – some property/class semantics
OWL – additional property/class semantics: more sophisticated (ontological) relationships
SWRL – rules for constructing virtual triples
Multiple Ways of Expressing Concepts in RDF

Note that there are many world views in how to express concepts: concepts as classes vs concepts as individuals vs concept as predicate values
Nuanced tagging

Concepts as objects can be interrelated: specific terms imply broader terms

Object ends up being tagging with terms ranging from general to specific.

Search can then be nuanced tagging can proceed in absence of perfect information
Faceted Search Explicated

IRI Data Library & Map Room Search Interface

Click on search terms in the categories on the left side of this page to display a list of data sets or maps that are associated with the terms you have selected. Each of the matching data sets or maps includes a title, a short description, and an icon. Click on a title or icon to see the data set or map page you want. Use the check boxes in the list of "Search Terms You Have Selected" to remove individual terms from the search, or the "New Search" link to remove all selected terms at once.

New Search

Search Terms You Have Selected:

- Person: S. Levitus

Search Results:

- DASILVA SMD94
  DASILVA Atlas of Surface Marine Data 1994: a five-volume atlas series depicting the seasonal and yearly variations of the surface marine atmosphere over the global oceans. Resolution: 1x1; Longitude: global; Latitude: global; Time: [Jan 1945, Dec 1993]; monthly

- NOAA NODC WOA01
  NOAA NODC WOA01: World Ocean Atlas 2001, an atlas of objectively analyzed fields of major ocean parameters at monthly, seasonal, and annual time scales. Resolution: 1x1; Longitude: global; Latitude: global; Depth: [0 m, 5500 m]; Time: [Jan, Dec]; monthly
Search Interface

- Items (datasets/maps)
- Terms
- Facets
- Taxa
{item} dc:title dc:description rss:link iridl:icon
dcterm:isPartOf {item2}
dcterm:isReplacedBy {item2}

{item} trm:isDescribedBy {term}

{term} a {facet} of {taxa} of {trm:Term},
{facet} a {trm:Facet}, {taxa} a {trm:Taxa},
{term} trm:directlyImplies {term2}
RDF Architecture

Virtual (derived) RDF

queries

queries

queries
IRI RDF Architecture

Data Servers

Ontologies

Start Point

RDF/XML-Schema Crawler
XSLT/GRDDL ingest

XML Schema to OWL translation
Owl Semantics
SWRL Rules
SeRQL CONSTRUCT

Search Queries

Search Interface

Sesame

Sesame

bibliography

MMI

JPL

Standards Organizations

Location Canonicalizer

Time Canonicalizer
Models, Crosswalks, and Objects

Structure of the RDF information that we are using to represent data objects in multiple frameworks (see full figure)
Semantic Crosswalk for metadata translation
IRI Climate and Society Map Room

The climate and society maproom is a collection of maps and other figures that monitor climate and societal conditions at present and in the recent past. The maps and figures can be manipulated and are linked to the original data. Even if you are primarily interested in data rather than figures, this is a good place to see which datasets are particularly useful for monitoring current conditions.

El Niño, La Niña and the Southern Oscillation
Pages for monitoring and understanding the impact of ENSO (El Niño Southern Oscillation).

Climate Monitoring and Analysis
Current and historical climate conditions around the globe.

Climate and Health Resource Room
Climate affects health in a number of ways. These effects may be direct, as with heat stress, or indirect, as with infectious diseases such as malaria and meningitis.

International Federation of Red Cross and Red Crescent Societies: Forecasts in Context
Working in collaboration with the International Federation of Red Cross and Red Crescent Societies, IRI has developed a map tool to provide context for global precipitation forecast information.
Semantic Web

- Universal ids (URIs)
- Multiple partial representations adding to be a more complete picture

John Godfrey Saxe (1816-1887)