GeoLink: Semantics and Linked Data for the Geosciences

The Web is the API.
Goal:

- Bring together experts from the geo sciences, computer science, and library science;
- to develop Semantic Web components;
- that support discovery and reuse of data and knowledge.
## Leadership

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robert Arko</td>
<td>Lamont-Doherty Earth Observatory</td>
<td>ICER14-40221</td>
</tr>
<tr>
<td>Suzanne Carbotte</td>
<td>Lamont-Doherty Earth Observatory</td>
<td>ICER14-40114</td>
</tr>
<tr>
<td>Cynthia Chandler</td>
<td>Woods Hole Oceanographic Institution</td>
<td>ICER14-40170</td>
</tr>
<tr>
<td>Michelle Cheatham</td>
<td>Wright State University</td>
<td>ICER14-40202</td>
</tr>
<tr>
<td>Douglas Fils</td>
<td>Consortium for Ocean Leadership</td>
<td>ICER14-40195</td>
</tr>
<tr>
<td>Timothy Finin</td>
<td>University of Maryland Baltimore County</td>
<td>ICER14-40139</td>
</tr>
<tr>
<td>Pascal Hitzler</td>
<td>Wright State University</td>
<td></td>
</tr>
<tr>
<td>Krzysztof Janowicz</td>
<td>University of California Santa Barbara</td>
<td></td>
</tr>
<tr>
<td>Matthew Jones</td>
<td>University of California Santa Barbara</td>
<td></td>
</tr>
<tr>
<td>Kerstin Lehnert</td>
<td>Lamont-Doherty Earth Observatory</td>
<td></td>
</tr>
<tr>
<td>Thomas Narock</td>
<td>Marymount University</td>
<td></td>
</tr>
<tr>
<td>Lisa Raymond</td>
<td>Woods Hole Oceanographic Institution</td>
<td></td>
</tr>
<tr>
<td>Mark Schildhauer</td>
<td>University of California Santa Barbara</td>
<td></td>
</tr>
<tr>
<td>Peter Wiebe</td>
<td>Woods Hole Oceanographic Institution</td>
<td></td>
</tr>
</tbody>
</table>
Repositories:

- Biological and Chemical Oceanography Data Management Office (BCO-DMO)
- Data Observation Network for Earth (DataONE)
- Interdisciplinary Earth Data Alliance (IEDA)
- International Ocean Discovery Program (IODP)
- Long Term Ecological Research (LTER) Network
- Marine Biological Laboratory/Woods Hole Oceanographic Institution (MBLWHOI) Library
- Rolling Deck to Repository (R2R) Program

Additional Collaborators:
- Advanced Cooperative Arctic Data and Information Service (ACADIS)
- CLIVAR and Carbon Hydrographic Data Office (CCHDO)
- Federation of Earth Science Information Partners (ESIP)
- Index to Marine and Lacustrine Geological Samples (IMLGS)
- U.S. Geological Survey - National Geochemical Database (NGDB)
History:

- “OceanLink” EAGER in 2013, prototype effort to federate two geo repositories (BCO-DMO, R2R) using Semantic Web.

- Earlier work on Linked Data and ontologies at WHOI (BCO-DMO), Consortium for Ocean Leadership, and others, with RPI.
Domains:

Modelling and publishing to date has focused on:

- Marine environments (Marine geology .. Marine ecosystems .. Biogeochemistry .. Paleo climatology)

- Fieldwork/ Expeditionary data

- Both sensor and sample data, from observing networks to the “long tail”

.. informed by many workshop reports and RCNs (C4P, CReSCyNT, iSamples, ..)
Use Cases:

- **Ocean ecosystems:** What’s happening in the subpolar North Atlantic (surface to 1000 m), where are the zooplankton, and when do they come out of diapause? Requires primary production, species composition, abundance and size distribution, nutrients, and basic hydrographic data.

- **Seafloor morphology:** Find and assess bathymetry data as baseline for investigations such as formation and aging of the ocean floor, global ocean circulation, and evaluation of biological habitats; as well as cruise planning and estimation of time-on-station for water column measurements.
Work Plan:

- Publish set of reusable ontology design patterns (ODPs) that describe
  - field expeditions
  - sensor + sample datasets
  - laboratory analyses
  - journal publications
  - conference presentations
  - theses/reports
  - funding awards

- Upgrade existing repositories to publish their content as [Linked Data](https://www.w3.org/2011/06/linked-data-vocab/), using those ODPs;

- Populate an integrated knowledgebase and exercise it against science use cases.
“Linked Data” principles:

• Use **URIs** (network names) to identify things;

• Use **HTTP** URIs, so these things can be dereferenced by both humans and machines;

• Describe these things using standard languages such as **RDF** and **SPARQL**; and

• Include URIs (ie. links) to other related things.

The Web is the API.
Connections to other BB projects:

- GeoLink will deploy an integrated knowledgebase (merged triplestore) at data.geolink.org that
  - Provides reference URIs for Expeditions, Persons, Awards, etc;
  - Infers new relationships + derives new knowledge;

- Partners could:
  - Build new (or retrofit existing) RDF/SPARQL-aware clients;
  - Mark up non-semantic services eg. add URIs to a Web Feature Service.
Work to date:

- Developed core ontology design patterns, including a simplified/integrated “view” to aid data managers;
- Selected initial set of useful controlled vocabularies, and cross-decked into ontology language (OWL) with network names (URIs);
- Deployed central knowledgebase, started to populate with partners’ data and build Web browser interface.
Work to date: (cont.)

Core classes

1. Award 7. Instrument 13. Program
5. Feature 11. PhysicalSample 17. Vessel
6. Format 12. Place
Work to date:

Class *PhysicalSample*

Properties

- :identifier
- :isPartOf *Dataset*
- :hasParent *PhysicalSample*
- :hasRepository
- :hasCruise *Cruise*
- :hasCollector *Person*
- :hasTectonicSetting
- :hasGeologicAge
- :hasSamplingMethod
- :hasMaterialType
- :hasFeature *Feature*
- :hasLocation *Place*
- :hasPosition
- :hasDepth

IHO/GEBCO Undersea Features
GVP Volcanoes
SCAR Composite Gazetteer
Example #1: VERTIGO Project

Cruise Description + Trackline at R2R

Sediment Trap Flux Dataset at BCO-DMO/WHOAS

Multibeam Sonar Dataset at R2R/NOAA

Journal Article on VERTIGO using DOI

NSF Funding Award

Investigator (K. Buesseler)

ORCID

GeoLink Project - EarthCube C4P, April 28, 2015
Example #2: Bering Sea Samples

Chemical analyses at USGS NGDB

Core Descriptions at IMLGS

Sample Identifier using IGSN

Cruise Description + Trackline at R2R

Investigator (D. Scholl) in ScienceBase
Challenges:

- Lack of key vocabularies published online using OWL with URIs;
- Lack of gazetteer data (eg. physiographic features) published online with URIs and proper geometries;
- Lack of universal Person and Organization identifiers with sufficient metadata;

and

- Need to map/match instances manually, at least in the beginning.
See you at the All-Hands Meeting!

Questions?

info@geolink.org