EarthCube Building Blocks:

OceanLink

Leveraging Semantics and Linked Data for Geoscience Data Sharing and Discovery
Overview

Goal

• Enable discovery of geoscience data and knowledge, and ultimately, integration

Strategy

• Publish content from existing network of repositories as Linked Open Data (LOD)
• Enable horizontal semantic integration
• Provide tools + services useful to working scientists
Ocean Science

- Research vessels collect data from the solid earth, water column, atmosphere
- Many repositories already interoperate
- Approach is extensible to other geo domains
Project Team

Lamont-Doherty Earth Observatory
  Robert Arko
  Suzanne Carbotte

Marymount University
  Tom Narock  (lead)

University of Maryland, Baltimore County
  Tim Finin

Woods Hole Oceanographic Institution
  Cynthia Chandler
  Lisa Raymond
  Adam Shepherd
  Peter Wiebe

Wright State University
  Pascal Hitzler
  Michelle Cheatham
  Adila Krisnadhi
Collections

- Biological & Chemical Oceanography Data Management Office (BCO-DMO)
- Rolling Deck to Repository (R2R)
  - cruise catalog + underway enviro sensor data
- Marine Biological Laboratory / Woods Hole Oceanographic Institution (MBLWHOI) Library
  - published articles, theses, tech reports, datasets
- AGU meeting abstracts
- NSF funding award abstracts
Ontology Design Patterns

- Core set of conceptual primitives from Ocean Science
  - Vessel
  - Cruise
  - Instrument
  - Dataset
  - Person
  - Organization
  - etc.

- Reuse existing standard vocabularies where they exist (DCAT, FOAF, PROV)

- Maximize reusability, minimize commitment
ODPs (cont.)

- Patterns published as OWL files with embedded axioms and local vocabularies
eg.

  *Cruise* must have a *Vessel*
  *Cruise* may have a *Person* in the *Role* of *Chief Scientist*

- Leverage existing alignment among repositories that use eg. NERC Vocabulary Server

- Inference to find relationships among cruises, datasets, people, publications, etc.
Work Plan

1. Model, align, inference over existing LOD collections (BCO-DMO + R2R)
   - Develop use cases eg. "find publications related to cruises at the Bermuda Rise that produced CTD profiles and/or seafloor mapping data"
   - Develop ODPs
   - Map existing collections to ODPs

2. Publish LOD for other collections (Library, AGU, NSF) and map to ODPs

3. Prototype end-user tools and services
   - Search/browse across federated LOD collections
   - Edit ontologies
   - Annotate LOD resources incl. provenance
Initial Results

“An Ontology Pattern for Oceanographic Cruises” (Krisnadhi et al.)

Technical Report and draft set of ODPs

Reuses existing patterns including
- Semantic Trajectory (Janowicz et al.)
- Information Object
- Simple Event Model

to model a Cruise and ship’s track

R/V Atlantis cruise AT22 (Scotian Shelf Survey, August 2012) Basemap: GMRT
Lessons

- Recurrent themes in EarthCube Workshop Reports eg.
  - Data are still difficult to discover and access
  - Data attribution and citation are critical
  - Reuse of data still hampered by need for *implicit* understanding

- Collaboration between Geo Science and Computer Science works best with Use Cases

- In-person working meetings are key to initial progress
  - Oct. 2013  Woods Hole
  - Nov. 2013  Baltimore
  - Jan. 2014  Washington
  (probably more)
Acknowledgements

“EAGER: Collaborative Research: EarthCube Building Blocks, Leveraging Semantics and Linked Data for Geoscience Data Sharing and Discovery”

NSF Funding Awards:

ICER 13-54990 LDEO
ICER 13-54693 UMBC
ICER 13-54778 WSU
ICER 13-54107 WHOI

September 15, 2013 - August 31, 2014
Thank you.

www.oceanlink.org