February 25, 2016

Some Introductory Comments on the Track Topic

Gary Berg-Cross

Ontolog, RDA US Advisory Committee, EarthCube Technical Advisory Committee
Outline of this Intro

1. Topic relevance – SI has a long history in the GeoSciences
2. Points from our Panel, EarthCube & INSPIRE Thrust
   1. Conceptual models, languages, legacies, raw data & heterogeneity, constrained interpretations, repositories, brokers, context etc.
3. Some Solution Ingredients in the era of Big Science and Data
4. Recap
5. Some References and Links
History – Many efforts going back to the 80s (commercial GIS) & 90s

• Complex Earth system issues such as climate change and water resources, mean that geoscientists must work across disciplinary boundaries; and access and understand data & systems outside of their fields.

• Examples:
  • “the language of latitude and longitude is universal, but the general language for describing all phenomena distributed over the surface of the Earth is not, and has not yet been fully defined.”

In EarthCube & INSPIRE some of the Challenges are Recognized – Complex, Heterogeneous & Federated Environments

Challenges to Interdisciplinary Science

From EarthCube Domain Workshops
... query across complex, heterogeneous, and federated environments
... community would benefit from access to a web clearing house/portal

Requires

Conceptual Modeling, Formalized Vocabulary, Intelligent Search, Data Publishing Support

Provides

W3C RDF
W3C OWL
W3C SPARQL

Semantic Web
A new generation of geoscience-aware intelligent systems with novel forms of reasoning and learning will be needed to address the challenging features of geoscience research questions:

- Spatio-temporal structure
- Intermittent, sparse data
- Heterogeneous and dispersed data
- Small sample size
- Multi-resolution, multi-scale observations
- High dimensionality
- Tolerance of measurement
- Uncertainty at all stages
- Process-centered models
- Combine diverse (physical, geological, chemical, biological, ecological, anthropogenic) phenomena
- Objects and processes with amorphous spatial/temporal boundaries
- Contextualized by rich background knowledge
- Hard to understand information
- Lack of ground truth

Figure 7. Geoscience data exhibits a variety of differentiating and challenging characteristics that require new research to extend the traditional intelligent systems approaches that are successfully used in commercial domains involving massive datasets.
Integration Challenges in GeoSpace Science

- Radio instruments
- Optical instruments
- Data
- Assimilation Models
- Numerical Simulations

- Data Scientists
- Theorists

- Ionosphere
- Solar
- Atmosphere
- Magnetosphere
Agent Brokering employs central mechanisms to help resolve such things as disparate vocabularies, support data distribution requests, enforce translatable standards and to enable uniformity of search and access in heterogeneous operating environments.
Some GeoScience Ontology Snapshots - Atmosphere & Geo Connections

- snapshots

Soil?

Gas flux?
Some Ontologies are Lists of Features

http://hydro10.sdsc.edu/cinergi_ontology/GeographicFeatures.owl

yagoGeoEntity

bank

cinergiGeoEntity

Community Inventory of EarthCube Resources for Geosciences Interoperability (CINERGI)
Physical Qualities
Graphic Overview of S/O (EarthCube) Manifesto

Guiding principles
1. Uses Cases
2. Lightweight - opportunistic (ODPs) Reduce Entry Barrier
   1. Semantic interoperability with semantic heterogeneity
4. Bottom-up & top-down approaches
5. Domain - ontology engineer teams
6. Formalized bodies of knowledge across Earth science domains
7. Reasoning services

Knowledge Infrastructure Vision

"Insertion" Architecture & Workflow Between

Horizontal Integration
Vertical Integration

Community Understanding of Semantic role and value
Recap & Some Thought to be Considered in the Track

• There is a long history of interest and increasing work we can leverage.
• Big Science & Big Data provide motivating challenges
• Semantic Web/LOD work is a driver and some useful things have been built
• But there seem practical & foundational challenges to make semantic approaches successful.
  • Many ontologies but may be too shallow and not well related
  • How many decisions need be made to handle varied contexts over time?
• Opportunities exist in the various Earth Sciences with large Projects like EarthCube.
• We should keep in mind the challenges of communicating across the Big Data, Semantic Web and Applied Ontology disciplines and projects.
Some References & Links


Supplementary
Community Priorities & Talking a Different Language

Enhance Data Standards, Models, Analytics, & ISO-Metadata catalogs.

Converting RDB into triples, RDF Vocabulary for annotating data.

Logical Axioms, deep knowledge, alignment.

Knowledge Infrastructure & Ontology applications:
- Smart Search, discovery & annotation
- Semantic services
- Knowledge Infrastructure

Sharing & Interoperability

We want Faster Processing, better visualization, better data management plans, open data, but simple semantics effort....

Patil & Polzer point

LOD is too complex/not rich enough.
Big data era means that diverse and voluminous data types are increasing & need to be described with semantics.

- For example, most of the database research self-assessment reports recognize that the **thorny question of semantic heterogeneity**, that is of **handling variations in meaning or ambiguity in entity interpretation**, remains open.

Exploit semantics of ontological relationships

Richer languages

Use Heuristics and Machine-learning

Modular Reuse

KE Tools

General Integration Approaches on Different Architectural Levels
Portion on SOCoP Ontologies in OntoHub
https://ontohub.org/repositories/socop/ontologies

Use Case Description Ontology (UCDO). Use cases are a commonly used technique for specifying the functional requirements of a project. This ontology is concerned with the formal representation of the flow of events that occurs in a use case, including alternative flows and exceptional flows. Nonfunctional requirements are specified separately.

- **Om** .owl
  - SOCoP
  - OWL implementation of ISO 19156 Observations schema

- **Sam** .owl
  - SOCoP
  - OWL implementation of ISO 19156 sampling features schema

- **Geoc** .owl
  - SOCoP
  - Simple lite geometry, geo features and relations

- **Geog** .owl