Ontology Summit 2016

Track Synthesis: Semantic Integration for the GeoSciences & Geographically Distributed Sensor and Control Systems

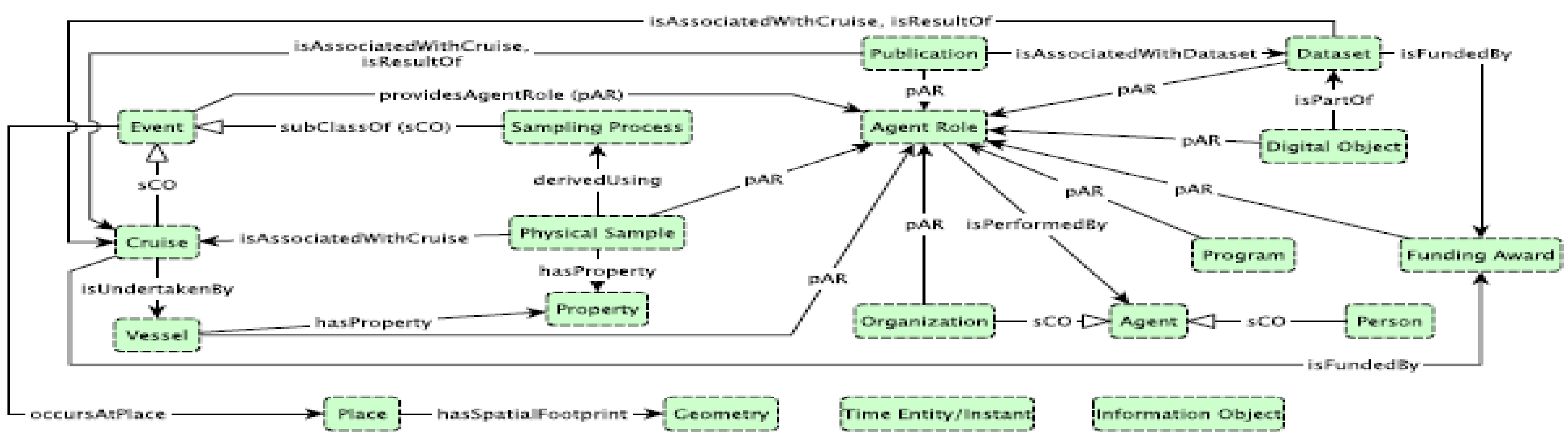
The mission of our track is to scope out challenges, opportunities, current & emerging practices in support of cross domain GeoSciences & Smart Grid Systems *semantic interoperability such as* unified view of data from different sources. *that is robust, well founded and practical.*

Co-Champions: Gary Berg Cross & Ken Baclawski

April 14, 2016

Outline

- Commonality across tracks
- Our Speakers & Session Topics
- Interoperability Challenges & Opportunities
- Synthesis of Semantic Science Challenges and Opportunities



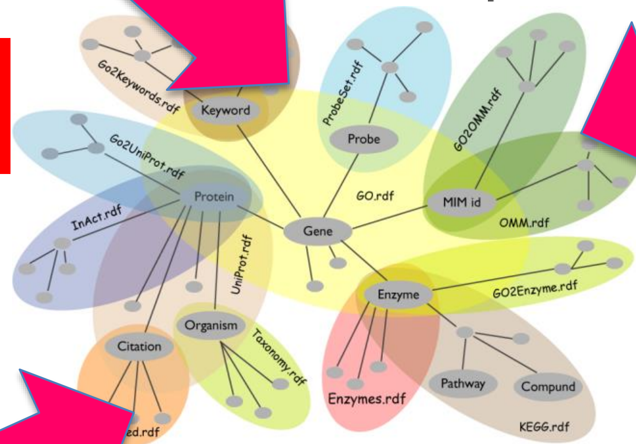
Each node represents a content pattern.

Krisnadhi "GeoLink Data"

Summit Track Themes Are Highly Related

Vocabulary harmonization is bottleneck and is impeded by lack of reference ontologies

Lack of small, ontological building blocks



Modeling axioms or knowledge representation language fragments cause difficulties in terms of an increase in reasoning complexity or reducing the reusability of ontologies

Challenges include: Domain information is heterogeous described in multiple schemas, different vocabularies & markup languages and ontologies with different level of granularity in the data & different conceptualization.

Our Speakers & Their Talks

1. Gary Berg-Cross Brief overview of GeoScience and Semantic Integration
2. Brandon Whitehead (University of Auckland, Auckland) An overview of semantic models in the geosciences: what do we have and where are we going?
3. Ruth Duerr (Ronin Institute) Semantics and the discovery and use of data and data services [(Bcube)]
4. Adila Krisnadhi (Wright State) Dealing with Semantic Heterogeneity in Data Integration using Modular Ontology Patterns (GeoLink Project)
5. Matthew Mayernik (UCAR) Building Geoscience Semantic Applications Using Established Ontologies
6. Steve Ray, Carnegie Mellon Silicon Valley, Semantic Interoperability Issues for the Smart Grid
7. Marshall Ma, RPI, SEM+: A Tool for Concept Mapping in Geoscience
8. Shirly Stephens and Torsten Hahmann, University of Maine, Semantic Alignment of the Groundwater Markup Language with the Emerging Reference Hydro Ontology HyFO

Interoperability Challenges

Big Science, Big Data and Big Industry provide many motivating challenges.

- Existing GeoScience standards, ontologies, models and associated were typically developed in isolation and major problems exist when they are combined. Some glue is needed to intergrate and harmonize these.
- The range of systems, data & semantic content is now broad but increasingly has to be integrated to be of use to Science & Society
- Activities such as NSF's EarthCube feature different GeoScience domains working on semantic interoperability issues
- These promote understanding of the state of the semantic practice across different domains

Issues

Success Stories

Some Broad Questions

- What range of semantic content is at least being shared and used on the Web?
  - From vocabularies to formally axiomatized ontologies
- What limits their use to support interoperability?
  - How do we go beyond “semantic tagging” with vocabularies to find relevant data to share and use?
    - E.g Tags used by NSIDC
- What ontologies are available/being used/required?
  - VIVO-ISF uses the Basic Formal Ontology (BFO) as its upper level ontology
- How can we find semantic content to advance interoperability?
- Where do we put Earth science ontologies (or semantic models; the word ontology has kind of lost its meaning) once they have been created?
  - e.g. LOV, ontology repositories
    - ESIP, bio portal and OntoHub OOR

Tools

Best Practices

Standards are part of the Approach

The traditional approach to interoperability is to create some standard (e.g. a controlled vocabulary, API, Smart Grid Standards etc.) which may be at a high level and/or at the domain level.

1. But standards at all levels are expressed in a variety of languages with varying degrees of formality and completeness.
2. Multiple standards in the same or overlapping domains can become impediments to interoperability.
    1. There need to be bridging concepts even within a single domain.
3. Even when a standard is expressed as an ontology it may be too shallow and not well related enough to support robust interoperability.

GeoScience Illustration: Semantic integration in Ground Water remains a challenge because

- Various types of standards are, for the most part, heterogeneous, meaning they:
  - are mostly fragmented and disconnected, describing either surface or groundwater.
  - lack foundational grounding.
  - use the same or similar terms but with differences in semantics.
  - are described using different formal (or non-formal) languages.

Opportunities Amidst Practical & Foundational Challenges

- There is a long history of interest and increasing work in geographically distributed systems & data that we can leverage.
- There is general sense of a convergence on standards for aligning components, including catalogs, vocabularies, services and information models.
  - But as we note there are limitations to these alignments
- The application of ontologies to provide semantics for this interoperability is seen as essential and more groups are ready to consider ontologies and ontological techniques to improve interoperability.
- We should keep in mind the challenges of communicating across the Big Data, Semantic Web and Applied Ontology disciplines and projects.
  - There remain misunderstandings about what can be accomplished , what the limitations are and how to work.
- Semantic Web/LOD work is a driver & some useful tools have been built.
  - Some useful practices have been developed that can be leveraged /expanded.
- But there seem practical and foundational challenges (such as semantic alignment of vocabularies, handling data and systems heterogeneity & development of reusable building blocks) to make semantic approaches successful, scalable and robust across and within domains.

Moving Forward in A GeoScience Domain: Matching, Alignment & Semantic Integration Techniques (Stephen & Hahmann)

Existing ontology matching and **alignment** techniques find similarities, equivalences and sub-sumption relations between two (or more) ontologies given that they are:

- syntactically and schematically integrated.
- of similar scope & no more expressive than OWL.

(Whereas) **semantic integration** between existing hydrologic ontologies and schemas additionally requires:

- **Translation** between ontology languages.
- More rigorous specification of the semantics in each ontology.
  - And perhaps **deeper semantics**

This can currently be done only by manual integration of the ontologies.... But use of a suitable **reference ontology** may automate this.

More Focused Questions & Semantic Challenges

- Upper-level and many domain ontologies are important for SI but there are challenges:
- Many of the upper and domain ontologies are hard to understand or have too many terms,
- are too abstract, with too complicated axioms to understand and yet remain too far from real data.
- Impose ontological commitments that may not be acceptable by all parties.
  - Brittle and/or costly, hard to extend,
  - And carelessly extension may break the whole thing ontology.

Some Best Practice Semantic Opportunities

1. We've considered reference ontologies that bridge across a domain or even several domains
2. Common conceptual models are needed as are organizing ontologies like ViVO
    1. We can leverage existing ontologies where possible to reduce modeling effort
    2. Constrain ontology needs & possibilities using information about
        - Particular entity types & relationships
        - Significant legacy dependencies
3. Ontology design patterns (ODPs) may represent reusable, modular solutions to frequently occurring modeling problem in a domain(s)
4. Minimalistic schema can act as a building block of a more complex ontology or partial ontologies
    - Existing controlled vocabulary can be accommodated as a pattern view and rules used to map to local vocabularies.
5. Content patterns abstract examples:
    - Agent, Agent Role, Event, Information Object, Identifier, Personal Info Item, Person Name, Property Value.

- Testing conformance to and integrity of standards using ontologies

Example of an Emerging Best Practice

Use reference ontologies:

- Foundational grounding (‘ontologically sound’)
- Broad coverage of the domain
  - Example entire hydro (both surface and subsurface water storage and flow)
- Provide detailed, rigorous axiomatization of all semantics in a language that affords automated verification and reasoning

Q. Can this and ODPs be combined?

Practical Questions
(largely from Brandon Whitehead)

1. Who maintains the ontology once it's released into the wild; i.e. published or... portaled?
2. How do we verify and validate these structures (ontology efficacy)? (i.e. if an ontology is created to do some thing, x; who verifies it actually does, x?)
3. Who owns the ontology once it is published?
4. Do ontologies need a license?
5. Do we treat it as code, IP, both?
6. How do we keep track of conceptual drift in ontologies?
7. How do we avoid ontology “hijacking”?
   – -Introducing ontological commitments that change the semantics of specific classes or properties in the original ontology.

Tools & Technologies

1. Need concept search supported by conceptual similarity like SEM+
    - "Tools may die but the ideas live on" Ruth Doerr
2. 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.
    – When searching for data current semantic brokers still yield:
        1. Invalid content or responses
        2. Unidentifiable document types
        3. Empty metadata elements - especially required elements
3. Need KE tools to build and bridge ontologies