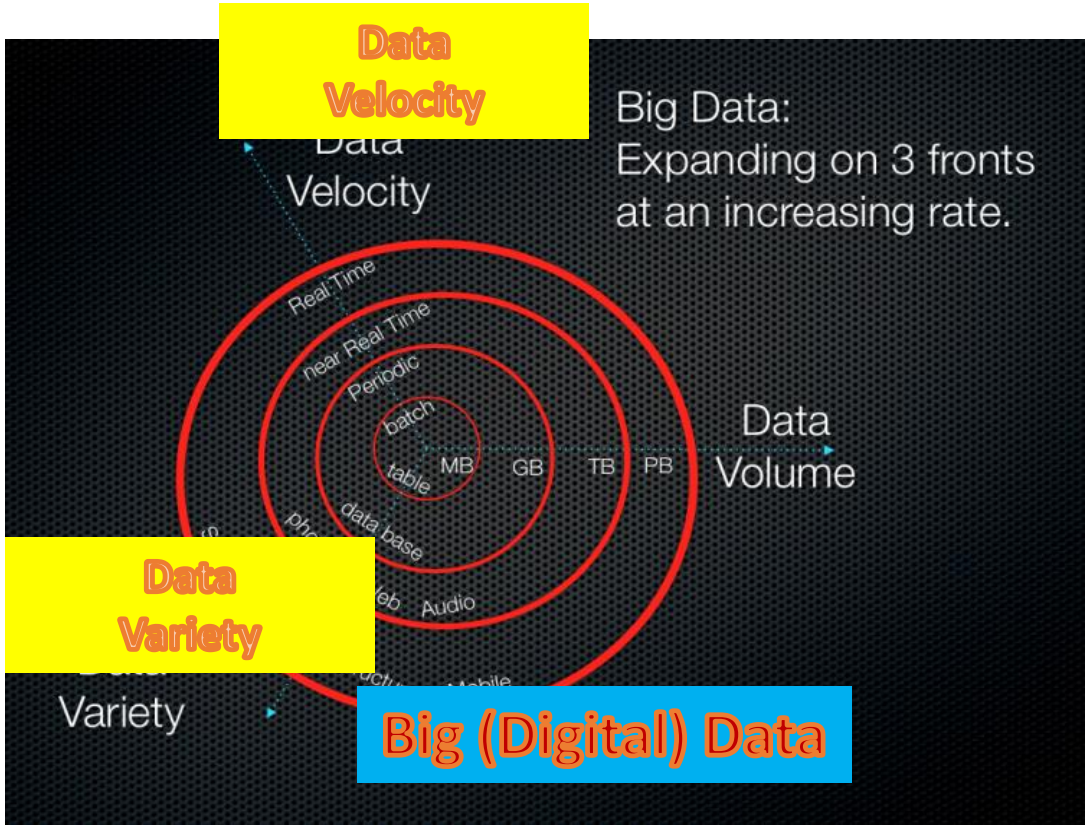
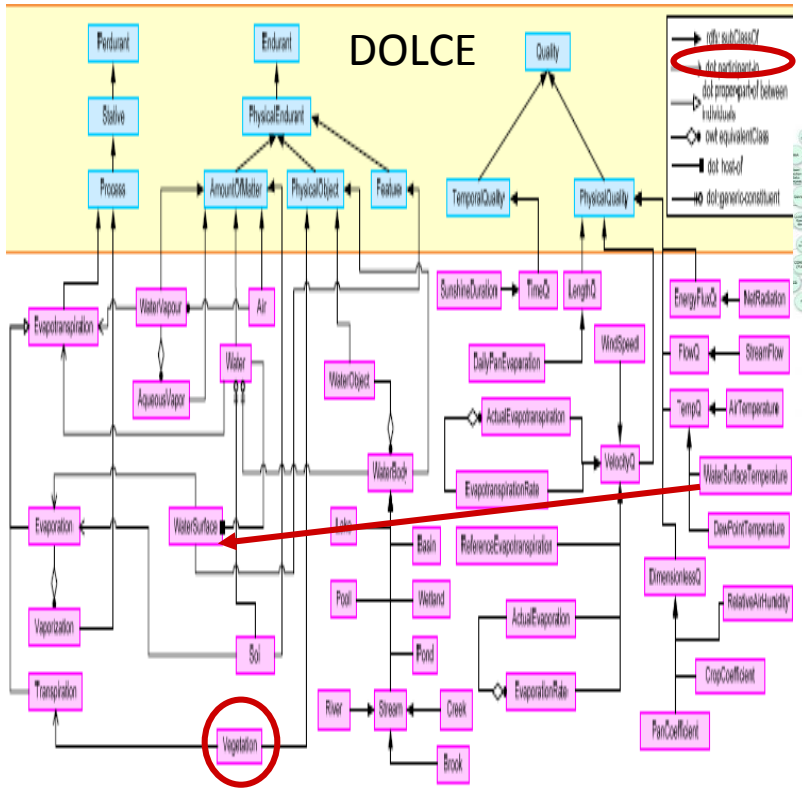
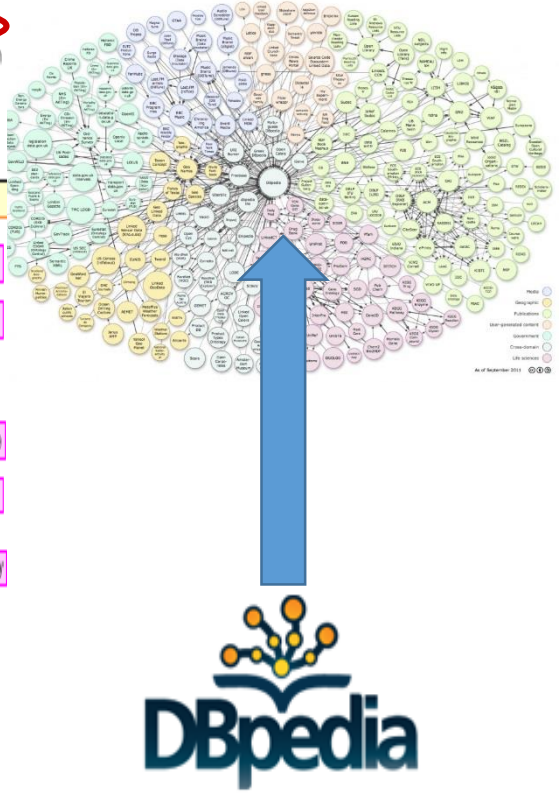
Ontology Summit 2014: Big Data & Semantic Web Meet Applied Ontology

Track A: Common, Reusable Semantic Content

Session 1: "Use and Reuse of Semantic Content - The Problems and Efforts to Address Them"







January, 23, 2014

Some Introductory Comments on the Track Topic

Gary Berg-Cross, SOCoP

Outline of this Intro

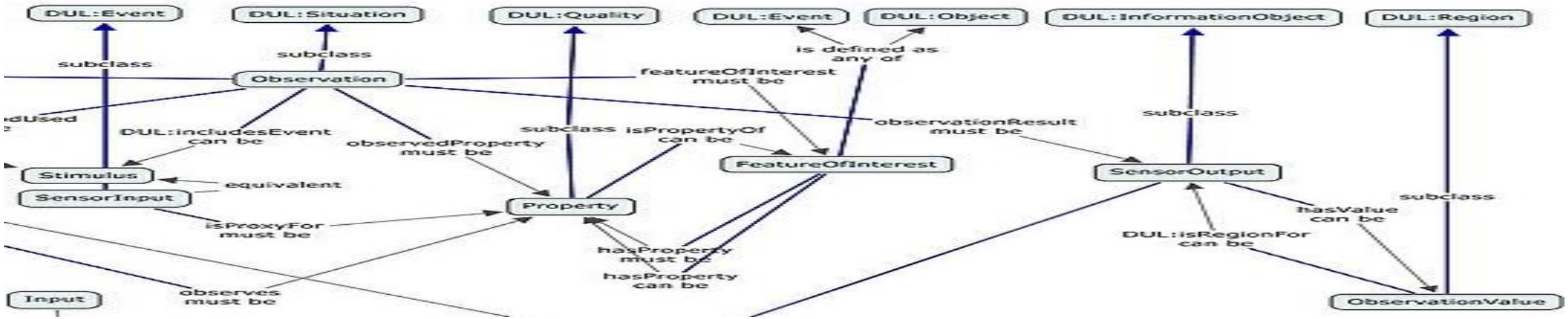
1. Topic relevance – a long history, Including Ontology Summits
2. Challenges Communicating
3. Line up with Big Data & Semantic Web & Services Issues/Challenges
4. Example from Hydrology
5. Example from EarthCube Semantic Manifesto
6. Lightweight Semantics, Methods (Ontology Design Patterns) & Enriched Schemas
7. Challenges for Reuse
8. Recap
9. Some References and Links

Historical Perspective: Ontological Building Blocks & Semantic Web

- Knowledge building has long been recognized as a bottleneck so K-reuse is very important and formalization of content as ontologies has been a way forward.
- "…the potential for achieving **semantic interoperability** across interconnected applications has become widely recognized….As this (SW) technology develops further, it will enable deployment of computer applications with increasing ability to make reliable knowledge-based decisions that currently require human effort. Programs with such enhanced capacity will increase the speed, efficiency, and sophistication of automated information analysis and exploitation…..
- The complementary technology for effectively representing the **semantic content of complex widely used concepts is also available**, but **agreement on standardized conceptual building blocks has not yet been reached**. "
  - **The UpperOntologySummit Joint Communiqué March 15, 2006**
    - **http://ontolog.cim3.net/cgi-bin/wiki.pl?UpperOntologySummit/UosJointCommunique**

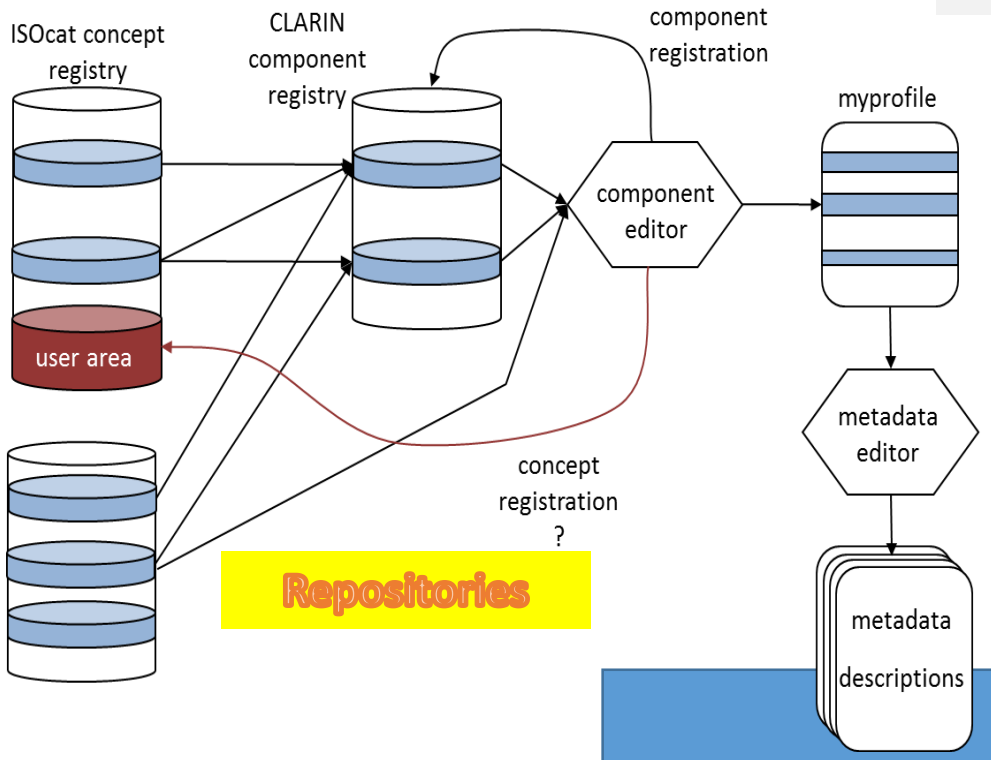
Sample Discussion Questions

1. What is an example of a small set of semantic content that the community might propose for reuse?
    1. Is there agreement on these or things like ODPs as building blocks?
2. What is an example of a large set that the community might propose for reuse?
3. Is it reasonable to expect reuse of an entire ontology like DOLCE and Semantic Sensor Network (SSN)?
    1. Under what conditions?
4. Is it better to expect alignment rather than exact content reuse?



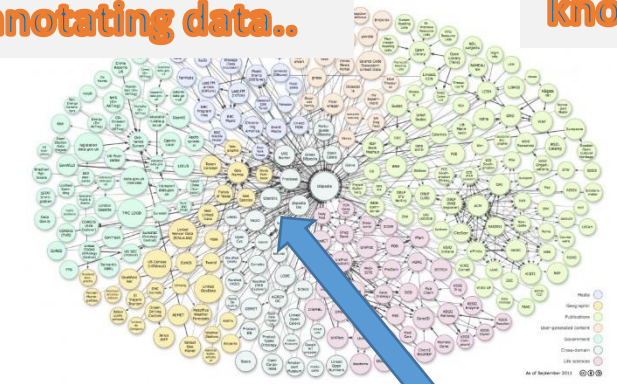
Community Priorities & Talking a Different Language

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



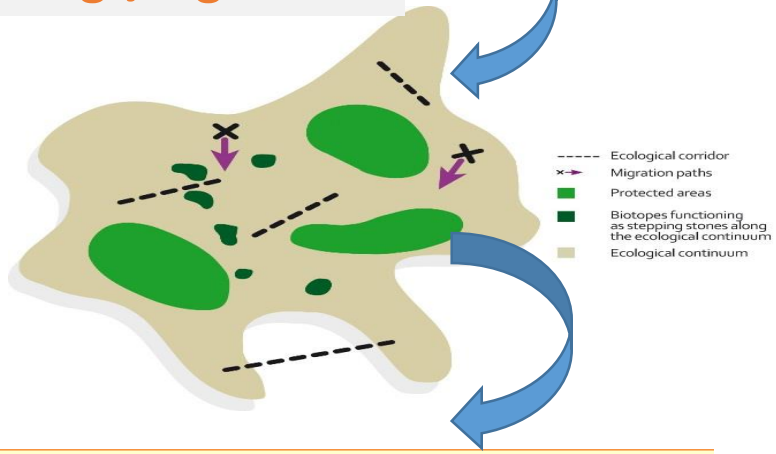
We need Faster Processing, better visualization, better data management plans, open data....

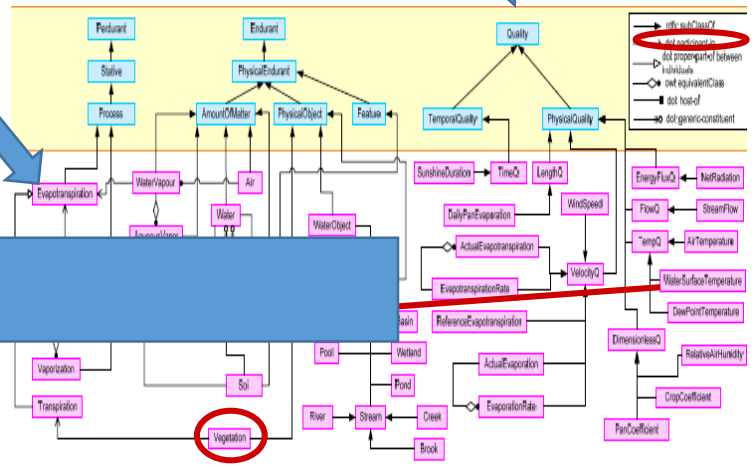
Converting RDB into triples, RDF Vocabulary for annotating data..



Logical Axioms, deep knowledge, alignment..

Geo Feature





Knowledge Infrastructure & Ontology applications:

- Smart Search,
- Sharing & Interoperability
- Semantic services
- Knowledge Infrastructure

Major Challenges - Big Data and LOD

Biomedical Big Data include:

- **Locating & liberating data** and software tools.
- Getting **access to the data** and software tools. (Discoverable)
- **Standardizing data and metadata**.
- **Extending policies and practices** for data and software sharing.
- **Organizing**, managing, and processing biomedical Big Data.
- **Developing new methods** for analyzing & integrating biomedical data.
- Training researchers who can use biomedical Big Data effectively.

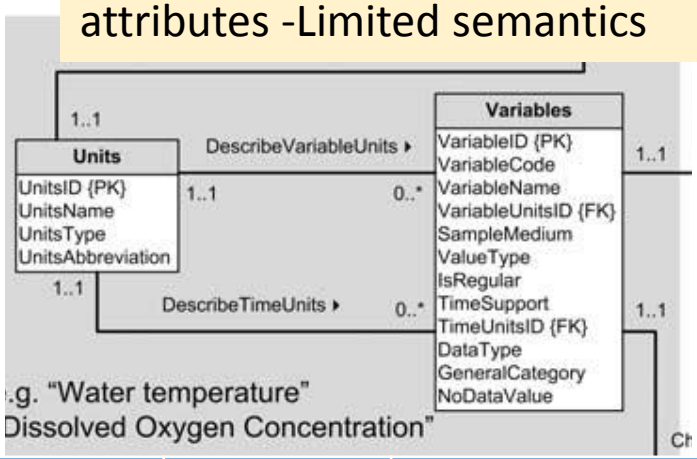
See http://bd2k.nih.gov/about_bd2k.html#sthash.ISpWsE4N.dpuf

- LOD is too complex/not rich enough.
- Too hard to master.
- Too few good tools.
- Needs **deep knowledge** and support of reasoning to fulfill its vision.
- Publishing linked data into a cloud does not ensure desired reusability.
  - Still needs better semantic relations (e.g. sameTypeAs,) **provenance**, quality, credit, attribution and methods to provide the *reproducibility* that enables validation of results.

Improved Semantic Content & its Representation helps with a number of these

Example fro Big Data Domain- Hydrology- Variables, Tags & "Ontology" Concepts

Observation DM uses RDB structure to
Integrate files & handle heterogeneity, Good MD attributes -Limited semantics



Concept ID	Concept Name	Ontology Layer
41	Chemical	1
42	Organic	2
43	PCBs	3
1001	Homolog Groups	4

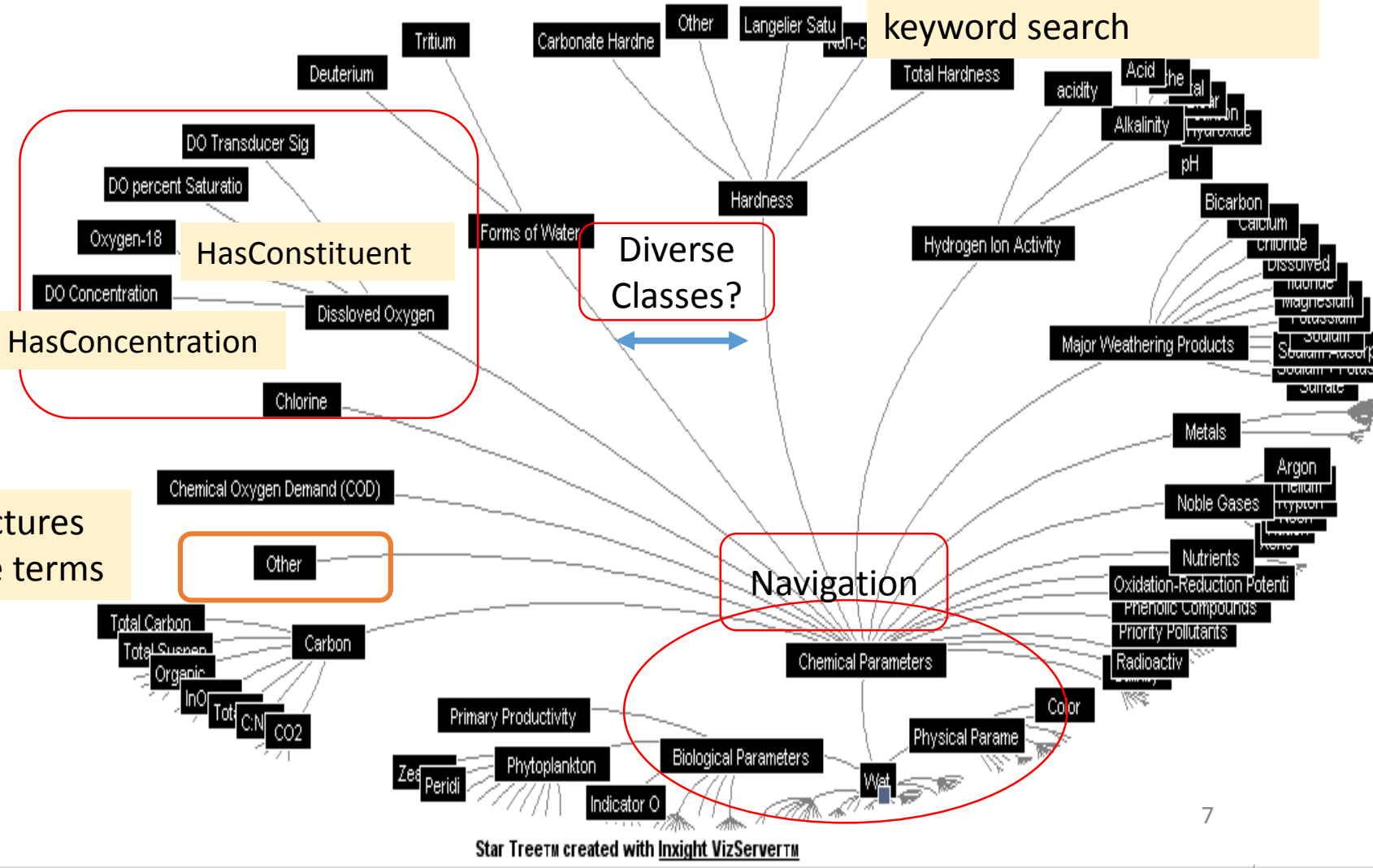
1220 Deca_Chloro_PCB 5

Classifier type structures
Connect to variable terms

CUAHSI Controlled Vocabulary

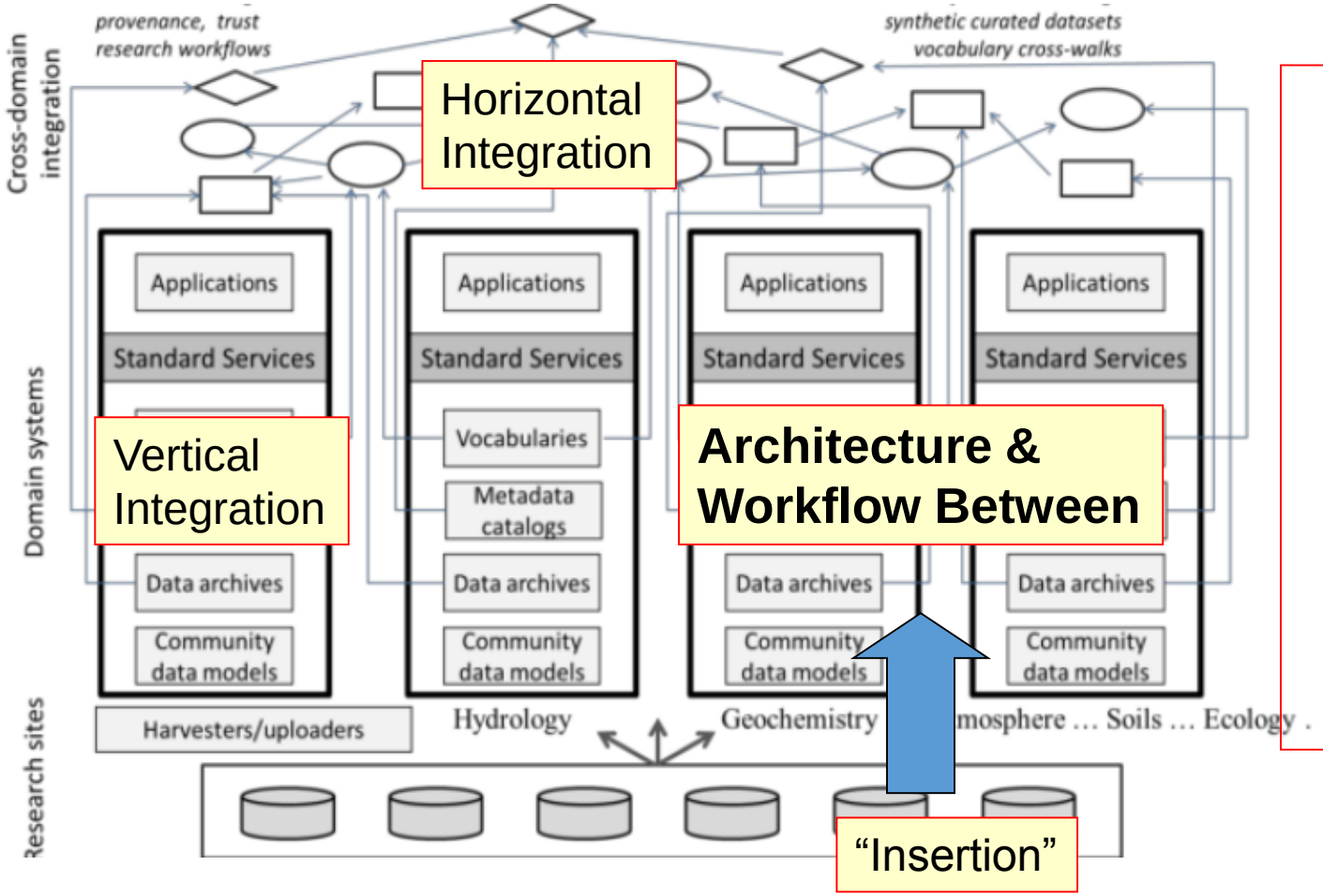
HydroTagger

Classifier type structures – for Navigation, tagging & keyword search



Graphic Overview of S/O (EarthCube) Manifesto

http://stko.geog.ucsb.edu/gibda2012/gibda2012_submission_6.pdf



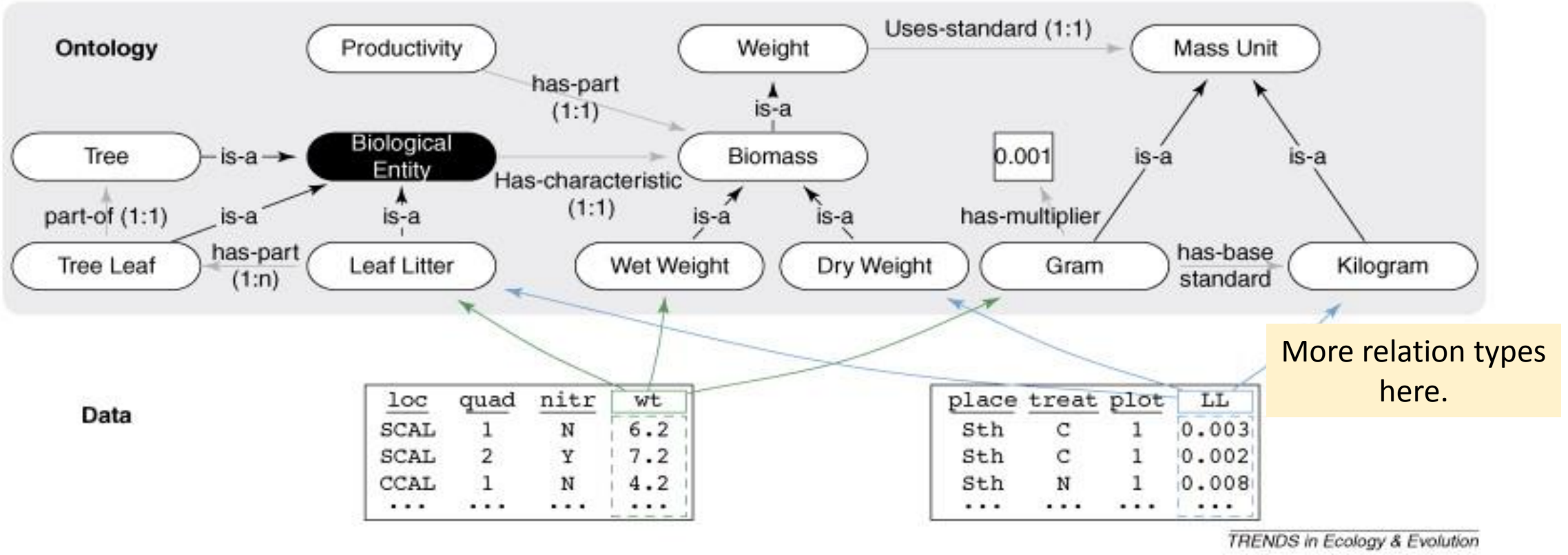
Guiding principles

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** → Community Understanding of Semantic role and value

Integrate with Lightweight Semantics (Top Down & Bottom Up)

- Low hanging fruit **leverages initial vocabularies** & existing **conceptual models** to ensure that a semantics-driven infrastructure is available for **early use.**
- Ontology Use can help handle heterogeneity



More relation types here.

Small modules are easier to deal with than large ontologies.

Adding Useful Relations Incrementally: Richer Schemata & Reusable Patterns

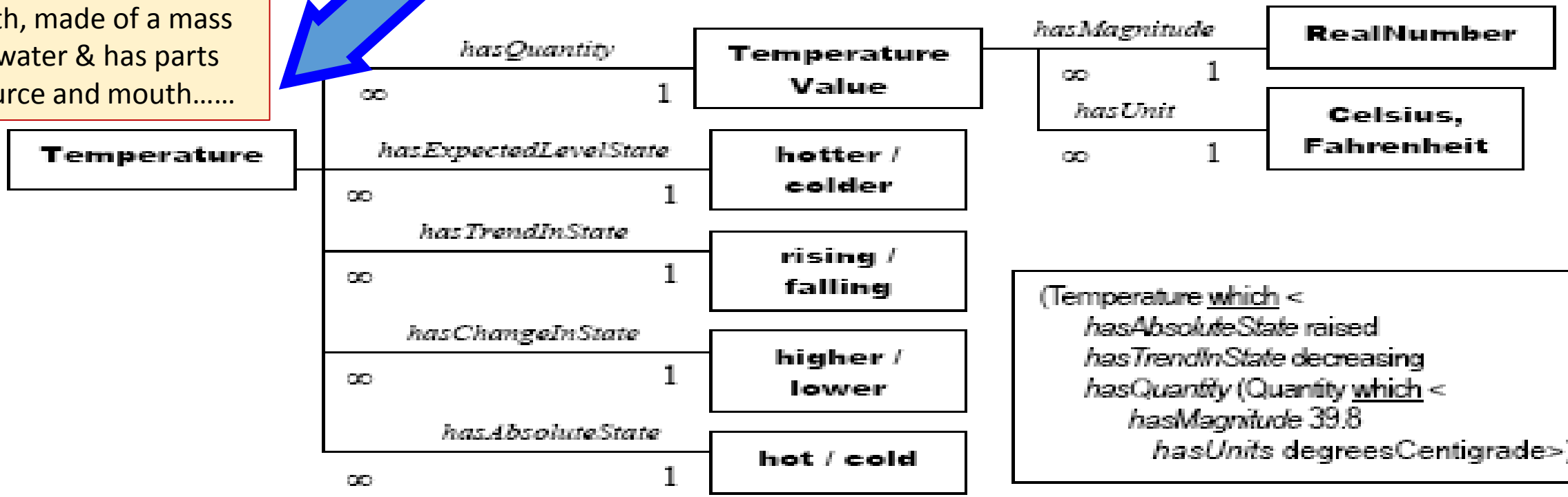
River, sub-surface water….

or height, salinity, acidity….

or salty, acidic….

Every River is a Water Body described by a path, made of a mass of water & has parts source and mouth……

Simple Feature-State Model (from GRAIL) becomes a richer schema



(Temperature which <
hasAbsoluteState raised
hasTrendInState decreasing
hasQuantity (Quantity which <
hasMagnitude 39.8
hasUnits degreesCentigrade>)>)

Example in GRAIL syntax

Example of Challenges – Semantic Mismatches, Inclusions & Alignments

Pragmatics of Intentions & goals

We have different goals so application & use are targeted. We need to adjust conceptualization to accommodate these.

Ontology level

- Different conceptualizations such as different class scope, Hierarchy level differences, coverage or granularity.
  - Scientists use different concepts & categories;
  - What does it mean to say that Concept P **includes** concept S?
  - What does it mean to say that concept P and S are **semantically close**?
  - Scientific understanding, often requires existing concepts to be revised or supplanted in the field
- Perspective – 4D vs. 3D, roads as straight lines or curves, time as interval or ratio…..
- Tacit assumptions

Language level

- Syntax and logical representation differences of the past should be handled by standardization & rule translations.
- Different expressivity (Owl vs. Common Logic) might be harder.

Recap

- There is a long history of interest and increasing work to leverage.
- There are problems in Big Data and Semantic Web/LOD work that quality semantic content can help with.
- But there remain challenges in reuse needing some foundational and practical work.
- Along with large & axiom rich domain and upper level ontologies, we should explore lightweight semantics & methods to provide easier entry.
- Opportunities exist in the Earth Sciences such as Hydrology and Ocean Science.
- We should keep in mind the challenges of communicating across the BD, SW and AO disciplines and projects.

Some References & Links

1. EarthCube Semantic Manifesto

Gary Berg-Cross, Isabel Cruz, Mike Dean, Tim Finin, Mark Gahegan, Pascal Hitzler, Hook Hua, Krzysztof Janowicz, Naicong Li, Philip Murphy, Bryce Nordgren, Leo Obrst, Mark Schildhauer, Amit Sheth, Krishna Sinha, Anne Thessen, Nancy Wiegand, and Ilya Zaslavsky

http://stko.geog.ucsb.edu/gibda2012/gibda2012_submission_6.pdf

2. CUAHSI www.cuahsi.org/
3. The Semantic Sensor Network Ontology

http://www.w3.org/2005/Incubator/ssn/wiki/Main_Page

4. UpperOntologySummit Joint Communiqué March 15, 2006
    - http://ontolog.cim3.net/cgi-bin/wiki.pl?UpperOntologySummit/UosJointCommunique
5. Hitzler, P., Janowicz, K., Berg-Cross, G., Obrst, L., Sheth, A., Finin, T., Cruz, I.: Semantic Aspects of EarthCube. Technical report, Semantics and Ontology Technical Committee. (2012)
   http://knoesis.wright.edu/faculty/pascal/pub/EC-SO-TC-Report-V1.0.pdf
6. Janowicz, K., Hitzler, P.: The Digital Earth as knowledge engine. Semantic Web Journal 3(3) (2012) 213–221
7. EarthCube http://www.nsf.gov/geo/earthcube/ and the community page at http://earthcube.ning.com/
8. VoCamps for ODPs http://vocamp.org/wiki/GeoVoCampDayton2012
9. Earth-Science-Ontolog Mini-Series http://ontolog.cim3.net/cgi-bin/wiki.pl?EarthScienceOntolog
10. S. Duce & K. Janowicz “Microtheories for Spatial Data Infrastructures” https://geog.ucsb.edu/~jano/duce_janowicz_microtheories_giscience2010.pdf
11. Christian Bizer: The Web of Linked Data (26/07/2009) http://en.wikipedia.org/wiki/Linked_Data Source
12. "Putting the Semantics in the Semantic Web: An overview of UIMA and its role in Accelerating the Semantic Revolution“ (Ferrucci 2006)

Supplementary

Track Questions & Related Issues Being Explored

1. How can we characterize or **measure semantic content reuse**, both between ontologies and by Big Data and Semantic Web communities?
2. What building blocks of common semantic content exists now to enable interoperability?
   - What additions are needed to move forward and how are these best achieved?
3. What is involved in reuse of Linked Data versus reuse of ontologies?
4. What is an example of a small set of semantic content that the community might propose for reuse?
   - Is there agreement on these or things like ODPs as building blocks?
5. What is an example of a large set that the community might propose for reuse?
6. Is it reasonable to expect reuse of an entire ontology like DOLCE and Semantic Sensor Network (SSN)?
   - If so under what conditions might this be reasonable?
   - Is it better to expect alignment rather than exact content reuse?
7. Is reuse about semantics alone or should it also address reasoning and data analytics?