Earth science ontologies: where do we go from here?

BENJAMIN ADAMS & MARK SCHILDHAUER*
(NCEAS, UCSB)

ONTOLOG SESSION 4, NOV. 1 2012

* Speaker

Introduction

- Already a diverse ecosystem of Earth Science vocabularies, thesauri, and ontologies
- How best to proceed to enable data discovery and integration within and across Earth Sciences?
- Some main considerations:
  - Scope and definitions of terms
  - Axiomatic structures
  - Governance
  - Intended uses and adoption

Many Vocabularies, Metadata, Ontologies

- Vocabularies
  - NASA global change master directory
  - LTER controlled vocabulary
  - AGI Thesaurus …
- Metadata standards
  - Infrastructure for Spatial Information in the European Community (INSPIRE)
    - http://inspire.jrc.ec.europa.eu/
  - GeoSciML (built on Geography Markup Language (GML))
    - See also SoTerML (Soil and Terrain Markup Language)
  - NetCDF CF climate forecasting metadata
  - Ecological Metadata Language EML …
- Ontologies
  - SWEET ontologies
  - Virtual Solar-Terrestrial Observatory (VSTO) ontology
  - Hydrologic Ontology for Discovery (CUAHSI-HIS) – tabular ontology (taxonomy)
  - Environmental Ontology (ENVO)
  - OGC O&M, and OBOE Extensible Observation Ontology …

Vocabularies, Metadata, Ontologies cont.

- This is just scratching the surface.
  - See, e.g., the collection of data standard references available through the Marine Metadata Interoperability website:
    - http://mmisw.org/
- We have a Babel of Earth Science data. Do we risk having a Babel of metadata and ontologies as well?

Scope and Definitions of terms

There already exist lots of terms, but we need terms that are actually "Enriched Concepts":

- With labels (names), definitions, & authorities
- "Naked terms" in an ontology / vocabulary not OK
  - E.g., what is "Litter" in the Sweet Ontology?
- Every term should be defined in natural language
- Every definition must have explicit source (authority)
  - Sources should be static reference to an unchanging authority (publication) or GUID – e.g. not a Wikipedia page URL
- Authority appropriate for science community

Axiomatic structures

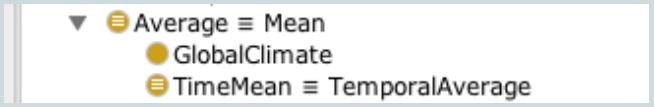
- Existing systems vary greatly in terms of their ability to facilitate types of automated reasoning
- Leverage varying capabilities of existing flat formats, SKOS, RDFS, OWL, etc.
- Linking to foundational ontologies with design patterns/principles might help:
  - Basic Formal Ontology (BFO)
  - Semanticscience Integrated Ontology (SIO)
    - OWL2 semantics

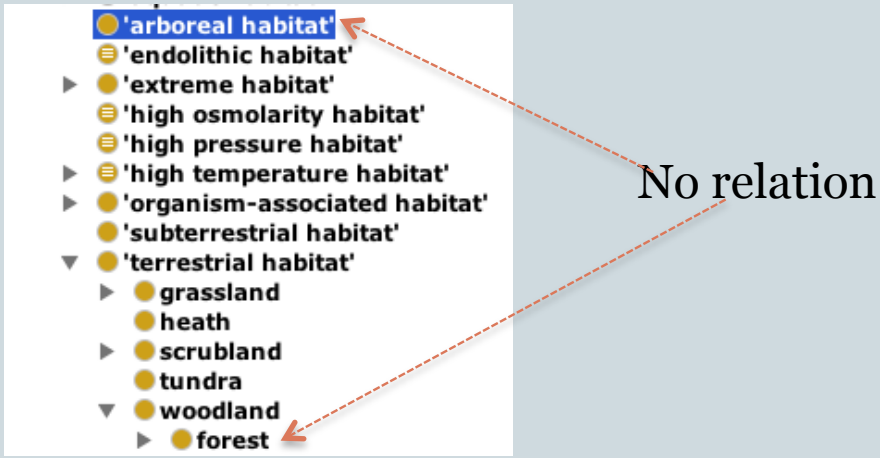
Ontologies preferred over SKOS?

- RDFS and OWL provide rich reasoning frameworks to work within
- Low hanging semantic fruit include better descriptions of how concepts are related:
  - Disjoint
  - Classes with more than one parent (not just taxonomies)
  - Property restrictions– domain & range constraints
  - Synonyms (a bit more controversial)

Evaluating ontologies

- Need better mechanisms for evaluating ontologies in terms of serving intended community of relevance.
  - I.e., feedback of working scientists, not ontology engineers
- Easy to open up existing ontologies and find questionable modeling





Governance issues

- What is the human process for contributing to the ontology?
  - suggesting new terms, correcting existing ones, revising, deprecating, versioning, curating, sustaining, etc …
- How do we maintain term provenance/ontogeny?
  - Term versioning and alignment
  - Terms with same definition and authority EQUALS synonym?
- Access to terms: free, open source or other?
  - E.g., can we get right-to-use for definitions from well-vetted but copyrighted, published sources such as AGI Thesaurus?

Intended Use

- Most ontologies, metadata languages are domain specific: a thematic scope, community of relevance, etc.== next-generation silo-ing
  - Intended for specific API – e.g., NetCDF CF
  - Intended for specific research community – e.g., ENVO (metagenomics)?
- But Earth and environmental science research often spans multiple domains.
- Ensuring that these representations can be integrated is important.

Conclusion

- Diversity of Earth Science terminology approaches is problematic for adoption and interoperability
- Is need for sustainable governance structure for developing and curating ontologies for the Earth Sciences
- Signal opportunity to collaborate if not coordinate on developing Earth Science ontologies
  - ONTOLOG, EarthCube, ESIP, MMI, DataONE, OGC, SONet, SOCoP, Phenoscape, VoCamp, EU efforts (INSPIRE?) etc. -- provide mechanisms to work together
- … or continue establishing a new Babel of Terms?