Observational data models as a unifying framework for biodiversity information

M. Schildhauer*[1], S. Bowers[2], M. Jones[1], Steve Kelling[3], Hilmar Lapp[4]

1. NCEAS, Santa Barbara, CA, USA
2. Gonzaga University, Spokane, WA, USA
3. Cornell University, NY, USA
4. NESCent, Durham, NC, USA

* Contact: schild@nceas.ucsb.edu

http://sonet.ecoinformatics.org

Acknowledgements: NSF OCI INTEROP 0753144

Growing challenges for biodiversity informatics

Investigations in biodiversity science often require integrating information from multiple scientific disciplines. While representing and organizing taxonomic names and concepts constitutes a significant challenge, there is also a critical need to integrate biodiversity information with relevant measurements and observations from other earth and life science domains. Observational data models show great promise for facilitating this integration.

Biodiversity use case

Investigator wants to explore relationship between biodiversity and ecosystem functioning (e.g. primary productivity) in forest trees. Data needs might include:

Vegetation plot information from repositories – Vegbank, Salvias, NVS --provide *in situ* association information, taxonomic, spatiotemporal, and other metadata (*VegX schema; references *EML and *Darwin Core/ABCD)

Vouchered specimen information from plots and supplementary collections (*Darwin Core/ABCD) enrich understanding of local associations and variation in forest composition

Functional trait information from e.g. LEDA/Salvias/TRY -- associated with taxonomic identities in plot data (*CNRS/TraitNet ontologies)

Phenotypic data from e.g. GenBank annotations (*GO, *EQ/PATO, *TO, *PO)

Phylogenetic relationships among taxa drawn from Treebase (*CDAO)

Climatic, geospatial, sensor data and in situ human observations from e.g. NCAR, NASA, misc independent researchers/citizens (* O&M, *SWE, *SWEET/VSTO, * EML)

* Represent de facto and emerging formalizations, as ontologies and other controlled vocabularies, of relevant concepts for interpreting how data are defined and inter-related

Utility of observational data models

Multiple communities within the earth and biological sciences are converging on the use of observational data models (e.g., ecology, evolution, oceanography, geosciences) to enable cross-disciplinary data discovery, interpretation and integration.

Observational data models provide a powerful, general, high level abstraction or "template" for describing a broad range of scientific data

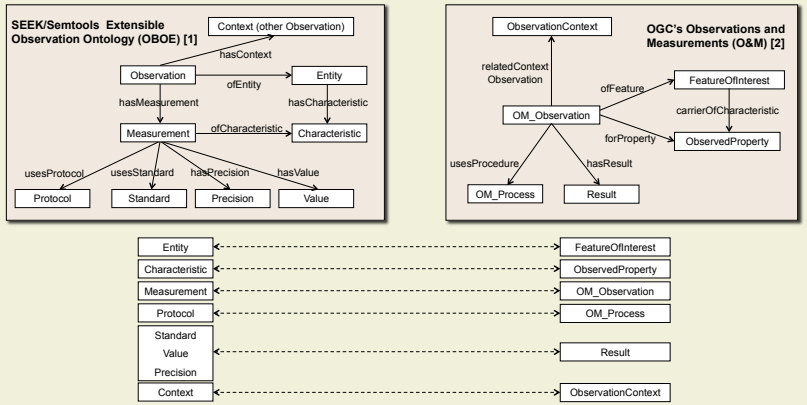
Controlled vocabularies can be linked to data through observational data models via *semantic annotation*, allowing for enhanced cross-disciplinary interpretation of scientific terminologies.

Reasoning capabilities such as hierarchy traversal, consistency checks, and equivalence determinations are enabled via semantic formats such as OWL (Web Ontology Language)

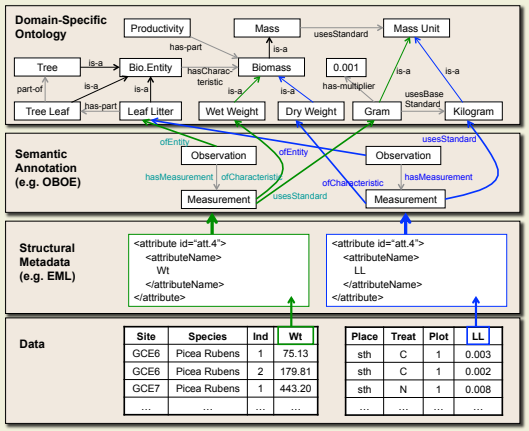
Semantic interoperability can be facilitated if observational data models and their controlled vocabularies are developed using compatible semantic and syntactic approaches

Collaborative development of observational data models is progressing through the "open" SONet effort (AKN, CUAHSI, OGC, SEEK/Semtools, SERONTO, TraitNet, VSTO), and the newly constituted "Joint Working Group on Observational Data Models and Semantics" (including SONet, Data Conservancy, Data ONE, and Phenoscape projects).

Similarities among Observational Data Models



Prototype Architecture for Applying Semantic Annotation



Site	Species	Ind	Wt	…
GCE6	Picea Rubens	1	75.13	…
GCE6	Picea Rubens	2	179.81	…
GCE7	Picea Rubens	1	443.20	…
…	…	…	…	…

(a) Dataset

observation "o1"
entity "Point_Location"
measurement "m1" key yes
Characteristic "GCE_Local-Code"
Standard "Nominal"
observation "o2"
entity "Tree"
measurement "m2" key yes
Characteristic "SpeciesName"
Standard "TaxonomicName"
measurement "m3" key yes
Characteristic "Local-ID"
Standard "Nominal"
measurement "m4"
Characteristic "Mass"
Standard "Ratio"
context identifying yes "o1"
map "Site" to "m1"
map "Species" to "m2"
map "Ind" to "m3"
Map "Wt" to "m4"

(b) Semantic annotation to dataset (a)

References:

[1] Shawn Bowers, Joshua S. Madin and Mark P. Schildhauer, A Conceptual Modeling Framework for Expressing Observational Data Semantics. In ER 2008, 41-54.

[2] OpenGIS observations and measurements encoding standard (O&M): http://www.opengeospatial.org/standards/om