Biodiversity Information Standards (TDWG)
http://www.tdwg.org

Observations Task Group Charter
A Task Group of Observations and Specimen Records Interest Group

1. Convenor

- Matthew Jones* (NCEAS, UC Santa Barbara)

2. Core Members

- Shawn Bowers* (UC Davis)
- Vishwas Chaven* (GBIF)
- Jerry Cooper* (Land Care Research)
- Philip Dibner* (OGCii)
- Steve Kelling* (Cornell, TDWG IG Convener)
- Jessie Kennedy* (Napier U)
- Josh Madin* (MacQuarie U)
- Éamonn Ó Tuama* (GBIF)
- Cynthia Parr* (U Maryland)
- Robert Peet* (U North Carolina)
- Mark Schildhauer* (NCEAS, UC Santa Barbara; *Scientific Observations Network)*
- Katharina Schleidt* (AlterNet)
- Dave Vieglais* (U Kansas)

3. Motivation

Broad-scale ecological studies often require information assembled from multiple disciplines including biology (e.g., genetics, physiology, and paleontology), the physical sciences (e.g., geography, meteorology, and hydrology), and increasingly the social sciences (e.g., economics and sociology). These 'synthesis' studies are becoming more common, and create new demands for seamless access to data from disparate disciplines. In such synthesis studies, data heterogeneity creates major informatics challenges that include the need to better discover, access, interpret, and integrate relevant data that have been collected by others. Some progress has been made using metadata standards (e.g., Ecological Metadata Language) and community-wide data networks, but these systems tend to leave the description of data semantics to natural-language fields that are difficult for computing systems to utilize.

To make further advances, ecology clearly needs robust, open, and generic software systems that address the semantic ambiguities of heterogeneous data. Observational data constitute the primary form of information that biodiversity and ecological scientists collect, and thus the Observation provides the best level of abstraction for modelling the semantics of these heterogeneous data. New strategies for managing observational data are needed so that systems can better interpret the semantics of observations to resolve issues impeding synthesis. For these systems to interoperate effectively, however, the scientific community must identify and agree on a core semantic model that can be used to accommodate and/or describe existing observational data representations. A community-sanctioned, core semantic model for ecological and environmental observational data is thus needed to enable interoperability among existing data resources, which will in turn provide the necessary foundation to support cross-disciplinary synthetic research in the environmental sciences. A core semantic model also provides the structure for linking to domain-specific scientific concepts, which can then be used in scientific applications.

4. Goals, Outputs, and Outcomes

The intended outcome of this working group is to derive consensus on modeling strategies for achieving observational data interoperability. To accomplish this goal we will solicit and follow recommendations from a broad community of researchers to develop a core semantic model for scientific observations onto which current and future data models can be mapped. Key to the success of the proposed core semantic model will be outreach to the broader environmental science communities and stakeholders. The goal will be to engage a diverse group of community members to contribute requirements and use cases, and provide feedback on proposed approaches, ultimately leading to ratification of a specification for observational data modeling.

5. Strategy

Our strategy will be to use the community resource tools available from TDWG to achieve these goals. We will use the TDWG standardization process to integrate existing efforts, such as the nascent Scientific Observations Network, a community-driven initiative to develop a core observational ontology, with other similar efforts (see History). The Scientific Observations Network will be working with several standards bodies, including TDWG, and possibly the OGC and W3C, and to achieve agreement among these groups on a common model.

6. Becoming Involved

Any TDWG member with an interest in organizing observational data can be involved, although there will be a concerted focus on maintaining relevance for biodiversity, ecological, and environmental information sources. Contact Matthew Jones (jones@nceas.ucsb.edu) for details.

7. History/Context

The Observations Task Group proposed here will directly build upon a recently held, NSF-funded workshop to discuss the various data models and ontologies used within the environmental sciences for managing observational data. Workshop participants included over twenty-five researchers, informatics specialists, and computer scientists representing various environmental-science disciplines, projects, and organizations. The models discussed by workshop participants included SEEK's Extensible Observation Ontology [MBS+07, OBOE], SPiRE's Evolutionary Trees and Natural History Ontology, NASA's Semantic Web for Earth and Environmental Terminology [SWEET], CUAHSI's Observations Data Model [ODM], LTER-Europe's Classes for Environmental Data Exchange [CEDEX], NatureServe's Observation Data Standard [ODS], OGC's model for Observations and Measurements, the TDWG Ontology [TAG], and the ontologies developed as part of the VSTO project, among others. A primary goal of this workshop was to determine whether there was sufficient accord among the various conceptualizations of “scientific observations” across groups to relate and unify these various models to determine overlap among existing observation models and to define requirements for a semantic observations data model and capabilities needed for broad-scale data interoperability. The workshop concluded that such accord among the observational models does exist and that a shared model for observational data would tremendously benefit scientists and application developers.

The workshop concluded that observational data are defined broadly as the outcomes of acts of measurement using particular protocols within the context of any objective scientific measurement activity. Examples include data from survey or monitoring efforts, controlled experiments, and sensor-derived measurements. In each case, the basic or atomic notion of an observation represents the outcome of some measurement taken of a defined attribute or characteristic of some “entity” (e.g., an organism “in the field”, a specimen, a sample, an experimental treatment, etc.), within some context (possibly given by other observations). Every observation entails the measurement of one or more properties of some real-world entity or phenomenon. As a result, a data model that focuses on the structure of observations can richly model the fundamental semantics of the scientific measurements that are being made. For example, a fundamental model of observations allows one to discover data based on the entities that were observed and on the context in which an observation was made [MBS+07]. In addition, such a model would enable the construction of a variety of data integration services that can mediate the wide variety of differences among observational data that might be relevant to a particular study. The NSF workshop on observational data modelling led to the formation of the Scientific Observations Network (SONet), which is seeking funding to pursue the standardization process for scientific observations data models and ontologies. The SONet initiative in turn led to this proposal for a TDWG Observations Task group.

8. Summary

The goal of this task group is to create a core semantic model for observational data in the ecological and environmental sciences. Holistic, integrative, and large-scale science would benefit from better data discovery, interpretation and integration both within and across disciplines that would be enabled from a common observational data model. This is necessary because there is an enormous quantity of observational data that is gathered by very diverse methodologies and stored in various formats. Additionally, a variety of observational data models have been developed, but none are sufficiently comprehensive.

It is essential that the characteristics of the data model accurately portray observational data via well defined terms that conform to existing standards. Thus, it is desirable to conduct this process within an established standards body such as TDWG. Since there is much commonality in observational data, the focus of this task group will be to develop a core observational data model that is domain independent, but extensible to be inclusive of specific requirements within a particular domain.

The development of the observational data model will be based on identifying the requirements needed for consumers and producers of the data model. The outcomes of this task group will provide data producers with tools and approaches that enable the design, creation, management, and publishing of observational data, and give consumers tools to more effectively discover, access, integrate, analyze, and report findings based on heterogeneous observational data sources.

9. Resources

- [MBS+07] Madin J, S Bowers, M Schildhauer, S Krivov, D Pennington, F Villa. (2007) An ontology for describing and synthesizing ecological observation data. *Ecological Informatics* 2 (3): 279-296. http://dx.doi.org/doi:10.1016/j.ecoinf.2007.05.004
- Observations Standards web site (http://obs.ecoinformatics.org/)
- Kelling Presentation at TDWG 2007 (http://www.tdwg.org/fileadmin/2007meeting/slides/Kelling_Obs%20Workshop_abs242.ppt)
- Ecological Metadata Language (http://knb.ecoinformatics.org/software/eml/)