



## Developing an Ontology for Ocean Biogeochemistry Data

Cynthia Chandler (1), Molly Allison (2), Robert Groman (2), Patrick West (3), Stephan Zednick (3), and Andrew Maffei (4)

(1) Woods Hole Oceanographic Institution, Marine Chemistry, BCO-DMO, Woods Hole, MA, United States (cchandler@whoi.edu), (2) Woods Hole Oceanographic Institution, Biology, BCO-DMO, Woods Hole, MA, United States, (3) Tetherless World Constellation, Rensselaer Polytechnic Institute, Troy, NY, United States, (4) Woods Hole Oceanographic Institution, Ocean Informatics Working Group, Woods Hole, MA, United States

A discipline as diverse as oceanography will benefit greatly when the research community meets the challenge of effectively incorporating Semantic Web technologies into the existing cyberinfrastructure. Semantically-enabled data delivery systems offer great promise for enabling new and better scientific research, but significant challenges must be met before the full potential can be realized. Evolving expectations for open access to research data combined with the complexity of global ecosystem science research themes present a significant challenge, and one that is best met through an informatics approach, wherein research scientists, information managers and computer scientists collaborate in small teams.

The Biological and Chemical Oceanography Data Management Office (BCO-DMO) is funded by the US National Science Foundation Division of Ocean Sciences to work with ocean biogeochemistry researchers in the US to improve access to data resulting from their respective programs. In an effort to improve data access, BCO-DMO staff members are collaborating with researchers from the Tetherless World Constellation (TWC at Rensselaer Polytechnic Institute) to develop an ontology that formally describes the concepts and relationships in the data managed by the BCO-DMO. The project required transforming a legacy system of human-readable, flat files of metadata to well-ordered controlled vocabularies, and finally to a fully developed ontology. To improve semantic interoperability, terms from the BCO-DMO controlled vocabularies are being mapped to controlled vocabulary terms adopted by SeaDataNet, the pan-European infrastructure for ocean and marine data management. Additionally, as part of their efforts to develop generic science ontologies, the team at TWC has facilitated the adoption of key concepts from the BCO-DMO ontology into ontologies developed for other science domains, and the adoption of concepts from other domains into the BCO-DMO ontology.

From the beginning of the project, development of the ontology has been guided by a use case based approach. The use cases were derived from data access related requests received from members of the research community served by the BCO-DMO. The resultant ontology represents the information stored in the metadata database and satisfies the requirements of the use cases. The BCO-DMO metadata database currently contains information that powers several different user and machine-to-machine interfaces to the BCO-DMO data repositories. One goal of the ontology development project is to enable subsequent development of semantically-enabled components (e.g. faceted search) to enhance the power of those interfaces and improve data access through enhanced data discovery.

In addition to sharing the ontology, we will describe the challenges encountered thus far in the project, the technologies currently being used, and the strategies associated with the use case based informatics approach.

URL: <http://bco-dmo.org/>