



Data Stewardship throughout the Ocean Research Data Life Cycle

Cynthia Chandler, Robert Groman, Molly Allison, Peter Wiebe, and David Glover

Woods Hole Oceanographic Institution, Marine Chemistry, BCO-DMO, Woods Hole, United States (cchandler@whoi.edu)

The Biological and Chemical Oceanography Data Management Office (BCO-DMO) works in partnership with ocean science investigators to publish data from research projects funded by the Biological and Chemical Oceanography Sections and the Office of Polar Programs Antarctic Organisms & Ecosystems Program (OPP ANT) at the U.S. National Science Foundation. Since 2006, researchers have been contributing data to the BCO-DMO data system, and it has developed into a rich repository of data from ocean, coastal and Great Lakes research programs. The end goals of the BCO-DMO are to ensure preservation of NSF funded project data and to provide open access to those data; achievement of those goals is attained through successful completion of a series of related phases.

BCO-DMO has developed an end-to-end data stewardship process that includes all phases of the data life cycle: (1) providing data management advice to investigators during the proposal writing stage; (2) registering their funded project at BCO-DMO; (3) adding data and supporting documentation to the BCO-DMO data repository; (4) providing geospatial and text-based data access systems that support data discovery, access, display, assessment, integration, and export of data resources; (5) exploring mechanisms for exchange of data with complementary repositories; (6) publication of data sets to provide publishers of the peer-reviewed literature with citable references (Digital Object Identifiers) and to encourage proper citation and attribution of data sets in the future and (7) submission of final data sets for preservation in the appropriate long-term data archive.

Strategic development of collaborative partnerships with complementary data management organizations is essential to sustainable coverage of the full data life cycle from research proposal through preservation of the final data products. Development and incorporation of controlled vocabularies, domain-specific ontologies and globally unique, persistent identifiers to unambiguously identify resources of interest curated by and available from BCO-DMO have significantly enabled progress toward interoperability with partner systems. Several important components have emerged from early collaborative relationships: (1) identifying a trusted authoritative source of complementary content and the appropriate contact; (2) determining the globally unique, persistent identifier for resources of interest and (3) negotiating the requisite syntactic and semantic exchange systems. An added benefit is the ability to use globally unique, persistent resource identifiers to identify and compare related content in other repositories, thus enabling us to improve the accuracy of content in the BCO-DMO data collection.

Results from a recent community discussion at the January 2013 Federation of Earth Science Information Partners (ESIP) meeting will be presented. Mindful of the NSF EarthCube initiative in the United States, the ESIP discussion was an effort to identify commonalities and differences in the way different communities meet the challenges of data stewardship throughout the full data life cycle and to determine any gaps that currently exist.

BCO-DMO: <http://bco-dmo.org>

ESIP: <http://esipfed.org/>