
Publishing and Citing Ocean Data

Lisa Raymond, WHOI, lraymond@whoi.edu (USA)

Cyndy Chandler, WHOI, cchandler@whoi.edu

Adam Leadbetter, BODC, alead@bodc.ac.uk

Ed Urban, SCOR, ed.urban@scor-int.org

Peter Pissierssens, IOC Project Office for IODE, p.pissierssens@unesco.org

Linda Pikula, NOAA, linda.pikula@noaa.gov

Data collected in the ocean sciences, whether generated from research or operational observations, are not always deposited in national or international repositories or data centers in a format that makes them retrievable and reusable. Often, there are insufficient incentives for data submission, resulting in low submission rates and even when submitted, a bare minimum of metadata. The problem is, in part, cultural. Publications of data do not carry the same weight in deciding promotion and tenure as papers that include analyses of the data. Promotion criteria do not take into account the innovation and complexity of data acquisition in challenging environments.

The advent of publisher and funding agency mandates to make data accessible is now the “stick” being used to force data accessibility. The “carrot” is the ability to cite the data accurately and to make re-use easier, making the openness advantageous to researchers. Scientists are now becoming aware that Digital Object Identifiers (DOIs) offer the means to easily cite their datasets and gain citation metrics.

The Use Cases

The Marine Biological Laboratory Woods Hole Oceanographic Institution (MBLWHOI) Library, the Scientific Committee on Oceanic Research (SCOR) and the International Oceanographic Data and Information Exchange (IODE) of the Intergovernmental Oceanographic Commission have developed and executed pilot projects related to two use cases: (1) data held by data centers are packaged and served in formats that can be cited and (2) data related to traditional journal articles are assigned persistent identifiers and stored in institutional repositories.

IODE has a history of fostering the establishment of standards and this collaboration is building a “community” of librarians, data managers and scientists to address the data publication paradigm.

The goal of the use cases has been to identify best practices for tracking data provenance and clearly attributing credit to data creators/providers so that researchers will make their data accessible. The assignment of persistent identifiers, specifically DOIs, enables accurate data citation.

Data Associated with Published Articles

The MBLWHOI Library has developed a workflow and metadata guidelines to deposit datasets in the Institutional Repository (IR), the Woods Hole Open Access Server (WHOAS). DOIs are assigned and registered with CrossRef. Libraries have been assigning DOIs to text documents for years and DOIs are now the de facto standard for datasets.

Elsevier Publishing sought collaboration with the MBLWHOI WHOI Library in 2011. The Library's ability to assign DOIs allowed a system to be set up that enables article records in ScienceDirect to display a banner that links to datasets deposited in WHOAS that are associated with Elsevier articles. The system works for DOIs assigned to datasets either before or after publication.

The Library is also collaborating with Biological and Chemical Oceanography Data Management Office (BCO-DMO) at the Woods Hole Oceanographic Institution. Tools and procedures have been developed to automate the ingestion of deposits from BCO-DMO. A Dublin Core compliant metadata record is deposited with a copy of each dataset into the WHOAS. The system also incorporates functionality for BCO-DMO to request a DOI from the Library. This partnership allows the Library to work with a trusted data repository to ensure high-quality data, while the data repository utilizes library services and is assured that a permanent archived copy of the data is associated with the persistent DOI.

Data Held by Data Centers

The Published Data Library (PDL) was recently implemented by the British Oceanographic Data Centre. It provides snapshots of specially chosen datasets that are archived using rigorous version management. The publication process exposes a fixed copy of an object and then manages that copy in such a way that it may be located and referred to over an indefinite period of time. Using metadata standards adopted across NERC's Environmental Data Centres, the repository assigns DOIs obtained from the British Library/DataCite to appropriate datasets.

Results

The project team has compiled a Cookbook (available online as IOC Manuals and Guides No. 64; <http://www.iode.org/mg64>) to enable other data centers and libraries to institute these services. The current version will be updated as more experience is gained. There will be a workshop with attendees who wish to start a data repository. Feedback from this class, and others who use the document, will help determine the weaknesses and strengths of the Cookbook. Further versions will be produced as appropriate. The intention is to share the Cookbook with the oceanographic community. It has already been distributed at an international meeting and to students at an e-Repository class at IODE. Response so far has been positive.

The British Oceanographic Data Centre and the MBLWHOI Library are actively assigning DOIs to appropriate datasets enabling accurate citation and preservation. This ensures accessibility to the broader community and provides an incentive to the original researcher to make the data available. Repositories that assign DOIs to data objects are providing services not yet provided by most large national or subject data repositories. These repositories are not meant to replace traditional data centers, but rather complement the data center by providing formal data publication services. Further, this project demonstrates that data publication that enables accurate citation of data sets is manageable for organizations of any size.