Emerging Role of Librarians in Data Publication

Lisa Raymond  
*Woods Hole Oceanographic Institution*

Cyndy Chandler  
*Woods Hole Oceanographic Institution*

Roy Lowry  
*British Oceanographic Data Centre*

See next page for additional authors

Follow this and additional works at: https://escholarship.umassmed.edu/escience_symposium

Part of the Library and Information Science Commons

Repository Citation


This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
Presenter Information
Lisa Raymond, Cyndy Chandler, Roy Lowry, Ed Urban, Gwenaelle Moncoiffe, Peter Pissierssens, Cathy Norton, and Holly Miller

Keywords
Data publication, Data citation, Data management

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.

Rights and Permissions
Copyright the Author(s)

This poster is available at eScholarship@UMMS: https://escholarship.umassmed.edu/escience_symposium/2012/posters/5
Emerging Role of Librarians in Data Publication
Lisa Raymond¹, Cyndy Chandler¹, Roy Lowry², Ed Urban³, Gwenaelle Moncoiffé², Peter Pissierssens⁴, Cathy Norton⁵, Holly Miller⁵
1. Woods Hole Oceanographic Institution (WHOI); 2. British Oceanographic Data Centre (BODC); 3. Scientific Committee on Oceanic Research (SCOR); 4. BCO-DMO; 5. IOC Project Office of IODE; 6. Marine Biological Laboratory (MBL)

Purpose
This poster demonstrates the procedures and tools developed to deposit datasets in an Institutional Repository (IR) and assign Digital Object Identifiers (DOIs).

Description
Current literature on the topic of data publication suggests that success is best achieved when there is a partnership between scientists, data managers, and librarians. The Marine Biological Laboratory/Woods Hole Oceanographic Institution (MBLWHOI) and the British Oceanographic Data Centre (BODC) have developed tools and processes to automate the ingestion of metadata from BCO-DMO for deposit with datasets into the Institutional Repository (IR) at Woods Hole Open Access Server (WHOAS) on the DSpace platform. The system also incorporates functionality for BCO-DMO to request a Digital Object Identifier (DOI) from the Library. This partnership allows the Library to work with a trusted data repository to ensure the data while the data are exchangeable, re-useable, and citable (by deposition in an accredited data center), through to the preparation of analysis data subsets for the purpose of scientific investigation and communication, data subsets which must in turn be openly accessible and citable for the purpose of traceability.

Method

Basic Workflow
1. A dataset is submitted by a scientist to BCO-DMO.
2. BCO-DMO staff enter metadata for the dataset in their database.
3. BCO-DMO staff store the dataset itself in their repository.
4. At some later time, a scientist contacts BCO-DMO with a request to cite a dataset.
5. BCO-DMO staff trigger a submission to WHOAS via a special 'submit to WHOAS and request DOI' button in BCO-DMO’s web interface.
6. BCO-DMO’s database system pulls data and metadata from its database and assembles a METS package.
7. BCO-DMO's database system submits the METS package to WHOAS via SWORD.
8. WHOAS ingests the metadata and returns a handle URL to BCO-DMO’s system.
9. BCO-DMO’s system parses the handle URL and generates a corresponding DOI.
10. BCO-DMO staff sends the DOI to the scientist who made the citation request.
11. After a short period following the submission, WHOAS notifies the BCO-DMO manager via email.
12. The WHOAS manager creates a DOI for the submitted dataset record.

Conclusions
The assignment of persistent identifiers enables accurate data citation. The Library can assign a DOI to appropriate datasets deposited in WHOAS. A primary activity is working with authors to deposit datasets associated with published articles. The DOI would ideally be assigned before submission and be included in the published paper so readers can link directly to the dataset, but DOIs are also being assigned to datasets related to articles after publication. WHOAS metadata records link the article to the datasets and the datasets to the article.

In addition, the Library has worked with researchers to deposit datasets in WHOAS that are not appropriate for national, international, or domain specific data repositories. These datasets currently include audio, text and image files.

Related Websites
PublishedOceanData http://publishedoceandata.net/
Woods Hole Open Access Server https://darchv.mblwholib.org/
Biological and Chemical Oceanography Data Management Office http://www.bco-dmo.org/
British Oceanographic Data Centre www.bodc.ac.uk
Information about the project, including reports from project meetings, can be found at http://www.iode.org/datapublishing

Acknowledgments
Funding provided by the Jewett Foundation
Ann Devenish, WHOAS Manager
Tobias Work, BCO-DMO Programmer
Alex Dorak, Library Programmer

Diagram showing the evolution of a scientific dataset through its different stages from raw data collection (fieldwork) to the generation of master files which must be preserved from accidental loss, self-described, exchangeable, re-useable, and citable (by deposition in an accredited data center), through to the preparation of analysis data subsets for the purpose of scientific investigation and communication, data subsets which must in turn be openly accessible and citable for the purpose of traceability.

SWORD METS Package
Simple Web-service Offering Repository Deposit (SWORD)
Metadata Encoding and Transmission Standard (METS)

- METS is a standard for encoding and packaging digital objects and metadata.
- SWORD is a lightweight protocol for depositing content from one location to another.
- SWORD submissions to DSpace involve creating a particular type of METS package that conforms to the DSpace METS Document Profile for Submission Information Packages (SIP) https://wiki.duraspace.org/display/DSpace/METSProfile.
- A METS package formatted per this profile is a zip file which contains these items: digital objects (documents, images, video, etc.) a mets.xml file that contains metadata about the digital objects this digital objects can be in any form.
- See the mets.zip file for an example of a valid METS package.
- The mets.xml file must conform to a specification defined in the DSpace METS profile.

The BCO-DMO submission interface:
The contents of a METS package:
The result of a successful submission in WHOAS: