

SCOR/IODE/MBLWHOI Library Collaboration on Data Publication

Lisa Raymond
Woods Hole Oceanographic Institution
360 Woods Hole Road, MS#8
Woods Hole, MA 02543
United States
lraymond@whoi.edu

Linda Pikula
NOAA Central Library
4301 Rickenbacker Causeway
Miami, FL 33419
United States
Linda.Pikula@noaa.gov

Roy Lowry
British Oceanographic Data Centre
6 Brownlow Street
Liverpool L3 5DA
United Kingdom
rkl@bodc.ac.uk

Ed Urban
Scientific Committee on Oceanic
Research
University of Delaware
Newark, DE 19716
United States
ed.urban@scor-int.org

Gwenaëlle Moncoiffé
British Oceanographic Data Centre
6 Brownlow Street
Liverpool L3 5DA
United Kingdom
gmon@bodc.ac.uk

Peter Pissierssens
UNESCO/IOC Project Office for IODE
Wandelaarkaai 7/61
8400 Ostende,
Belgium
p.pissierssens@unesco.org

Cathy Norton
Marine Biological Laboratory
7 MBL Street
Woods Hole, MA 02543
United States
cnorton@mbledu

ABSTRACT

This poster describes the development of international standards to publish oceanographic datasets. Research areas include the assignment of persistent identifiers, tracking provenance, linking datasets to publications, attributing credit to data providers, and best practices for the physical composition and semantic description of the content.

Categories and Subject Descriptors

H.3.7 [Information Storage and Retrieval]: Digital Libraries
collection, standards, user issues.

General Terms

Design, Standardization

Keywords

data repositories, data citation, persistent identifiers

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference '10, Month 1–2, 2010, City, State, Country.
Copyright 2010 ACM 1-58113-000-0/00/0010...\$10.00.

1. Introduction

Motivated by publisher and funding agency mandates, and a desire to properly attribute data sets to originating investigators, 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 of UNESCO (IOC) are working together to develop and execute 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 referred to in the articles and stored in institutional repositories, such as DSpace. IODE has a history of fostering the establishment of standards and this collaboration is building a "community" of Oceanographic Data and Information Exchange scientists to address the data publication paradigm. Initial publisher inquiries about this research project include Pangaea/Elsevier [5], BioMed Central, Institute of Physics, and Dryad.

2. Use Cases

The primary issue under investigation in use case (1) is how to express the continuous nature of data dynamically managed in a relational database management system in the quantized manner required for citation. The group that is managing this use case, the British Oceanographic Data Centre is also working to document best practice for the physical composition (e.g. file formats) and semantic description of the content of such snapshots to ensure

confident re-use of the data in decades to come. IODE is establishing a repository, PublishedOceanData that will serve these datasets. PublishedOceanData is an e-repository setup by IODE as a parallel repository to its OceanDocs system and targeted at datasets rather than documents. It was implemented by the IODE to support pilot projects undertaken by a joint IODE/SCOR/MBLWHOI Library working group on data publication and data citation. Its aim is to offer an open access e-repository facility where datasets snapshots labeled with Digital Object Identifiers (DOIs) and published by IODE's National Oceanographic Data Centres are preserved indefinitely in a digital library infrastructure.

The goal of use case (2) is to identify best practices for tracking data provenance and clearly attributing credit to data collectors/providers for data published in journal articles. To improve efficacy of data directly associated with a scientific article those data must be discoverable, citable and freely available on the Internet. Resources, standards, and workflows must be defined to support publisher and funding agency mandates, while being manageable by the scientists who are publishing their data. "In most fields, the rewards come from publication, not from data management" [1, p.7]. For the data to be discoverable, appropriate metadata, defined using community-accepted metadata standards, must be associated with the data source. Data will be made citable by the assignment of a persistent identifier as well as provenance and attribution metadata. The availability of the data will be assured by submission to a data repository that has stability and permanence. "There are three primary (and related) motivations for developing a robust data curation infrastructure: enabling new discoveries by exposing data for use in data-driven research, ensuring access to and preservation of scholarly output, and meeting existing or forthcoming requirements of funding agencies or institutions regarding data management, retention, and access. Libraries have demonstrated expertise in several areas that could be productively applied to the practice of data curation, and in some cases, cyberinfrastructure development" [3, p.8]. The MBLWHOI Library is developing Use Cases [6] to test the concept of depositing data associated with published articles and assigning DOIs to the datasets in our Institutional Repository (IR), the Woods Hole Open Access Server (WHOAS). The IR (<https://darchive.mblwhoilibrary.org/>) is built on the DSpace platform and metadata and procedures used to deal with text documents have been adapted to expand functionality to datasets. Datasets are linked to the article metadata record and the article metadata record contains links to the associated datasets. The repository system meets the stated requirements to make the data accessible, citable and permanent. This process is not meant to replace data submission to appropriate national, international or domain specific data centers.

This presentation reports on the progress of the MBLWHOI Library project to deposit data and assign persistent identifiers in

the Woods Hole Open Access Server (WHOAS), as well as the development of IODE's PublishedOceanData repository (<http://www.publishedoceandata.org/>). We will also report on the strategies for meeting the cultural and technical challenges. There are many good reasons for scientists to publish data, but also many reasons given for not making their data publicly available online [2]. "Incentives for researchers to share their data include the ethos of open science and peer review; the value of collaborating with others, for which data may be the "glue;" benefits to reputation; and reciprocity" [1, p.7]. It is important to note that data centers, publishers, libraries, and an intergovernmental, international oceanographic data exchange which facilitates the setting of international standards are working together to test the processes, metadata, and policies [4].

3. ACKNOWLEDGMENTS

Funding provided by the George Frederick Jewett Foundation.

4. REFERENCES

- [1] Borgman, Christine L. 2010. *Research Data: Who will share what, with whom, when, and why?* Presented at the China-North American Library Conference, 17 Aug 2010. <http://works.bepress.com/cgi/viewcontent.cgi?article=1237&context=borgman>
- [2] Costello, M. J. 2009. Motivating Online Publication of Data. *BioScience*, 59,5 (May 2009),418-427. DOI=<http://dx.doi.org/10.1525/bio.2009.59.5.9>
- [3] *Digital Research Data Curation: Overview of Issues, Current Activities, and Opportunities for the Cornell University Library* A report of the CUL Data Working Group. Cornell University, Ithaca, NY, 2008. http://ecommons.library.cornell.edu/bitstream/1813/10903/1/DaWG_WP_final.pdf
- [4] *SCOR/IODE/MBLWHOI Library Workshop on Data Publication*, UNESCO Headquarters, Paris, France, 2 April 2010. IOC Workshop Report No. 230. UNESCO, Paris, 2010. <http://www.scor-int.org/Publications/wr230.pdf>
- [5] *SCOR/IODE Workshop on Data Publishing*, Oostende, Belgium, 17-19 June 2008. IOC Workshop Report No. 207. UNESCO, Paris, 2008. <http://www.scor-int.org/Publications/wr207.pdf>
- [6] *A Woods Hole Data Repository: Addressing the Issues of Provenance, Attribution, Citation, and Accessibility*. Project Report. MBLWHOI Library, Woods Hole, MA, 2010. <http://tw.rpi.edu/proj/portal.wiki/images/3/3b/JewettSummary.pdf>