

Publishing and Citing Ocean Data

Lisa Raymond¹, Cynthia L. Chandler¹, Adam Leadbetter², Ed Urban³, Peter Pissierssens⁴, Linda Pikula⁵

1. Woods Hole Oceanographic Institution (WHOI) 2. British Oceanographic Data Centre (BODC) 3. Scientific Committee on Oceanic Research (SCOR) 4. IODE Project Office of IOC 5. NOAA Central Library

Introduction

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 a requirement that is forcing scientists to make their data available. However, the ability to cite the data accurately and to make re-use easier, is 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.

Data Publication

The Scientific Committee on Oceanic Research (SCOR), the Marine Biological Laboratory/Woods Hole Oceanographic Institution (MBLWHOI) Library, and the International Oceanographic Data and Information Exchange (IODE) of the Intergovernmental Oceanographic Commission have assembled a team of librarians, data managers, and scientists who are collaborating to identify best practices for tracking data provenance and clearly attributing credit to data collectors/providers.

A key element in the data publication process is the assignment of persistent identifiers to enable access to data sets and accurate data citation. The MBLWHOI Library can assign a DOI to appropriate datasets deposited in the Institutional Repository (IR) Woods Hole Open Access Server (WHOAS) by scientists from the participating institutions. The DOI is ideally assigned before submission of the article in which the data are used, so the DOI can be included in the published paper and readers can link directly to the dataset. However, DOIs are also being assigned to datasets that support papers that have already been published. WHOAS metadata records link the article to the datasets and the datasets to the article.

Data Associated with an Article

1 Deposit Data and Metadata

The screenshot shows the BCO-DMO MapServer Geospatial Interface. A map of the Gulf of Maine is displayed with various sampling locations marked. A sidebar on the right lists datasets, including 'Bottle Data | BCO-DMO'. Below the map, there is a 'Dataset: Bottle Data' section with details such as 'Project: Investigations of *Alexandrium fundyense* dynamics in the Gulf of Maine', 'Principal Investigator: Dennis McGillicuddy', and 'Data version: 21 March 2013'. A 'Current state' section indicates 'Final no updates expected'.

2 DOI Assigned

This block contains two screenshots. The left screenshot shows the WHOAS (Woods Hole Open Access Server) interface for a dataset titled 'Bottle Data collected from R/V Okeanos, R/V Tioga and R/V Endeavor cruises from the Gulf of Maine, Cape Cod Bay and Georges Bank from May 28, 2003 to August 04, 2010'. It lists metadata such as contributor, coverage, date, and description. The right screenshot shows a ScienceDirect article titled 'Deep Sea Research Part II: To in Oceanography' with a citation to the dataset: 'Georges Bank: A leaky incubator of *Alexandrium fundyense* blooms' by D.J. McGillicuddy et al. (2012). A DOI is highlighted in the citation: 'doi:10.1016/j.dsr2.2012.11.002'.

3 Dataset can be cited

This screenshot shows the WHOAS interface for the dataset 'Georges Bank: a leaky incubator of *Alexandrium fundyense* blooms'. It provides detailed metadata including title, author (D.J. McGillicuddy, D.W. Townsend, B.A. Kester, M.A. Anderson, Donald M.), abstract, description, and citation information. A DOI is highlighted: 'doi:10.1016/j.dsr2.2012.11.002'.

Data Held by Data Centers

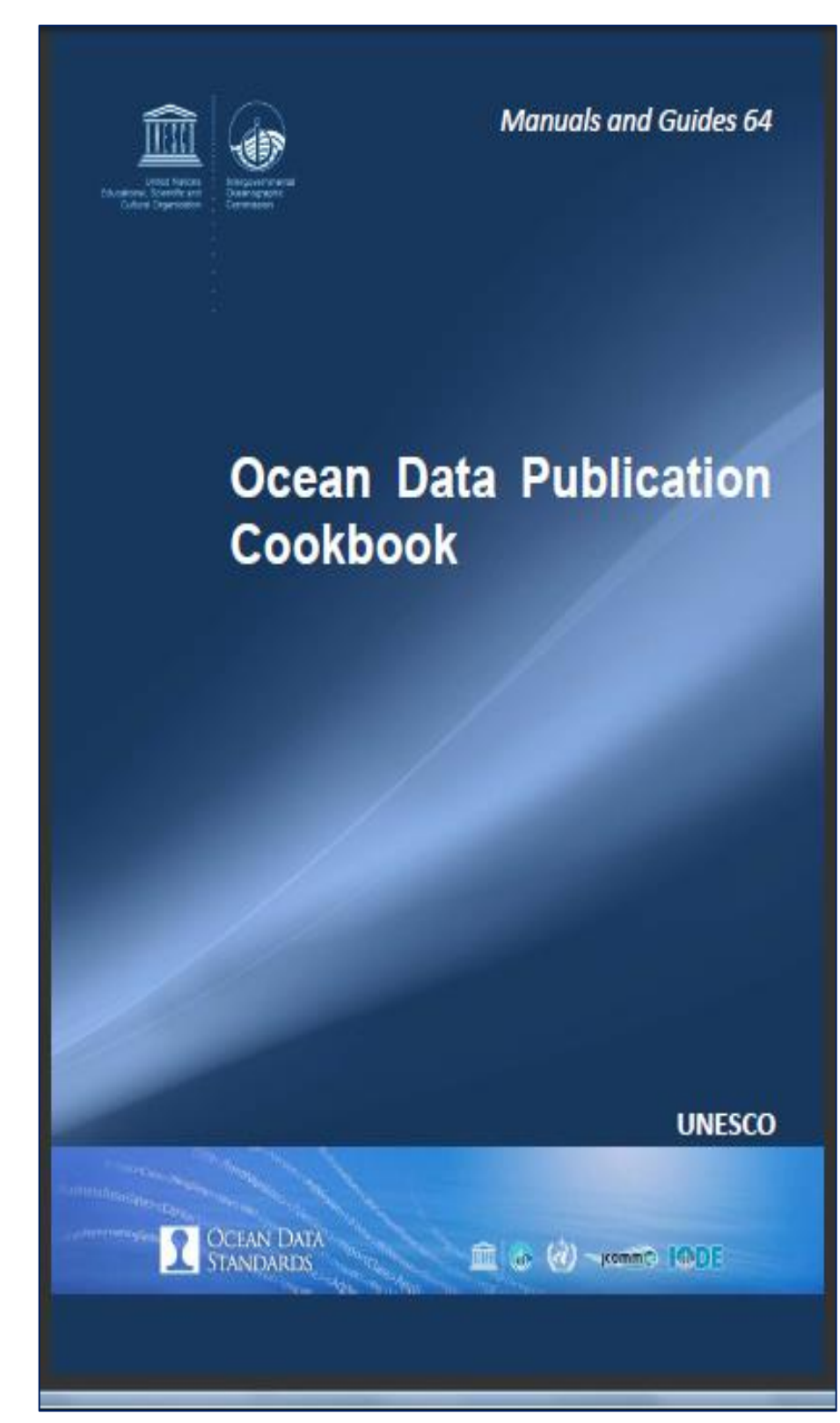
The Published Data Library (PDL) was recently implemented by the British Oceanographic Data Centre. It provides snapshots of specially chosen datasets that are archived using rigorous version management. The publication process exposes a fixed copy of an object and then manages that copy in such a way 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.

The screenshot shows the Published Data Library (PDL) interface for the dataset 'AMT14 Rates of nitrogen fixation in surface waters'. It lists the dataset title, creator (Dr. A. R. Rees), subject (Biological Sciences), and publication date (17th January 2012). A DOI is highlighted: 'doi:10.5285/41479c42-4dfb-4da9-be97-4c532ce13922'.

Results

The project team has compiled a Cookbook (available online as IOC Manuals and Guides No. 64) to enable other data centers and libraries to institute these services. Feedback from those who use the Cookbook, will help determine the weaknesses and strengths of the document, and the online version will be updated as more experience is gained.

WHOAS is now included in the Thomson Reuters Data Citation Index, providing citation metrics and credit for the creation of research data.



iodo.org/mg64

Related Websites

Woods Hole Open Access Server <https://darchive.mblwhoilibrary.org/>
Biological and Chemical Oceanography Data Management Office <http://www.bco-dmo.org/>
BODC Published Data Library https://www.bodc.ac.uk/data/published_data_library/
Information about the project, including reports from project meetings, can be found at <http://www.ioode.org/datapublishing>.

Presented at IMDIS 2013
Lucca, Italy, September 23-25, 2013

