

FORUM

A New Approach to Data Publication in Ocean Sciences

PAGE 484

Data are collected from ocean sciences activities that range from a single investigator working in a laboratory to large teams of scientists cooperating on big, multinational, global ocean research projects. What these activities have in common is that all result in data, some of which are used as the basis for publications in peer-reviewed journals.

However, two major problems regarding data remain. First, many data valuable for understanding ocean physics, chemistry, geology, biology, and how the oceans operate in the Earth system are never archived or made accessible to other scientists. Data underlying traditional journal articles are often difficult to obtain. Second, when scientists do contribute data to databases, their data become freely available, with little acknowledgment and no contribution to their career advancement. To address these problems, stronger ties must be made between data repositories and academic journals, and a “digital backbone” needs to be created for data related to journal publications.

Links Between Data Repositories and Academic Journals

The Scientific Committee on Oceanic Research (SCOR) and the International Oceanographic Data and Information Exchange (IODE) of the United Nations Educational, Scientific and Cultural Organization's Intergovernmental Oceanographic Commission (IOC) are discussing how to provide better access to ocean data through increased submission to approved, open, online resources. Such new infrastructure and new approaches to data publication could help scientists who observe the ocean and model its processes. Most important, it is now timely to

- increase the availability of data used to create figures, tables, and statistical analyses in traditional journal articles;
- reinforce linkages between data lodged in data centers and science publications, particularly “data briefs”; and
- encourage the publishing of journals that specialize in “data publications” or “data briefs.”

Data publications are short descriptions (as short as a few paragraphs of text), not interpretations, of data sets. They provide persistent pointers to the data in an approved data repository as well as

references citable in papers that use the data, and in authors' curricula vitae.

Getting Journals on Board

Several journals in the ocean sciences already welcome the publication of data briefs. They include *Marine Micropaleontology*; *Geochemistry*, *Geophysics*, *Geosystems*;

Ecological Archives; and *Earth System Science Data*.

Other journals also acknowledge the benefits of submitting the data underlying traditional publications to approved databases. In 1993, AGU first established its “Policy on Referencing Data in and Archiving Data for AGU Publications” (see http://www.agu.org/pubs/policies/data_policy.shtml). The policy emphasizes the importance for authors to submit data that are the basis for their papers to a recognized data archive. It also states AGU's commitment to ensuring the long-term archiving and protection of data. Data sets associated with articles are available at http://www.agu.org/pubs/esupp_browse.html, and access to these data does not require membership in AGU or subscription to an AGU journal.

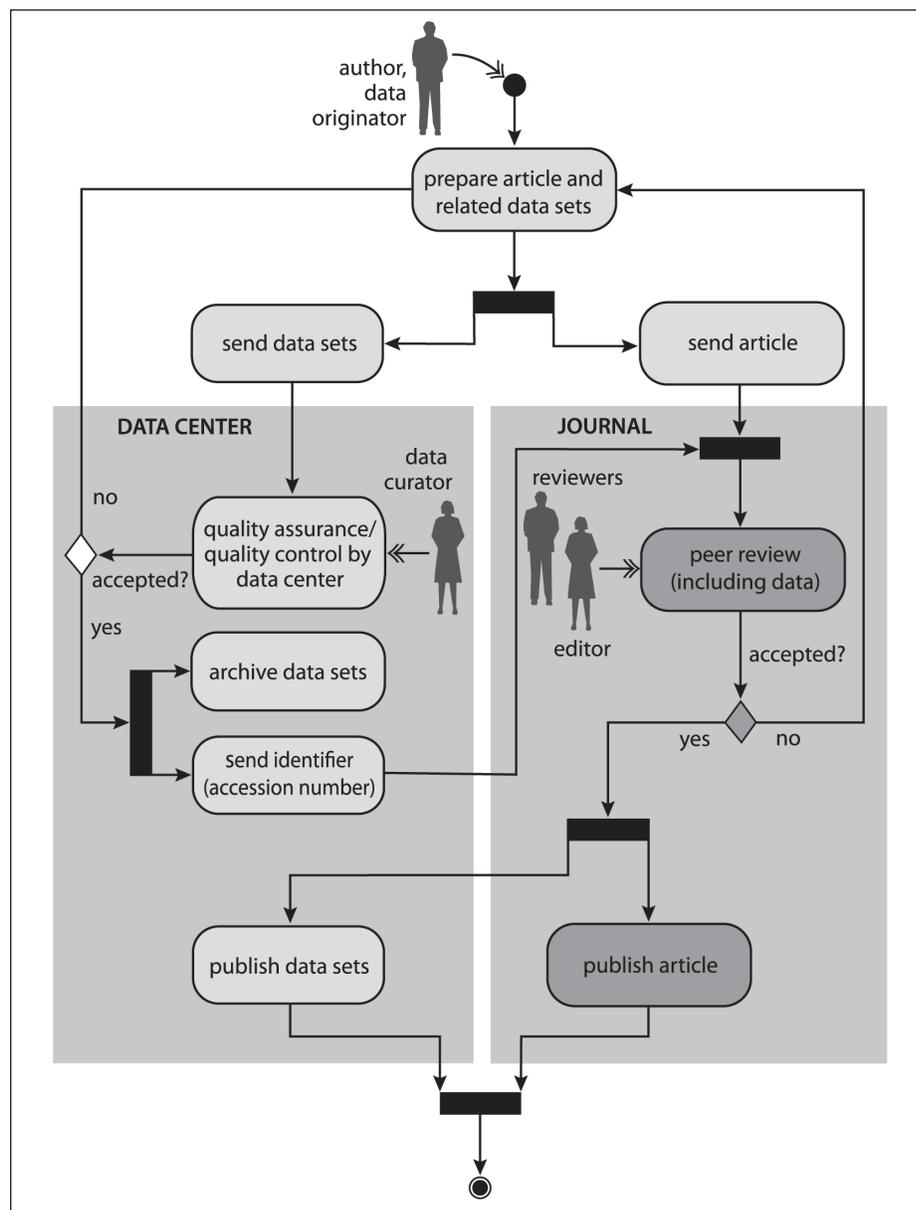


Fig. 1. Suggested work flow for peer-reviewed data publications in the ocean sciences. Image modified from Scientific Committee on Oceanic Research/International Oceanographic Data and Information Exchange [2008]. At the end point of this flowchart are freely available peer-reviewed papers and data sets.

Submission of data associated with journal publications is a standard practice in other domains, such as molecular biology, in which the gene sequences that are described in peer-reviewed publications must be submitted to GenBank or related archives. To help make such submissions standard in the ocean sciences, SCOR and IODE are working with editors and publishers of journals to discuss how to implement greater use of data publication.

Building a Digital Backbone

To archive and administer data related to journal publications, additional infrastructure in data management systems is required. Such infrastructure must be implemented with minimal costs to avoid impeding the publication process (see Figure 1). The “eRepository” technology developed by the digital library community delivers some of the functionality needed for this infrastructure. However, it does not provide added value—in terms of harmonization with other data in the system, quality control, and metadata enhancement—associated with the IODE network of national data centers.

A workable compromise would be to use eRepository technology as “front-end” processes of data centers that serve ingested data sets “as is” in the short term, as well as providing added value to data sets through existing data management infrastructure in the medium and long terms. This new infrastructure should improve the data publication review process through closer collaboration between data centers and journal editors.

SCOR and IODE are working with existing data centers, libraries, and journals to promote the development of the infrastructure required to provide ocean sciences publications with an effective “digital backbone.”

Other groups are also spearheading efforts to link academic journals to data repositories. Ongoing cooperative activities are along three lines:

1. SCOR and IODE are continuing to work with editors of ocean science journals to establish pilot projects along the lines described in Figure 1.

2. The Marine Biological Laboratory/Woods Hole Oceanographic Institution (WHOI) library is working with the U.S. Biological and Chemical Oceanography Data Management Office (BCO/DMO) at WHOI on a pilot project on how libraries and data centers could work together to provide the digital backbone for traditional journal publications, ensuring that data sets have appropriate associated metadata and are easily accessible.

3. The British Oceanographic Data Centre is working on a pilot project to repackage existing data holdings into data sets appropriate for assignment of persistent identifiers to provide a mechanism for concrete links to scientific publications.

The work flow diagram in Figure 1 will be revised as scientists, data managers, and journal editors gain experience from the pilot projects. Important questions raised by the ocean science community include the following:

- What should be the details of quality control in data centers? A simple action would be to ensure that submitted data are machine readable. Other actions might be to ensure that data sets include a minimal set of metadata.
- What happens to data associated with articles that are not published? Such data may still be valuable to other scientists, and archiving should ensure that the data originator receives appropriate credit.
- What processes will be needed to ensure that data are archived, assigned a

persistent identifier, and accessible before the associated paper is published? The timing surrounding the implementation of this process is especially important as publication times become faster and review drafts of papers become available through electronic publishing.

- What are the rights and responsibilities of data archives during the review process, in terms of data release, data protection, timing, etc.?

- What existing persistent identifier should be assigned to data referenced in journal articles? Digital object identifiers (DOIs) have become an almost de facto standard in journal publishing, but other options exist. Whichever identifier is used, the issue of the “least publishable unit” for assignment of an identifier must be tackled.

More details about the SCOR/IODE activity are available at <http://www.iode.org/datapublishing>. The authors welcome input on this topic from the geosciences community.

Reference

Scientific Committee on Oceanic Research/International Oceanographic Data and Information Exchange (SCOR/IODE) (2008), SCOR/IODE workshop on data publishing: IOC Project Office for IODE, Oostende, Belgium, 17–18 June 2008, *Workshop Rep. 207*, 23 pp., United Nations Educ., Sci., and Cult. Organ., Paris.

—ROY LOWRY, British Oceanographic Data Centre, Liverpool, UK; ED URBAN, Scientific Committee on Oceanic Research, University of Delaware, Newark; and PETER PISSIERSSENS, United Nations Educational, Scientific and Cultural Organization’s Intergovernmental Oceanographic Commission Project Office for the International Oceanographic Data and Information Exchange, Oostende, Belgium; E-mail: p.pissierssens@unesco.org