

University of Massachusetts Medical School

eScholarship@UMMS

---

University of Massachusetts and New England  
Area Librarian e-Science Symposium

2012 e-Science Symposium

---

Apr 4th, 12:00 AM - 2:00 PM

## 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*

Ed Urban

*Scientific Committee on Oceanic Research*

Gwenaelle Moncoiffe

*British Oceanographic Data Centre*

*See next page for additional authors*

Follow this and additional works at: [https://escholarship.umassmed.edu/escience\\_symposium](https://escholarship.umassmed.edu/escience_symposium)

 Part of the [Library and Information Science Commons](#)

---

### Repository Citation

Raymond, L., Chandler, C., Lowry, R., Urban, E., Moncoiffe, G., Pissierssens, P., Norton, C., & Miller, H. (2012). Emerging Role of Librarians in Data Publication. *University of Massachusetts and New England Area Librarian e-Science Symposium*. <https://doi.org/10.13028/9bx7-kr42>. Retrieved from [https://escholarship.umassmed.edu/escience\\_symposium/2012/posters/5](https://escholarship.umassmed.edu/escience_symposium/2012/posters/5)

Creative Commons License



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 3.0 License](#). 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](mailto:Lisa.Palmer@umassmed.edu).

---

**Presenter Information**

Lisa Raymond, Cyndy Chandler, Roy Lowry, Ed Urban, Gwenaelle Moncoiffe, Peter Pissierssens, Cathy Norton, and Holly Miller



# Emerging Role of Librarians in Data Publication

Lisa Raymond<sup>1</sup>, Cyndy Chandler<sup>1</sup>, Roy Lowry<sup>2</sup>, Ed Urban<sup>3</sup>, Gwenaëlle Moncoiffé<sup>2</sup>, Peter Pissierssens<sup>4</sup>, Cathy Norton<sup>5</sup>, Holly Miller<sup>5</sup>

1. Woods Hole Oceanographic Institution (WHOI); 2. British Oceanographic Data Centre (BODC); 3. Scientific Committee on Oceanic Research (SCOR); 5. IOC Project Office of IODE; 6. Marine Biological Laboratory (MBL)

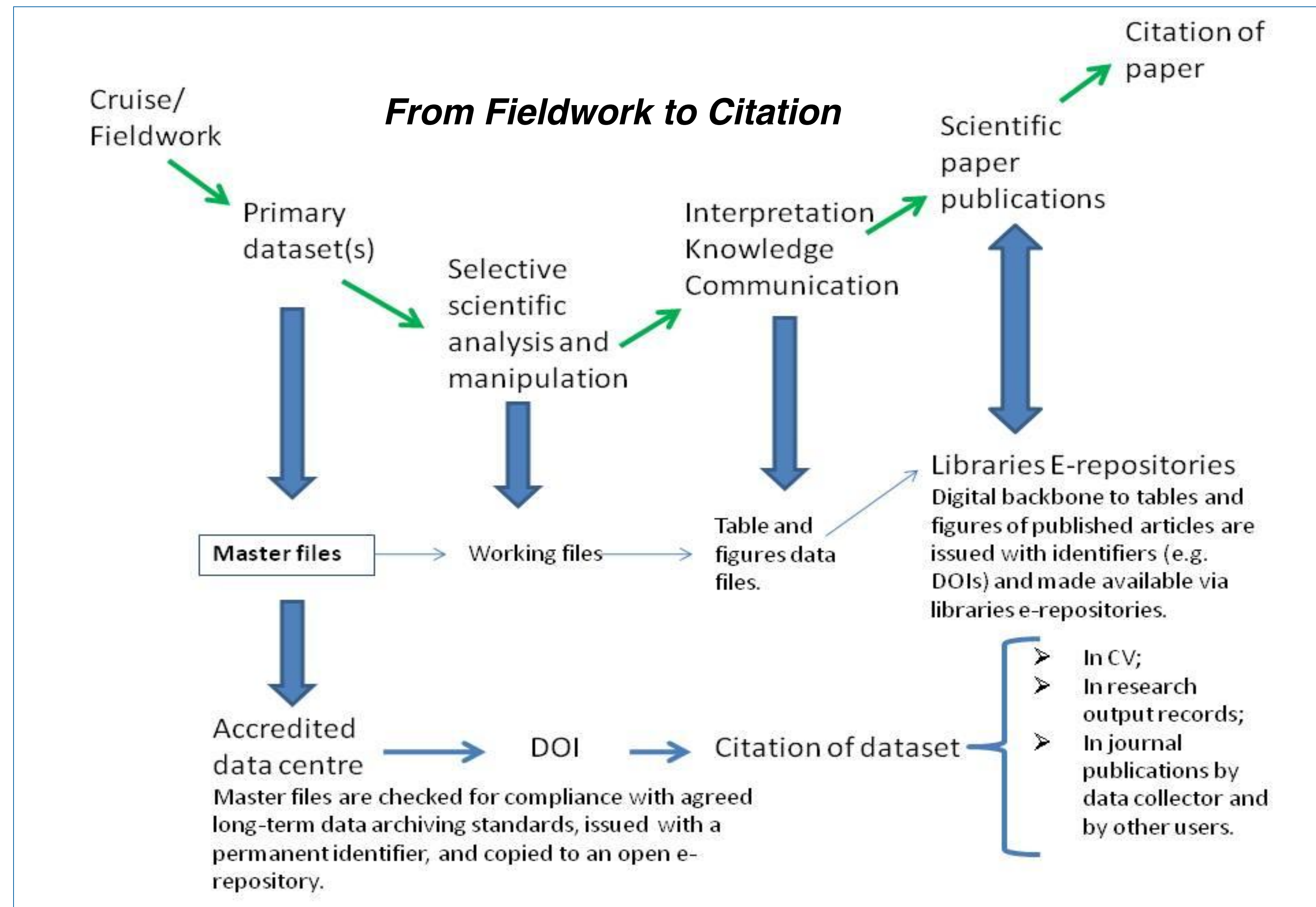


Diagram showing the evolution of a scientific dataset through its different stages from raw data collection (fieldwork) to the generation of master files which must be preserved from accidental loss, self-described, exchangeable, re-usable, and citable (by deposition in an accredited data center), through to the preparation of analyzed 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.

## 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) Library and the Biological and Chemical Oceanography Data Management Office (BCO-DMO) have developed tools and processes to automate the ingestion of metadata from BCO-DMO for deposit with datasets into the Institutional Repository (IR) 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 high-quality data while the data repository utilizes library services and is assured of a permanent archive of the copy of the data extracted from the repository database.

This research is being conducted by a team of librarians, data managers and scientists that are collaborating with representatives from the Scientific Committee on Oceanic Research (SCOR) and the International Oceanographic Data and Information Exchange (IODE) of the Intergovernmental Oceanographic Commission (IOC). The goal is to identify best practices for tracking data provenance and clearly attributing credit to data collectors/providers.

## Method

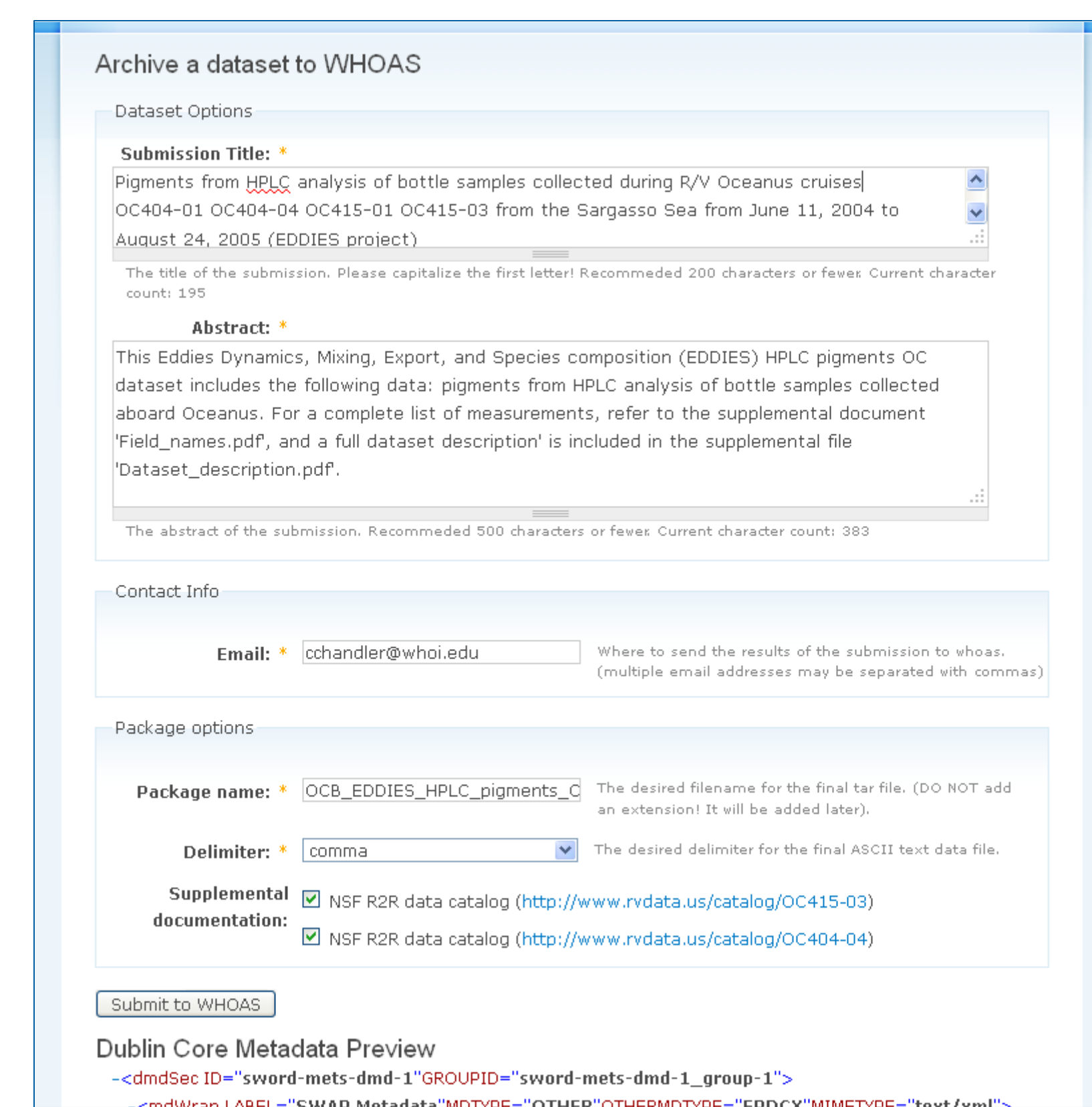
### Basic Workflow

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 data automatically and returns a handle URL to BCO-DMO's system.
9. BCO-DMO's system parses the handle URI 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 WHOAS manager via email.
12. The WHOAS manager creates a DOI for the submitted dataset record.

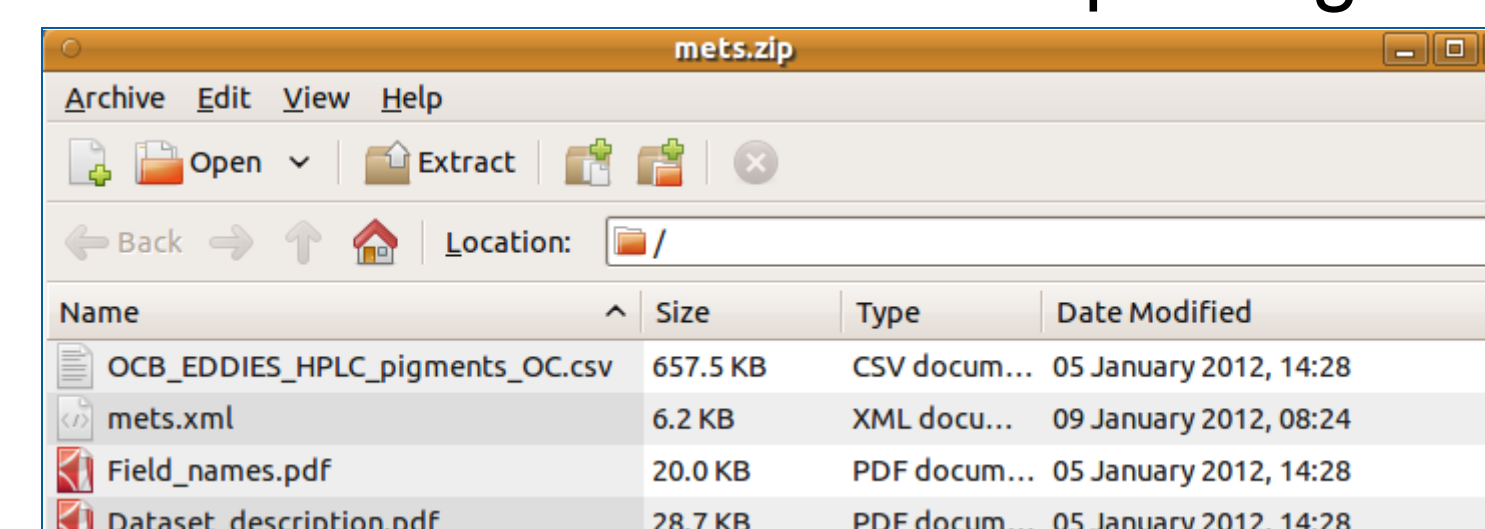
**SWORD METS Package**  
**Simple Web-service Offering Repository Deposit (SWORD)**  
**Metadata Encoding and Transmission Standard (METS)**

- 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/DSpaceMETS+SIP+Profile>.
- A METS package formatted per this profile is a .zip file which contains these items:
  - a mets.xml file that contains metadata about the digital objects
  - The 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.

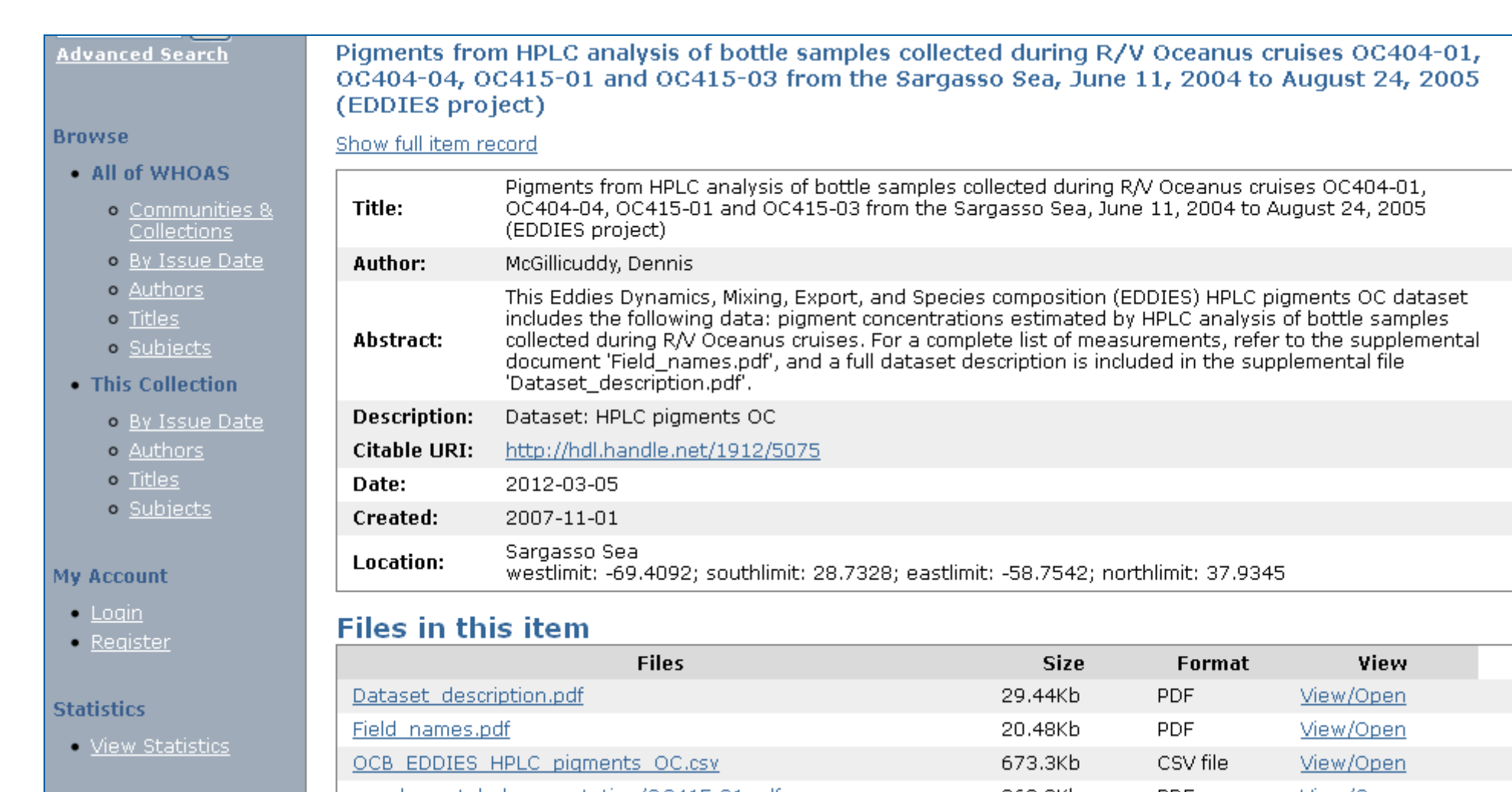
The BCO-DMO submission interface:



The contents of a METS package:



The result of a successful submission in WHOAS:



## 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.

The assignment of DOIs has enabled another important collaboration with Elsevier, publisher of educational and professional science journals. Elsevier can now link from articles in the Science Direct database to the datasets available from WHOAS that are related to that article. The data associated with the article are freely available from WHOAS and accompanied by a Dublin Core metadata record.

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://darchive.mblwhoilibrary.org/>
- Biological and Chemical Oceanography Data Management Office <http://www.bco-dmo.org/>
- British Oceanographic Data Centre [www.bodc.ac.uk](http://www.bodc.ac.uk)
- Information about the project, including reports from project meetings, can be found at <http://www.ioode.org/datapublishing>

## Acknowledgments

Funding provided by the Jewett Foundation

Ann Devenish, WHOAS Manager  
 Tobias Work, BCO-DMO Programmer  
 Alex Dorsk, Library Programmer

