Dataset to DOI using SWORD

Ann Devenish (MBLWHOI Library), Cyndy Chandler (BCO-DMO), Alex Dorsk (MBLWHOI Library), Lisa Raymond (MBLWHOI Library), Tobias Work (BCO-DMO)
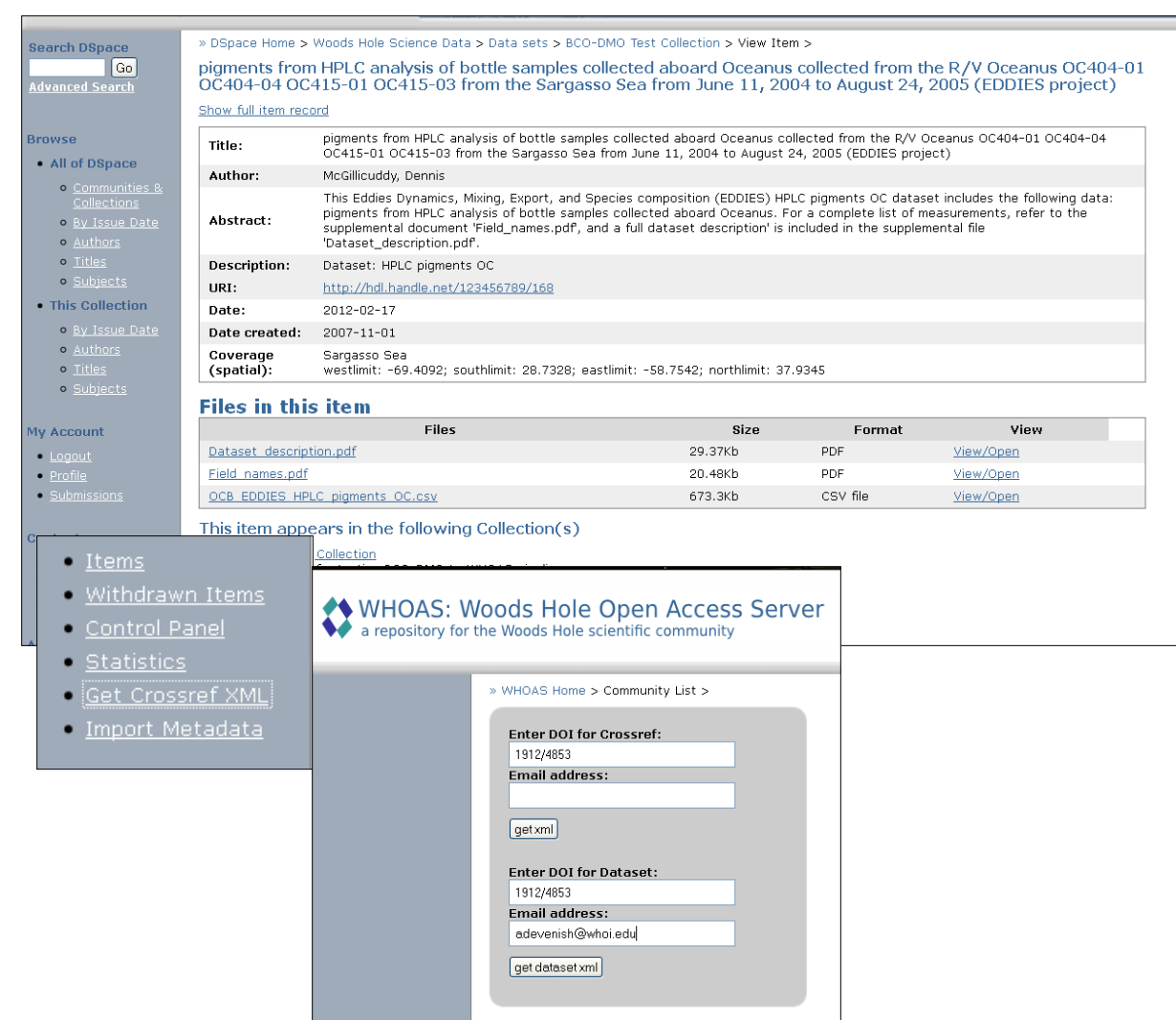
Introduction

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 datasets and metadata from BCO-DMO for deposit into the Woods Hole Open Access Server (WHOAS) institutional repository. The system also incorporates functionality for BCO-DMO to request a DOI from the Library and streamlines the process by which XML files are generated for DOI deposit with CrossRef, a DOI registration agency.

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.

Simplified Workflow

Datasets are submitted to BCO-DMO by researchers.

BCO-DMO staff create metadata and store the datasets.

Authors wishing to cite a dataset in an article send a request to BCO-DMO.

BCO-DMO staff click the "submit to WHOAS" button.

A METS package (dataset and metadata) is assembled and pushed from BCO-DMO to WHOAS, using SWORD.

Upon successful ingest into WHOAS, a Handle is returned to BCO-DMO, and Library staff receive automated notice of the ingest.

BCO-DMO software parses the Handle to determine the DOI and advise the author.

Library staff assign the DOI in WHOAS.

Library staff click the "Get CrossRef XML" link, generating an XML file for upload to CrossRef.

The SWORD crosswalk file was customized to handle additional metadata fields as the default DSpace crosswalk did not contain all of the mappings needed.   Longsight added these mappings:

dc.relation → http://purl.org/dc/elements/1.1/relation

dc.coverage.spatial → http://purl.org/dc/elements/1.1/coverage/spatial

dc.coverage.temporal → http://purl.org/dc/elements/1.1/coverage/temporal

dc.date.created → http://purl.org/dc/elements/1.1/date/created

dc.subject → http://purl.org/dc/elements/1.1/subject

dc.description.sponsorship → http://purl.org/dc/elements/1.1/description/sponsorship

dc.type → http://purl.org/dc/elements/1.1/type

dc.publisher → http://purl.org/dc/elements/1.1/publisher

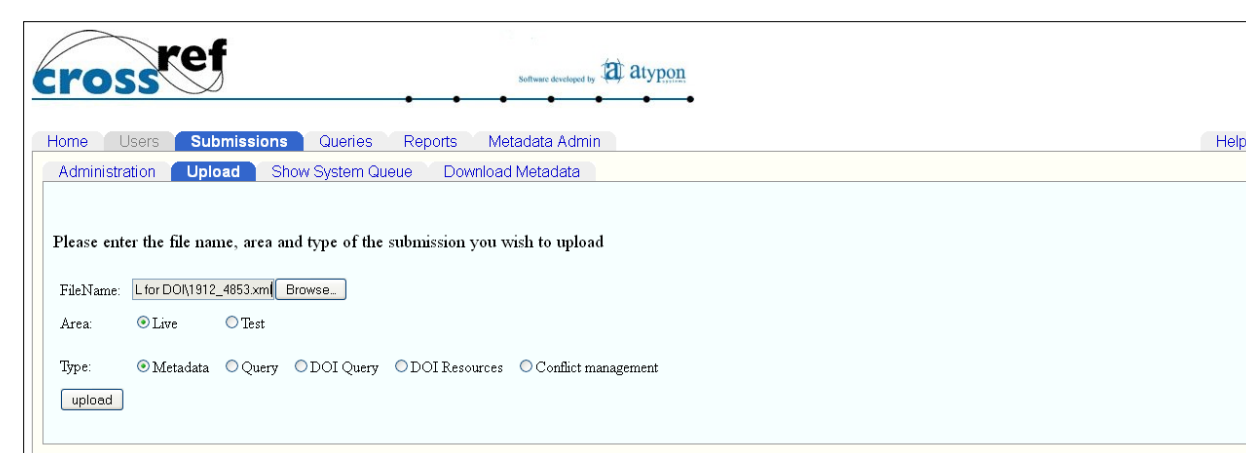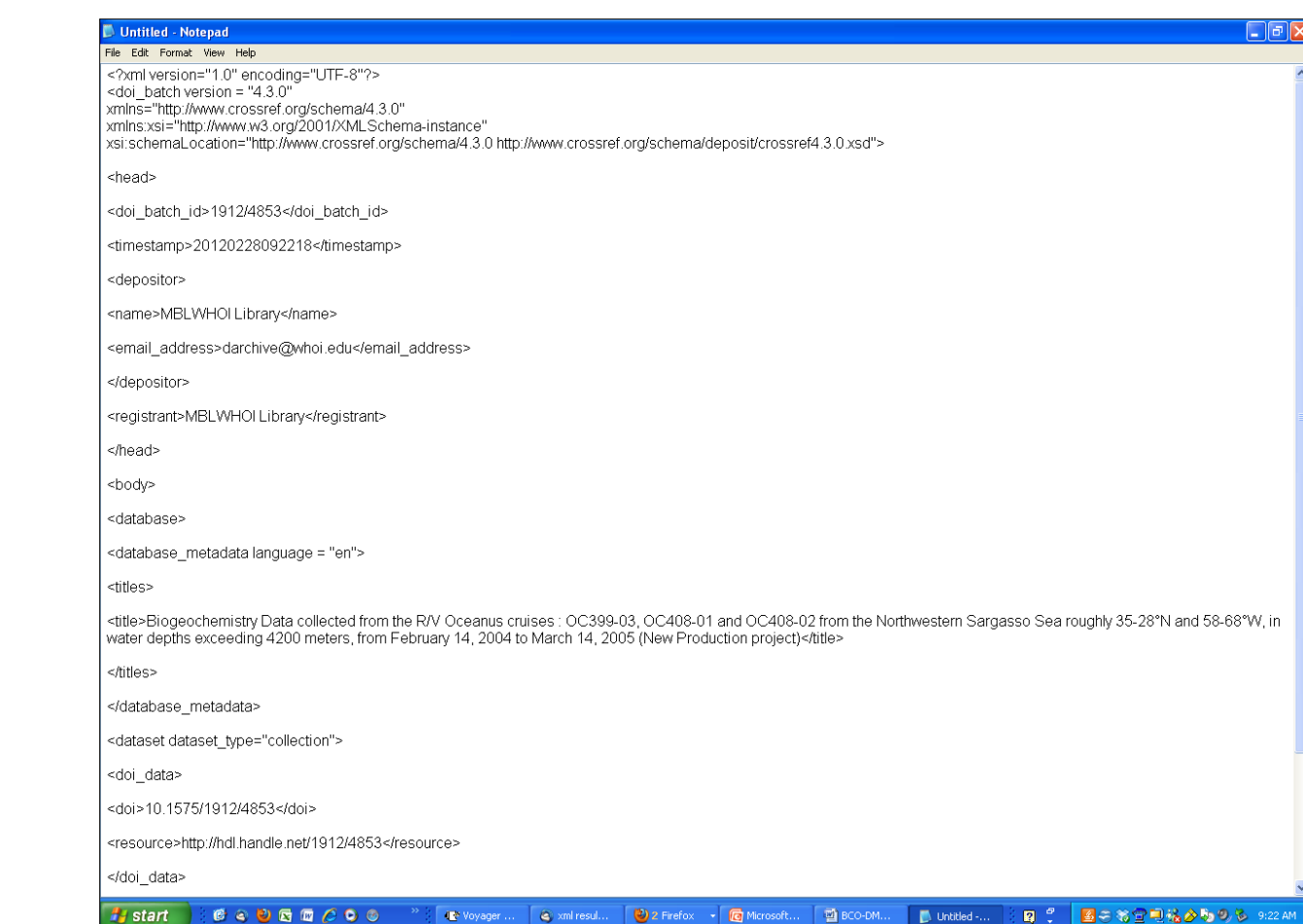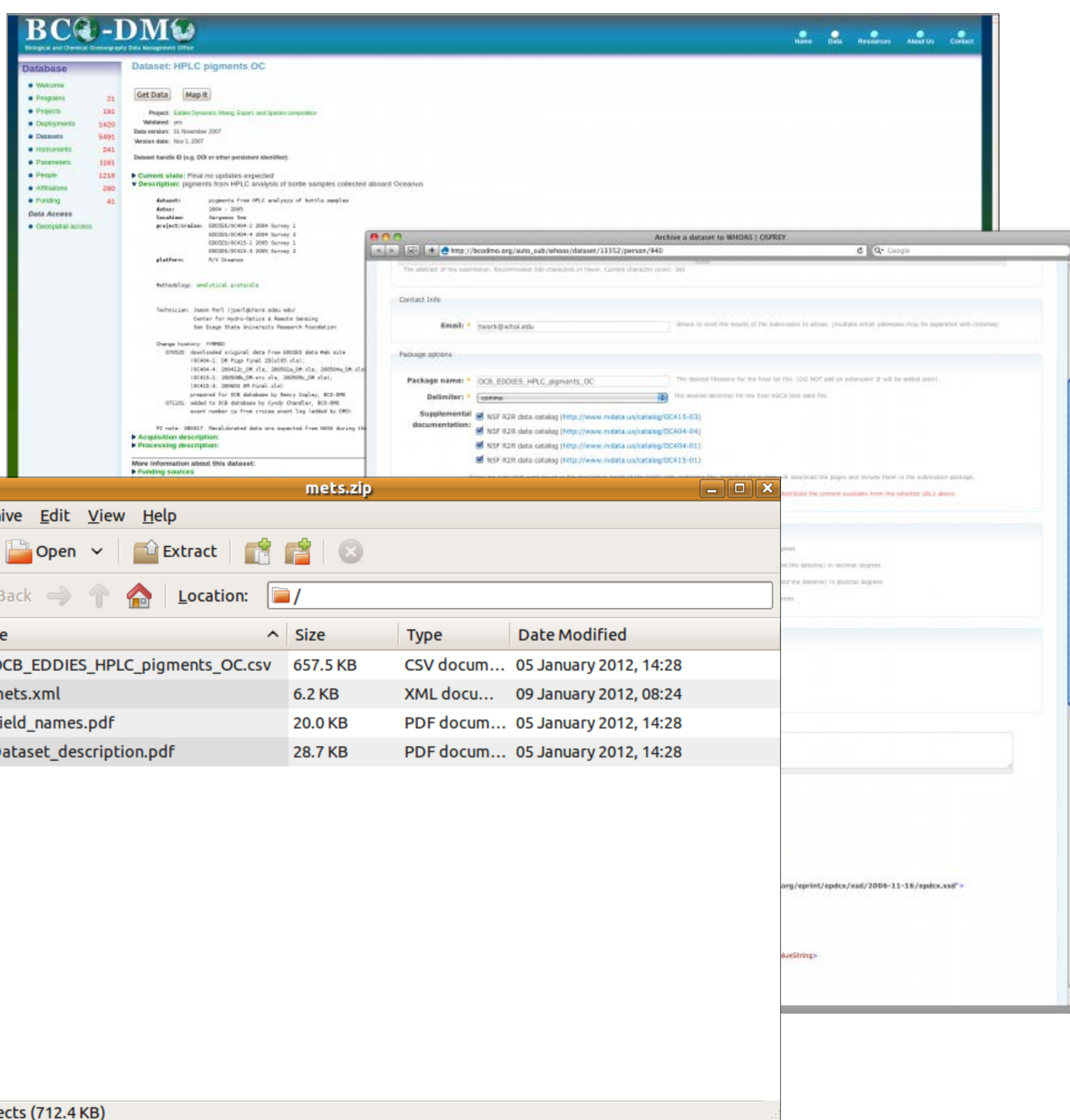dc.description → http://purl.org/dc/elements/1.1/description

Configuration changes made within DSpace make it possible for BCO-DMO's system to make submissions. Metadata elements were added to the Dublin Core elements registry for SWORD :

dc.identifier.slug

dc.language.rfc3066

dc.date.updated

dc.description.version

dc.rights.holder

Other changes :

Creation of a unique user in DSpace to use for submissions:

bcodmo@gmail.com.

Granting of administrator rights to that user on the BCO-DMO collection

Removing any other users from the workflow steps for the BCO-DMO collection.

Altering the workflow so that submissions are automatically accepted; Handle is automatically returned to BCO-DMO's submission system via email.

Definitions:

BCO-DMO: Biological and Chemical Oceanography Data Management Office, based at the Woods Hole Oceanographic Institution .

CrossRef: a DOI registration agency.

Handle: an unique and persistent identifier of a digital object, used by DSpace

DOI: digital object identifier.

Longsight Group: vendor providing external IT support for WHOAS.

SWORD: Simple Web service Offering Repository Deposit, a lightweight protocol for repository deposit, creation of which was funded by Joint Information Systems Committee (JISC).

WHOAS: Woods Hole Open Access Server, a DSpace repository.