

Sarah Carrier
July 30, 2009

Dryad Curation Manual, Summer 2009

Introduction

Dryad is being designed as a "catch-all" repository for numerical tables and all other kinds of published data that do not currently have a home. A major design consideration with these data is to avoid placing an undue burden of metadata generation on individual researchers while at the same time capturing sufficient metadata to enable data discovery and reuse. Dryad's curator will also reduce the burden upon depositors by ensuring the accuracy, completeness, and consistency of author-provided metadata. Furthermore, the curator will provide a service to authors by ensuring the permanence and usability of their data through the use of appropriate tools and preservation metadata.

Purpose of this document

The purpose of this document is to direct the curatorial management of data and metadata in Dryad, and will detail the current curation workflow as of summer 2009. This document will guide the curation of individual submissions as they are received, as well as guide the ongoing curation of all objects stored in Dryad.

The Digital Curation Centre defines data curation as: "The activity of managing the use of data from its point of creation to ensure it is available for discovery and re-use in the future." Data curation can also include managing vast data sets for daily use; updating it to keep it readable, etc. Therefore, the term data curator is applicable to a large range of professional backgrounds, from minimal management of digital materials, to the addition of metadata, to managing institutional repositories.¹

Dryad's curator will be an information specialist who has knowledge of and experience working with information standards. Overall, their responsibility will be the preservation, archiving, and the provision of access to digital data stored in Dryad. Furthermore, the curator will maintain documentation of cataloging and curation practices. Their work, as judged by the quality of the data in Dryad, will be assessed by quality control measures such as: completeness, accuracy, conformance to standards and user expectations, and timeliness (i.e., the use of current terminology).

Curation tasks

Prevailing rules and policy

1. Issue of ADDING metadata
 - a. Metadata can be added if an optional field is not completed.

¹ <http://www.dcc.ac.uk/FAQs/data-curator#q1>

- i. If the metadata is KNOWN, for example, if there are clearly keywords associated with an article, and the depositor did not include them, then the curator will add them.
 - ii. For metadata that is not explicitly known to the curator, for example, other subject keywords, or temporal/geographical metadata, this will only be added for “high use” or “high quality” data packages, as determined by use and prominence. Currently, such information is not tracked in Dryad, and therefore a list of “high use” data is not available.
- 2. Issue of EDITING metadata
 - a. This should be considered an "ownership" issue – the depositor “owns” the metadata they create.
 - b. The curator can correct what is "obviously" wrong – for example, misspellings, metadata in the wrong field, etc.
 - i. Feedback from management board: author/depositor-provided metadata should be edited by the curator.
 - c. Significant changes to author-supplied metadata will have to be done AFTER contacting the author. What constitutes a major change may have to be considered on a case-by-case basis. DELETING metadata, for example removing a keyword, will always be a significant change requiring contact with the depositor.
- 3. If a file is in a proprietary format, it should be converted to the "least common denominator" format - and something that is ideally open source. For example, Excel files should be converted to tab delimited files and stored in Dryad.

Minimum Criteria for Curation

What is the minimum amount of curation that a submission requires? What is the minimum effort to curate the entire contents of Dryad?

For BOTH data packages and publications, it is important to double check if the right metadata has gone into the appropriate fields. Also, it is important to check for correct spelling and punctuation in all metadata fields. This task can also be accomplished with the use of an external spell check tool, and this tool should be able to be implemented in batch operations. This is the absolute minimum amount of effort.

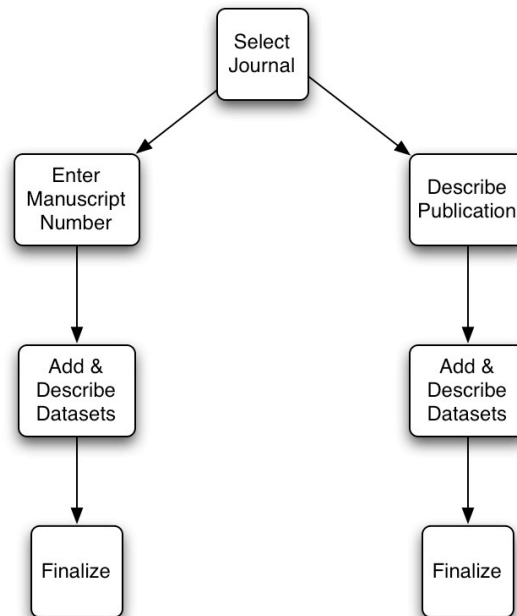
For preservation purposes, the minimum effort entails checking to see if the formats are valid, that the files can be opened and used, and that proprietary formats are converted to the “lowest common denominator.” Currently, file conversion will not be taking place, but will be added to the workflow in the near future.

Deposition Process - July 2009

Currently, the depositor must manually create metadata records for both the publication and the data package. During the submission process, the link between a publication and associated data files is done automatically, with the option of editing.

1. Describe the publication.
2. Upload and describe the associated data packages.
3. Approve data packages for publication.

Deposition Process - future



1. Select the journal in which the article appears using the dropdown menu.
2. If the journal is partner journal, enter the manuscript number. This will automatically prepopulate the metadata fields for the publication corresponding to the manuscript number.
3. If the article is not from a partner journal, and "Other" is chosen, the next step is to describe the article in as much detail as possible.
4. Upload data packages and describe them. They can be uploaded individually, or all together as a zip file.
5. Edit the publication metadata, or choose to finalize the submission.

There are two current submission workflows that the author/depositor would undertake: one, where the publication metadata is automatically imported, and two, where the publication metadata has to be manually entered before upload of data.

Required fields to describe the publication: title, authors, journal name. Required fields for the data package: title (authors, keywords, etc. inherited from publication).

Places where curator enhances metadata creation, and tasks that would be beyond the MINIMUM effort of curation:

1. Confirm accuracy and correct author-created metadata – could require EDITING depositor metadata
2. ADD optional metadata
3. SUPPLEMENT author-created metadata, i.e., add more keywords if they are known, etc.

General Steps to Curate a NEW Submission

The curator is not going to be ADDING metadata to the majority of objects in Dryad, rather they are simply checking for accuracy and completeness. Currently, publication metadata will need to be double-checked because it is not being supplied by the journals. This will change when partner journals are fully integrated into Dryad. The curator may need to find the DOI for articles once they are published, and add this information to the publication metadata.

Currently, items are published immediately once a depositor approves them. New submissions must be tracked semi-manually by the curator, and if edits are required, they are therefore done “live.” The default DSpace workflow system should be turned on, however, which means that new submission WILL go into a queue before publishing. The curator then receives an email saying that they have a new task, with a link that takes them to “My Dryad.”

There will be two major workflows for new submissions: one for the publications, and one for the data packages. While inheritance of metadata takes place automatically during deposition (the data file inherits metadata from the publication), currently changes made to the publication metadata AFTER it has been published are not reflected in the records for associated data files. Therefore, the curator must make changes manually for each data file as well.

PUBLICATIONS

Steps that the curator will follow:

1. Find ORIGINAL article - go to publisher's website. This is hopefully as simple as copying the DOI provided by the depositor and pasting it into the address bar. If a DOI has not been provided, then it needs to be found. To find the publication without a DOI, the title of the article, the journal name, etc. can be used.
 - a. Here the curator will find the majority of the information needed to check the accuracy and completeness of the PUBLICATION metadata.
 - b. Double check the author's names - if there is only a first initial, other sources will need to be referenced in order to find the full author's name, for example, ISI Web of Science and/or PubMed.
 - c. Double-check the TITLE, JOURNAL NAME, the SUBJECT KEYWORDS, CITATION information (year, volume, issue, pages, etc.) - check for accuracy.
 - d. THE CITATION STRING for the ORIGINAL ARTICLE: we have not yet chosen a standard for this - in the meantime, we will be using the AMERICAN NATURALIST's citation format, for example: Belyea, L. R., and J. Lancaster. 1999. Assembly rules within a contingent ecology. *Oikos* 86(7): 402–416. The name of the journal must be spelled out.
2. Is there a corresponding author? This information needs to be checked, and added if it was not supplied - the name should be on the article or on the publisher's website.

DATA FILES

Metadata is inherited by from the publication, but the depositor can change or add to it. Depositors also have the option to create a unique title and add a description to a data file. Currently, there is an option to choose and embargo, but the majority of items being deposited to Dryad will have no embargo. After journal integration, the embargo issue will come into play, and manually setting embargoes will be added to the workflow. If an embargo is chosen, it will be based on the publication date of the original article.

1. Metadata that is added by the depositor, or edits made to inherited publication data should be double-checked for spelling, punctuation, and accuracy (to the best of the curator's knowledge).
2. Currently, we are accepting anything for the data file title and the description.
3. Double check the FORMAT of the file - is this correct? Can it be downloaded/opened/etc.? Tools can be utilized to assist with this process, namely DROID² and PRONOM³.

Detailed Steps to Curate Submissions

1. You will receive an email saying that you have a new task. The link in the email will take you directly to a page that lists items needing to be curated.

Submissions & workflow tasks

Workflow tasks

These tasks are items that are awaiting approval before they are added to the repository. There are two task queues, one for tasks which you have chosen to accept and another for tasks which have not been taken by anyone yet.

Tasks you own

Task	Item	Collection	Submitter
No tasks are assigned to you			

Tasks in the pool

Task	Item	Collection	Submitter
<input type="checkbox"/> Awaiting editor's attention	aa	Publications	email: Ryan Scherle
<input type="button" value="Take selected tasks"/>			

Submissions

You may [start a new submission](#). The submission process includes describing the item and uploading the file(s) comprising it. Each community or collection may set its own submission policy.

2. You have the option to view the item(s) before publishing. This is when the editing of metadata, if necessary, takes place. Click on the task as listed, which gives you the following options:

² DROID (Digital Record Object Identification) is a software tool developed by The National Archives to perform automated batch identification of file formats. DROID is a platform-independent Java tool, which is freely available to download under an open source license.

³ PRONOM is an on-line information system about data file formats and their supporting software products. Originally developed to support the accession and long-term preservation of electronic records held by the National Archives, PRONOM has been made available as a resource for anyone requiring access to this type of information.

Actions you may perform on this task:

If you have reviewed the item and it is suitable for inclusion in the collection, select "Approve".	<input type="button" value="Approve item"/>
If you have reviewed the item and found it is not suitable for inclusion in the collection, select "Reject". You will then be asked to enter a message indicating why the item is unsuitable, and whether the submitter should change something and resubmit.	<input type="button" value="Reject item"/>
Select this option to change the item's metadata.	<input type="button" value="Edit metadata"/>
Return the task to the pool so that another user may perform the task.	<input type="button" value="Return task to pool"/>
<input type="button" value="Cancel"/>	

3. Choose to edit the metadata.
4. Click the "Show full item record" link under the title to see all of the metadata contributed by the depositor.

Dryad Home > Main > Publications > View Item >

The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*

[Show full item record](#)

Dryad Identifier	http://dx.doi.org/10.1038/ng766 http://hdl.handle.net/10255/dryad.120
Authors	Gisele Passador-Gurgel; Greg Gibson; Rebecca M. Riley; Wei Jin; Kevin P. White; Russell D. Wolfinger

5. The metadata is displayed as follows:

The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*

[Show simple item record](#)

dc.contributor.author	Gisele Passador-Gurgel	
dc.contributor.author	Greg Gibson	
dc.contributor.author	Rebecca M. Riley	
dc.contributor.author	Wei Jin	
dc.contributor.author	Kevin P. White	
dc.contributor.author	Russell D. Wolfinger	
dc.date.accessioned	2008-03-28T14:09:28Z	
dc.date.available	2008-03-28T14:09:28Z	
dc.date.issued	2001	
dc.identifier.citation	Jin, W., Riley, R. M., Wolfinger, R. D., White, K. P., Passador-Gurgel, G., & Gibson, G. (2001). The contributions of sex, genotype and age to transcriptional variance in <i>Drosophila melanogaster</i> . <i>Nat Genet</i> , 29(4), 389-395. doi: 10.1038/ng766.	en_US
dc.identifier.uri	http://dx.doi.org/10.1038/ng766	
dc.identifier.uri	http://hdl.handle.net/10255/dryad.120	

- Choose "Edit this item" from the sidebar navigation under Context, if the metadata needs to be edited:

DRYAD

Dryad Home > Main > Publications > View Item >

Search Dryad

[Advanced Search](#)

Navigation

[Depositing Data](#)

[Using Data](#)

[Dryad Partners](#)

[About Dryad](#)

My Account

[Logout](#)

[Profile](#)

Context

[Edit this item](#)

[Export Item](#)

The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*

[Show simple item record](#)

dc.contributor.author	Gisele Passador-Gurgel	
dc.contributor.author	Greg Gibson	
dc.contributor.author	Rebecca M. Riley	
dc.contributor.author	Wei Jin	
dc.contributor.author	Kevin P. White	
dc.contributor.author	Russell D. Wolfinger	
dc.date.accessioned	2008-03-28T14:09:28Z	
dc.date.available	2008-03-28T14:09:28Z	
dc.date.issued	2001	
dc.identifier.citation	Jin, W., Riley, R. M., Wolfinger, R. D., White, K. P., Passador-Gurgel, G., & Gibson, G. (2001). The contributions of sex, genotype and age to transcriptional variance in <i>Drosophila melanogaster</i> . <i>Nat Genet</i> , 29(4), 389-395. doi: 10.1038/ng766.	en_US
dc.identifier.uri	http://dx.doi.org/10.1038/ng766	

- Choose "Item Metadata:"

Edit Item

- [Item Status](#)
- [Item Bitstreams](#)
- [Item Embargo](#)
- [Item Metadata](#)
- [View Item](#)

8. Metadata fields appear as follows. Simply make the changes in the “Value” field. If you need to remove a field, click the box to the left of the field Name. When finished, choose “Update” at the top or the bottom of the page.

Metadata

Remove	Name	Value	Language
<input type="checkbox"/>	dc.contributor.author	Gisele <u>Passador-Gurgel</u>	<input type="text"/>
<input type="checkbox"/>	dc.contributor.author	Greg Gibson	<input type="text"/>
<input type="checkbox"/>	dc.contributor.author	<u>Rebecca M. Riley</u>	<input type="text"/>
<input type="checkbox"/>	dc.contributor.author	<u>Wei Jin</u>	<input type="text"/>
<input type="checkbox"/>	dc.contributor.author	Kevin P. White	<input type="text"/>
<input type="checkbox"/>	dc.contributor.author	Russell D. <u>Wolfinger</u>	<input type="text"/>
<input type="checkbox"/>	dc.contributor.correspondingAuthor	Greg Gibson	<input type="text"/>

9. Adding metadata:

Add new metadata

Name:

dc.contributor.advisor

Value:

Language

Add new metadata

10. When finished, choose Update:

PLEASE NOTE: These changes are not validated in any way. You are responsible for entering the data in the correct format. If you are not sure what the format is, please do NOT make changes.

Update Return

11. When viewing the publication record, the associated data files are listed as links. Each one needs to be viewed one by one to determine if there is unique metadata that the publication doesn't have, or if any edits are made to the publication, they must be made to each file individually and manually. The process for editing data metadata is the same as above.

SOME ISSUES TO NOTE:

1. When editing the author's names for an item in the "Edit Metadata" feature, it rearranges the order - therefore you lose the first author if was listed first. The rearranging seems to be random. When this happens, you need to delete all of the author fields, and recreate them in the correct order.

Other tasks

1. Adding a data file to a publication that has already been published in Dryad:
 - a. Create a new submission, using fake information for the publication, and uploading the "real" data file.
 - b. Once it is published, you can edit the dc.relation fields to connect the new file to the old package, and then withdraw the fake publication.
 - i. NOTE: there is a temporary glitch where the item has to be DELETED, and not withdrawn, in order for the item to not pollutes the search results.
2. When needing to withdraw an item: Choose "Edit this item" from the sidebar navigation under Context. The default is "Item Status," and it is here that you would withdraw an item.

Edit Item

[Item Status](#)

[Item Bitstreams](#)

[Item Embargo](#)

[Item Metadata](#)

[View Item](#)

Welcome to the item management page. From here you can withdraw, reinstate or delete the item. You may also update or add new metadata / bitstreams on the other tabs.

Item Internal ID:	128
Handle:	10255/dryad.120
Last Modified:	2009-06-16 16:40:22.322
Item Page:	http://datadryad.org/repo/handle/10255/dryad.120
Item's Authorizations:	Edit Authorizations (system administrators only)
Withdrawn item from the repository:	Withdraw...
Completely expunge item:	Permanently delete (system administrators only)

[Return](#)

Data Dictionary

What follows is a listing of the metadata fields used in Dryad.

DATA FILE METADATA

Field Label	Formal Definition	User Definition	Requirement	Cardinality	Generation
Authors (dc.contributor.author)	Entity primarily responsible for making the content of the resource.	The entity or entities responsible for the creation and development of the data package.	Required	Repeatable	Inherited (after publication metadata is acquired or created, the names are inherited. List can be edited.)
Descriptive Title (dc.title)	A name given to the resource.	Descriptive title of the data file.	Required	Non-repeatable	Default is the name of the actual file, with the option of editing.
Date of Issue (dc.date.issued)	Issue Date of formal issuance (e.g., publication) of the resource.	Date of original publication.	Required	Non-repeatable	Automatically inherited from publication.
Embargo (dc.date.embargoedUntil)	A date after which the dataset will be made	A date after which the dataset will be	Optional	Non-repeatable	This is only used for datasets

	public.	made public.			under embargo. Will be set manually by curator.
Type (dc.type)	Type of the resource.		Required	Non-repeatable	[The default will be "Dataset" and done automatically.]
Subject Keywords (dc.subject)	The topic of the resource.	Data file keywords.	Required	Repeatable	Automatically inherited from publication, but can be edited and added to.
Description (dc.description)	Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.	Description of data file.	Optional	Non-repeatable	Manual
Described By Publication (dc.relation.ispartof)	A related resource.	Identifier of the published article with which dataset is associated.	Required	Repeatable	Automatically inherited from publication metadata.
Rights Statement (dc.rights.uri)	Information about rights held in and over the resource.	Statement regarding rights held in and over the resource.	Required	Repeatable	Automatic
Geographic Areas (dc.coverage.spatial)	Spatial topic may be a named place or a location specified by its geographic coordinates.	The spatial description of the data set specified by a geographic description and geographic coordinates.	Optional	Repeatable	Automatically inherited with the option to add/edit. Will be enhanced by HIVE.
Geologic Timespans (dc.coverage.temporal)	Temporal period may be a named period, date, or date range.	The temporal description of the data file including start date and end date of the collection/creation of the data file.	Optional	Repeatable	Automatically inherited with the option to add/edit.

Taxonomic Names (dwc.Scientific Name)	The full name of lowest level taxon to which the cataloged item can be identified (e.g., genus name, specific epithet, subspecific epithet, etc.).	The full name of lowest level taxon to which the cataloged item can be identified (e.g., genus name, specific epithet, subspecific epithet, etc.).	Optional	Repeatable	Currently manual, but will be enhanced by HIVE.
---------------------------------------	--	--	----------	------------	---

DSpace metadata automatically assigned to a data file:

- dc.date.accessioned
- dc.date.available
- dc.description.provenance
- dc.identifier.uri -> this is the Dryad handle that is automatically assigned
- File Metadata (bitstream format indicator) - Code indicating the type of file. This is automatically detected by DSpace, but can be modified manually.

PUBLICATION METADATA

Field Label	Formal Definition	User Definition	Requirement	Cardinality	Generation
Authors (dc.contributor.author)	Entity primarily responsible for making the content of the resource.	Author(s) of the article.	Required	Repeatable	Automatic for partner journals, manual for non-partners.
Article Title (dc.title)	A name given to the resource.	Title of the article.	Required	Non-repeatable	Automatic for partner journals, manual for non-partners.
Date of Issue (dc.date.issued)	Date of formal issuance (e.g., publication) of the resource.	Date of publication.	Required	Non-repeatable	Automatic for partner journals, manual for non-partners.
Publisher (dc.publisher)	An entity responsible for making the resource available.	Journal publisher.	Optional	Repeatable	Manual
Full Citation (dc.identifier.citation)	Details of the bibliographic item that contains the resource along with the position of the resource within it.	The citation information for the journal article.	Required	Repeatable	Automatic for partner journals, manual for non-partners.
Journal (dc.relation.isPartOfSeries)	A related resource in which the described	Name of journal.	Required	Non-repeatable	Automatic for partner journals,

	resource is physically or logically included.				manual for non-partners.
DOI (dc.identifier.uri)	An unambiguous reference to the resource within a given context.	The Digital Object Identifier of a journal article.	Required	Non-repeatable	Automatic, but for some articles the DOI will not be available and the curator must find it; manual for non-partners.
Type (dc.type)	Type of the resource.		Required	Non-repeatable	Automatic.
Language (dc.language.iso)	Language of the resource.		Optional	Non-repeatable	Currently the default is English.
Subject Keywords (dc.subject)	The topic of the resource.	Article keywords.	Optional	Repeatable	Automatically assigned for partner journals, manual for non-partners.
Abstract (dc.description.abstract)	An account of the resource.	Article abstract.	Required	Non-repeatable	Automatically assigned for partner journals, manual for non-partners.
Rights Statement (dc.rights.uri)	Information about rights held in and over the resource.	Statement regarding rights held in and over the resource.	Required	Repeatable	Automatic
Geographic Areas (dc.coverage.spatial)	Spatial topic may be a named place or a location specified by its geographic coordinates.	The spatial description of the data set specified by a geographic description and geographic coordinates.	Optional	Repeatable	Currently manual, later will be semi-automatic through use of HIVE.
Geologic Timespans (dc.coverage.temporal)	Temporal period may be a named period, date, or date range.	The temporal description of the article including start date and end date of the collection/creation of the data.	Optional	Repeatable	Currently manual.
Primary Contact (dc.contributor.corresponding)	An entity responsible for making contributions to	Corresponding author.	Optional	Non-repeatable (Dryad only allows one)	Automatically assigned for partner journals,

Author)	the resource.				manual for non-partners.
Manuscript Number (dc.identifier.manuscriptNumber)		Manuscript number.	Optional	Non-repeatable	If this is available, it will be automatic.
Taxonomic Names (dwc.Scientific Name)	The full name of lowest level taxon to which the cataloged item can be identified (e.g., genus name, specific epithet, subspecific epithet, etc.).	The full name of lowest level taxon to which the cataloged item can be identified (e.g., genus name, specific epithet, subspecific epithet, etc.).	Optional	Repeatable	Currently manual, but will be enhanced by HIVE.

DSpace metadata automatically assigned for the publication:

- dc.date.accessioned
- dc.date.available
- dc.description.provenance
- dc.identifier.uri -> assigned by Dryad as a handle
- *dc.relation.haspart* - automatically assigned by DSpace when data objects are uploaded for the publication, but please note that this can be edited manually to create relationships AFTER a publication/data file is published

Resources

A list of resources for the curator has been compiled on the Dryad wiki for the use of team members. Links are included to the original files. Please see “Resources for the curator.”⁴

GLOSSARY⁵

Bitstream: The binary form of data.

Derivative File: A derivative file is a version of the original file in a file format considered to be more stable than the original file format. For example, if a work is submitted as a Word document we may create a derivative in a plain text format.

Digital Preservation: The series of managed activities necessary to ensure authenticity of and continued access to digital materials for as long as possible.

DSpace: Open-source digital repository software developed by the Massachusetts Institute of Technology Libraries and Hewlett-Packard. DSpace is used by institutions

⁴ https://www.datadryad.org/wiki/Summer_2009_Curation_Workflow#Resources_for_the_curator

⁵ Adapted from UT Digital Repository glossary <http://repositories.lib.utexas.edu/glossary>

worldwide to capture, store, index, and redistribute digital data in a variety of formats. More detailed information about DSpace can be found at <http://www.dspace.org>.

Embargo: An instruction to restrict access to a work for a specified period of time.

File Format Migration: Moving or converting data from one file format to another format considered to be more stable. For the purposes of the UT Digital Repository, the new file is considered a derivative file. For additional information about the preservation of original and derivative files, please see the Preservation Policy.

Institutional Repository: An online database that archives, in digital format, the scholarly output of an institution and makes that output freely available.

Metadata: Information used to identify and describe works within the Repository. This information facilitates searching the Repository and includes the title, author, and date of issue, as well as technical information such as the file size and file format.

OAI-PMH: Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH) is a harvesting protocol for sharing data between online services such as repositories. Additional information about OAI-PMH can be found at <http://www.openarchives.org/pmh>.

Open Archival Information System (OAIS): An international standard (ISO 14721:2003) that describes the components and processes necessary for a digital archive.

OAster: Using OAI-PMH, OAster harvests descriptive data from repositories, digital libraries, and online journals to create a searchable catalog of digital resources. Additional information about OAster can be found at <http://www.oaister.org>.

Persistent URL: A permanent web address that will not change. The Repository is using the Handle System® for persistent URLs. Additional information about the Handle System® can be found at <http://www.handle.net>.

Recommended File Formats: These are file formats that the Repository considers stable and therefore compatible with long-term preservation efforts. In general, these formats have the following characteristics: openly documented; supported by a range of software platforms; widely adopted; lossless data compression or no compression; non-proprietary; and does not contain embedded files or embedded programs.

Repository Curator: The Repository Curator is responsible for managing all aspects of the Repository, including coordinating the various communities and collections within the Repository.

Storage Media Refreshment: Copying data from one long-term storage medium to another of the same type, with no change whatsoever in the bitstream (the binary form of the data).