3D: DOIs, Data Files, and Digital Commons
Outline

• Data
  • Getting started with data
  • Review & survey
  • Workshop documentation
  • Digital Commons, Dataverse, dPanther
  • Dataverse workflow
  • LibGuides
  • Pilot program implementation

• DOIs
  • DOI implementation
  • Crossref & DataCite
  • Metadata
  • Degree disciplines
  • DOI Workflow
Who is this workshop for?

• **People who are interested in:**
  
  • implementing and/or maintaining a data and/or DOI initiative
  
  • sharing their experience and knowledge on data and DOIs
  
  • experts on either who can contribute to the conversation!
Where to start with data

- Review of current data
  - ETD
    - Research data
  - RTD
    - Audio (fine arts)
    - Computer program

- Survey
Q1 - What form(s) of data have you collected while completing your thesis or dissertation?

- Numerical: High
- Multimedia: Moderate
- Text: High
- Models: Low
- Software: Moderate
- Discipline specific: Moderate
- Instrument specific: Moderate
- Other: Low
How much space does your data occupy?

1 GB
16g
10 MB
40GB
A lot
2 TB
I don't know
1MB
1gb
1 GB
1 GB
1TB
1TB
2 gbs
116 GB
300GB
3000GB
300MB
Q4 - Would you consider including your data with your ETD submission?

<table>
<thead>
<tr>
<th>Answer</th>
<th>%</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>50.00%</td>
<td>14</td>
</tr>
<tr>
<td>No</td>
<td>50.00%</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>28</td>
</tr>
</tbody>
</table>
ETD Data Pilot Program

Overview

In this pilot program, graduate students submitting Master’s Theses and Doctoral Dissertations in the Fall 2016 and Spring 2017 will be provided with the opportunity to archive finalized research data sets associated with their published ETD. These data sets will be stored in the Libraries iPanther system and virtually connected to the ETD in Digital Commons, the university’s institutional repository. A collaboration between the FIU Libraries and the University Graduate School, this pilot program seeks to

1. Establish data submission guidelines
2. Develop strategies for educating graduate students on data archiving
3. Assess data storage capabilities and needs for long term storage and access to data.

1. Data Submission Guidelines

Good documentation and metadata along with archiving in a preferred file format, can help ensure continued long-term access and re-usability of your data. When submitting data, students should submit the documentation, metadata and data as a compressed file in the supplemental field on their ETD submission form in Digital Commons.

Documentation

Good documentation of data can help ensure that data can be understood and interpreted by any user. Documentation should start at the beginning of a project and continue throughout the research.

When submitting data, students are required to include appropriate documentation in a readme.txt file.

The documentation should include:

- How data was created
- What the data means
- The data’s content and structure
- Any manipulations that may have taken place

Sample documentation:

Metadata

Metadata, is a subset of core data documentation. Though metadata standards vary across disciplines all metadata provides standardized structured information explaining the purpose, origin, time references, geographic locations, creator, access conditions and terms of use of data.

Along with the documentation of your project and data you must include a metadata.txt file that includes:

- Title: The name given to the data/dataset
- Title of Theses or Dissertation

- Creator: Who created the data?
- Subject: What is the data about? (usually expressed as keywords)
- Description: What is the resource about? Description may include but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content
- Date: When was the resource created?
- Format: What is the format of the data (PDF/A, Excel Spreadsheet, etc)
- Source: Where did the data come from
- Language: What language is the data in?
- Coverage: Coverage will typically include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity)
- Rights: Who owns the data?

Format

Choosing an appropriate format for the data is also an important aspect of long term access and reusability. Below are basic guidelines when preparing data for submission:

- Non-proprietary
- Open, documented standards
- Commonly used by your community/discipline
- Standard character encoding (ASCII, UTF-8)
- Unencrypted

Below is a table that outlines acceptable data formats. Data not in these formats will only be accepted if deemed appropriate upon review.

<table>
<thead>
<tr>
<th>Type of data</th>
<th>Acceptable formats for sharing, reuse and preservation</th>
<th>Other acceptable formats for data preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative tabular data with extensive metadata</td>
<td>SPSS portable format (.sav) delimited text and command (txt) file (.txt) containing metadata information</td>
<td>proprietary formats of statistical packages e.g. SPSS (.sav), Stata (.dta), MS Access (.mdb), acdtab</td>
</tr>
<tr>
<td>Quantitative tabular data with minimal metadata</td>
<td>comma-separated values (CSV) file (.csv) tab-delimited file (.tab) including delimited text of given character set with SQL</td>
<td>delimited text of given character set, only characters set present in the data should be used as delimiters (.bdi) widely-used formats, e.g. MS Excel (.xlt) .csv, MS Access</td>
</tr>
</tbody>
</table>
Digital Commons
- https://digitalcommons.fiu.edu/etd/3505/

Dpanther
- http://dpanther.fiu.edu/dPanther/home

Dataverse
- https://dataverse.org/about
<table>
<thead>
<tr>
<th>Functionality</th>
<th>Dataverse</th>
<th>dPanther</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metadata</strong></td>
<td>Supports discipline specific metadata including Geospatial Metadata, Social Science &amp; Humanities Metadata, Astronomy and Astrophysics Metadata, Life Sciences Metadata (more information). <a href="http://guides.dataverse.org/en/latest/user/appendix.html#user-appendix">http://guides.dataverse.org/en/latest/user/appendix.html#user-appendix</a></td>
<td>MODS/DC/MARC, limited vocabularies</td>
</tr>
<tr>
<td><strong>Individual Data Upload</strong></td>
<td>Yes. Built in web submission form</td>
<td>Yes, but requires development for simple submission form. Also, each user would need individual permissions management by staff.</td>
</tr>
<tr>
<td><strong>User Management</strong></td>
<td>Authentication via Shibolih, user restrictions customizable</td>
<td>Manage user permissions individually, unsure if authentication with LDAP/Shibolih possible or how much development it would take</td>
</tr>
<tr>
<td><strong>Auto-generated email notifications for users</strong></td>
<td>Yes - System generated emails to user when actions are taken in system</td>
<td>No - although it may be possible. Needs further investigation from developers</td>
</tr>
<tr>
<td><strong>Support</strong></td>
<td>Support: Support team available via email, Dataverse online community, Biweekly Community Call with Dataverse Development Team and other institutions, and Online chat (website says it offers training to developers, waiting on response from support to confirm this and any cost info)</td>
<td>Internal DCC and developer provided support.</td>
</tr>
<tr>
<td><strong>OAI-PMH Compliant</strong></td>
<td>Yes; can set up dataverse as a harvester from other OAI-PMH compliant data repositories. Exposes metadata through OAI-PMH and DDI</td>
<td>Yes - metadata is exposed via OAI-PMH for harvesting. No - metadata is not harvested into dPanther via OAI-PMH</td>
</tr>
<tr>
<td><strong>Built In Metrics</strong></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Built In Geolocation</strong></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Versioning</strong></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Access Restriction (IP or authentication)</strong></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Embargo</strong></td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>Automatically generated citations for individual data set</td>
<td>No</td>
</tr>
<tr>
<td><strong>SEO</strong></td>
<td>Some development required for enhanced discoverability <a href="http://guides.dataverse.org/en/4.7.1/installation/config.html#letting-search-engines-crawl-your-installation">http://guides.dataverse.org/en/4.7.1/installation/config.html#letting-search-engines-crawl-your-installation</a></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Migration</strong></td>
<td>Batch Export via API</td>
<td>dPanther developers can batch export if requested, however it takes time to script based on export criteria</td>
</tr>
<tr>
<td><strong>Additional functions for users</strong></td>
<td>Allows users to upload and save data before publishing, allows data uploader to share unpublished data with others.</td>
<td>support for Large Data upload via rsync and integration with an external application, the Data Capture Module (DCM).</td>
</tr>
</tbody>
</table>
Dataverse workflow

1. Ready to Submit
   - Yes: Sign Deposit Agreement
   - No: Consultation with Data Services Team

2. Consultation with Data Services Team
   - Sign Deposit Agreement
   - Mediated Deposit

3. Sign Deposit Agreement
   - Yes: Appraise (meets our policies?)
   - No: Email: Please correct and/or resubmit

4. Appraise (meets our policies?)
   - Yes: Review Files and Documentation
   - No: Email: New Dataverse Submission to Review

5. Review Files and Documentation
   - Yes: Finalize Submission/Publish Data
   - No: Email: Submission Finalized, here is DOI and any changes

6. Finalize Submission/Publish Data
   - Generate DOI

7. Generate DOI
   - Email Submission Finalized, here is DOI and any changes

8. Email Submission Finalized, here is DOI and any changes
   - Email: Submission Finalized, here is DOI and any changes

9. Email: Submission Finalized, here is DOI and any changes
   - Email: Submission Finalized, here is DOI and any changes

10. Email: Submission Finalized, here is DOI and any changes
    - Email: Submission Finalized, here is DOI and any changes

11. Email: Submission Finalized, here is DOI and any changes
    - Email: Submission Finalized, here is DOI and any changes

12. Email: Submission Finalized, here is DOI and any changes
    - Email: Submission Finalized, here is DOI and any changes

Key:
- Researcher Task
- Data Services Task
- Automated Task
• ETD Data Pilot Program
  https://library.fiu.edu/etd_data_pilot

• Data LibGuides
  https://fiu.libapps.com/libguides/
Steps to Implementation

- Committee - Research & Scholarship Strategies Group
- Review
- Survey
- Grad school/admin buy-in
- Product research & comparison
- Workflow
- Funding – Tech Fee 2017
- Pilot – to begin Spring 2019
- Workshops
- Documentation
- Implementation
Implementation of DOIs

A digital object identifier (DOI) is a unique and permanent online content identification, linked on the Internet. It is a unique alphanumeric string, assigned and governed by a registration agency, and identifies an object permanently, even if the object changes its location, its owner or other characteristics.
### DOI Registration Agencies used by FIU

<table>
<thead>
<tr>
<th>Agency</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crossref</td>
<td>Scholarly and professional research content. Journal articles, books, conference proceedings, etc. Reference linking and searchable metadata database. Journal Article: &quot;Quantum tomography: Measured measurement&quot;, Markus Aspelmeyer, <em>nature physics</em> January 2009, Volume 5, No 1, pp11-12; <a href="https://doi.org/10.1038/nphys1170">doi:10.1038/nphys1170</a></td>
</tr>
</tbody>
</table>

https://www.doi.org/RA_Coverage.html
• Research and review
• Workflow
• Implementation – by November 2017 all Electronic Theses and Dissertations registered with Crossref in Digital Commons
DOI Metadata

Registering journal articles?
Try Metadata Manager, now in Beta

These forms allow you to enter metadata and register DOIs. Review the help documentation for details.

Note: Your Crossref username and password will be required at the end of this process to submit data to the system. Please insure your browser allows JavaScript so that the data validation will function.

Step 1: Select Data Type

Data Type Selection

Step 2: Input the Dissertation metadata.

Dissertation information
Dissertation DOI:* URL:
Add Similarity Check as created URL

Author:
First Name:* Last Name:* ORCID: http://orcid.org/
Approval date:
Year:* Month:* Day:

Institutions:
Name:* Location: Acronym:
Degree:
Date:
Dissertation Identifier:
type: [d, i, c]:

Submit Dissertation Cancel

Most recent metadata schema
https://www.crossref.org/schemas/crossref4.3.0.xsd

Schema Documentation
http://data.crossref.org/reports/help/schema_doc/4.4.0/4.4.0.html

https://www.crossref.org/webDeposit/
Degree

Master of Science (MS)

Major/Program

Please Select Department

Major/Program

Select department if desired value is no longer available in the above menu.

Medical Laboratory Science
The stress-buffering potential of the work environment

Doreen C. DuMond, Florida International University

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>School/College</td>
</tr>
<tr>
<td>2</td>
<td>College of Arts and Sciences</td>
</tr>
<tr>
<td>3</td>
<td>Field/Majors/G Gradate Degrees/ Academic Degrees/ Academic Programs/ Academic Majors/ Campus Programs/ Modality Programs</td>
</tr>
<tr>
<td>4</td>
<td>Biology (offered jointly with FAU)</td>
</tr>
<tr>
<td>5</td>
<td>Chemistry (offered jointly with FAU)</td>
</tr>
<tr>
<td>6</td>
<td>Community Psychology</td>
</tr>
<tr>
<td>7</td>
<td>Computer Science</td>
</tr>
<tr>
<td>8</td>
<td>Economics</td>
</tr>
<tr>
<td>9</td>
<td>Environmental and Urban Systems (offered with the School of Technology)</td>
</tr>
<tr>
<td>10</td>
<td>International Studies</td>
</tr>
<tr>
<td>11</td>
<td>Mathematical Sciences</td>
</tr>
<tr>
<td>12</td>
<td>Physics (offered jointly with FAU)</td>
</tr>
<tr>
<td>13</td>
<td>Psychology (offered jointly with FAU)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specialization/Concentration/Major/Track</td>
<td>Degree</td>
<td>Exit Options</td>
<td>Campus Programs</td>
</tr>
<tr>
<td>Master</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master of Science in</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master of Science in</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master of Arts in</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

University Catalog 1982 - 1983 Miami, Florida
Florida International University
MASTER OF ARTS
in
AFRICAN NEW WORLD STUDIES

College of Arts and Sciences
African-New World Studies

Major/Program
African and African Diaspora Studies

MASTER OF SCIENCE
in
ENVIRONMENTAL AND URBAN STUDIES

College of Engineering and Applied Sciences
Environmental and Urban Systems

Major/Program
Environmental Studies
The image depicts a DOI (Digital Object Identifier) workflow. The workflow includes the following steps:

1. **Curating/Item selection**
2. **Pre-Processing**
   - Metadata Creation/Cleanup/Editing
   - Imaging Resolutions/File formats
3. **Assign Unique Identifier/Create Tracking system record**
4. **DOI Creation**
5. **Disseminate through IR/Digital Commons**
6. **File preservation in FDA**
7. **Reports sent to tracking sys**
8. **QC & finalize**
Questions and comments?

Kelley Rowan  
Digital Archives Librarian  
305-348-1079  
krowan@fiu.edu

Rebecca Bakker  
Digital Collections Librarian  
305-348-6485  
rbakker@fiu.edu

Jill Krefft  
Institutional Repository Coordinator  
jkrefft@fiu.edu

Ivy Torres-Morales  
Institutional Repository Associate  
itorresv@fiu.edu