Managing and Archiving Your Research Data
Introduction

1. Overview of Research Data and the Data Management Lifecycle - Strategies for managing your data

2. ETD Data Archiving Service – The why and how
What is Research Data?

Research data comes in many different forms and definitions of research data can vary based on research community.

- **NIH Definition**
  
  “recorded factual material commonly accepted in the scientific community as necessary to document, support, and validate research findings.”

- **NSF Definition**
  
  “determined by the community of interest through the process of peer review and program management.”

- **NEH Definition**
  
  “Materials generated or collected during the course of conducting research.”
Examples include:

- Images and Video
- Mapping/GIS Data/Geodatabases
- Numerical measurements
- Survey responses
- Focus group or interview transcripts
- Economic indicators and demographics
- Polls
- Computer modeling
- Simulations
- Observation and/or field studies
- Code or software
- DNA or Blood Samples
- Physical Collections (plant specimens)
What Data Is Not…

- preliminary analyses
- drafts of scientific papers
- plans for future research
- peer reviews, or communications with colleagues
- physical objects (e.g., laboratory samples)
- trade secrets
- commercial information
- materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law
Research Data Lifecycle

Planning
Collection and Creation
Processing and Analysis
Access and Reuse
Archiving and Preservation

Documentation and Metadata
Why is Data Management Important?

Data management throughout the data life-cycle will:

- Save you time and effort later
- Ensure research data and records are accurate and complete making data understandable to people outside the project
- Allow your data to be usable by others
- Encourages data citation to increase the impact of the research
Planning
Planning

Data Management Plan (DMP)

- A written document that describes the data you expect to acquire or generate during the course of a research project, how you will manage, describe, analyze and store those data, and what mechanism you will use at the end of your project to share and preserve your data.
Questions to ask during the planning phase

- What type of data will be produced?
- How much data will there be?
- Who will use the data now, and in the future?
- How long should the data be retained?
- Where will this data be retained?
- What file formats will be used?
- What documentation/metadata standards will be used?
- Are there privacy/security requirements?
The DMPTool helps researchers create data management plans (DMPs). It provides guidance from specific funders who require DMPs, but the tool can be used by anyone interested in developing generic DMPs to help facilitate their research (http://dmptool.org)
Funder Requirements

This guide provides information and links to the top FIU Federal Funding Agencies as well as private funders.

- [http://libguides.fiu.edu/PublicAccess](http://libguides.fiu.edu/PublicAccess)
Collecting and Organizing Data
Collecting and Organizing Data

Keeping track documents and datasets is critical while you are still conducting the research. There are two main strategies that can assist you in organizing your data while collecting research:

1. Use a Naming Convention for individual files

2. Grouping your files into meaningful datasets (File Structure)
Collecting and Organizing Data

Using a Naming Convention

These are some **best practices** when creating a naming convention:

- Use file naming *consistently*

- Make sure the names *clearly* represent what the file is (descriptive) – Identify the activity of project in the file name

- Use *short* informative words or phrases and try to keep file names under 25 characters

- Avoid using these symbols "/\:*<>&[]"

- Dates should be formatted like this: YYYYMMDD (e.g., 20150209) Put dates at the beginning or the end of your files, not in the middle, to make it easy to sort files by name

- Use underscores (_) not spaces to separate terms
What To Do:

- Consider including a Unique Identifier (Project Name or Grant #)
- Version #

*Be Consistent and Document!*
Naming Conventions

There are tools available that can assist you with retrospectively renaming your files based on a convention

- PSRenamer
- Renamer (For Mac OS Only)
- Wildrenamer
File Structure

- **Hierarchical:** Items organized in folders and subfolders
- **Tag-based/Piling:** Each item assigned one or more tags
- **Hybrid:** A mix of Hierarchical and tag-based

**Folder Structure Example: Piling (using a flat folder structure)**

Example 1: Study

Example 2: Study, Poster, Paper

When you use a looser, less hierarchical folder structure, your file names may need more detail.

**Folder Structure Example: Filing using a hierarchical structure**

- Project Name
- Surveys
- Data
- Analysis
- Visualizations
- Paper
- Instrument 1
- Raw
- Processed
- Instrument 2
- Poster
- Paper

When you use a hierarchical folder structure, you may want to create a readme file to remind yourself (and others) which files belong in which folders.

Heather Coates, 2015
Final Thoughts:

• Document your system and use it consistently
• Include important contextual information:
  Data
  Collection Method
  Collector…..

Example file structure:
/[Project]/[SubProject]/[Type_of_File]/[Data_Collector_Name]/[YYYYYMMDD]
Documentation and Metadata
Documentation and Metadata

- **Documentation** is “all about the use”
  - Makes reference to data in the context of their use

- **Metadata** is “data about data”
  - Describes key attributes of each data element or collection of elements
Project Level Documentation

- What's important to document?
  - Context of data collection
  - Data collection methodology
  - Structure and organization of data files
  - Data validation and quality assurance
  - Data manipulations through data analysis from raw data
  - Data confidentiality, access and use conditions
Data Level Documentation

- Variable names and descriptions
- Definition of codes and classification schemes
- Codes of, and reasons for, missing values
- Definitions of specialty terminology and acronyms
- Algorithms used to transform data
- File format and software used
Metadata Concept Map by Amanda Tarbet is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.
All metadata standards provide standardized structured information explaining

- Purpose
- Origin
- Time references
- Geographic locations
- Creator
- Access conditions and terms of use of data

For a full list of metadata standards you can visit: Digital Curation Centers Disciplinary Metadata List
http://www.dcc.ac.uk/drupal/resources/metadata-standards
Data Pilot Program Documentation and Metadata Template:

http://libguides.fiu.edu/ld.php?content_id=24464305
Documentation and Metadata

Important things to do while you collect or create your data

1. Make a note of all file names and formats associated with the project, how the data is organized, how the data was generated (including any equipment or software used), and information about how the data has been altered or processed.

2. Include an explanation of codes, abbreviations, or variables used in the data or in the file naming structure.

3. Keep notes about where you got the data.
Privacy Rights and Accessibility
Privacy Rights

- For research with human subjects, be sure to adhere to all rules and regulations set forth by FIU’s IRB.
- Researchers also must to adhere to privacy laws regarding personal health information as outlined by HIPPA.
Privacy Rights

- Anonymizing data sets
  - Direct Identifiers: names, addresses, telephone numbers, social security numbers, drivers license number etc.
  - Indirect identifiers: detailed geographic information, organization to which the respondent belongs, detailed occupational titles, places where respondent grew up.

More information about identifiable information can be found here:

[ICPSR Approach to Confidentiality](#)
Access Rights

Access/Permission Checklist

✓ Do you need to provide access to all the data produced under a grant?

✓ Does your data include any private information, medical information, or other information with possible confidentiality concerns?

✓ Does your research project have sufficient permissions necessary to disseminate the project data.
  ✓ Did any of the data come from a third party source?
  ✓ If so, did the project obtain permission to disseminate?
  ✓ Are there any restrictions you need to include in your permission statement?
Why provide access to your data?

Data are becoming valued scholarly products instead of a byproduct of the research process:

Benefits of sharing data:

1. Increase the impact of your work;
2. Lead to new collaborations or projects
3. Enables verification of your published results
Archiving and Preservation
Archiving and Preservation

**Archiving** is used when a file is to be preserved as-is, often at the end of a project and acts as a static (and usually final) record. [source - DataONE education module]

**Preservation** is the management and maintenance of digital objects (the files, or groups of files, that contain information in digital form) so they can be accessed and used by future users.
The format and software in which research data are created usually depend on how researchers choose to collect and analyze data, often determined by discipline-specific standards and customs.

Ensuring long-term usability of data requires consideration of the most appropriate software and file formats.
Formatting your Data

Use these basic guidelines when choosing a data format when submitting to a repository:

- Non-proprietary
- Open, documented standards
- Commonly used by your community/discipline
- Standard character encoding (ASCII, UTF-8)
- Unencrypted
Examples of preferred format choices:
- PDF/A, not Word
- ASCII, not Excel
- MPEG-4, not Quicktime
- TIFF or JPEG2000, not GIF or JPG
- XML or RDF, not RDBMS

UK Data Archive provides clear guidelines for acceptable formats for preservation and archiving. [http://www.data-archive.ac.uk/create-manage/format/formats-table](http://www.data-archive.ac.uk/create-manage/format/formats-table)
Converting your Data

- Data may need to be converted from the original format to a preferred data preservation format in preparation for long-term storage.

- Conversion is best done by the researcher familiar with the data, to ensure data integrity during conversion.
ETD Data Pilot Program
All graduate students submitting a thesis or dissertation through the graduate school are eligible to submit and archive their *finalized data* sets alongside their ETD. All data sets will be made openly accessible through the Library's dPanther system and Digital Commons.

Students will submit finalized data sets along with a ReadMe_[LastName].txt which serves as the metadata and documentation as a compressed file.
Submission Checklist

- **Is your data right for submission?**
  - Is this data related to your Thesis or Dissertation?
  - Is this data your original data?
  - Can this data be made publicly accessible through the IR?

- **Data Documentation:**
  - Do you have documentation for your data? If not, use the Readme_[LastName] to create appropriate metadata and documentation.
  - Is your data file naming convention clear and consistent to an outside user. Is your data grouped into datasets that make sense?

- **Sharing and Permissions:**
  - Have all collaborators, advisors, or other interested parties agreed to sharing the data publicly through the IR?
  - Is the data anonymized to protect any personally identifiable information?
ETD Data and other Data Management Resources

ETD Data Pilot Program
http://libguides.fiu.edu/etd_data_pilot

Research Data Management 101
http://libguides.fiu.edu/rdm

Data Management (FIU GIS Center)
http://maps.fiu.edu/gis/services/data-management
Love Your Data Week
February 13 - 17

- Love Your Data Week is a social media campaign to help raise awareness and build a community to engage on topics related to research data management, sharing, preservation, reuse, and library-based research data services. We will share practical tips, resources, and stories to help researchers at any stage in their career use good data practices.

- This year’s theme is data quality. Follow our daily posts on the FIU Libraries Blog, FIU Libraries Facebook page and the Digital Collections Center’s Facebook page. Use #LYD17 and #loveyourdata to join the conversation.
Contact Us

- Jill Krefft, IR Coordinator, jkrefft@fiu.edu
- Jamie Rogers, Assistant Director Digital Collections Center, rogersj@fiu.edu
- Brandie Thomas, ETD Coordinator, bthomas@fiu.edu
- Stephanie Brenenson, Graduate Studies Librarian, brenenso@fiu.edu
Thank you!