Data Management Plans 101

Dr. Nick Ruhs
STEM Data & Research Librarian
Florida State University Libraries
April 23, 2020
Overview

Defining Data

Research Data Lifecycle

Defining Data Management

Federal Funder Requirements

Resources to help build your DMP

Resources from FSU Libraries
What is research data?

“...the recorded factual material commonly accepted in the scientific community as necessary to validate research findings.”

(2 CFR 200.315(3))
## Examples of Research Data

<table>
<thead>
<tr>
<th>Obvious</th>
<th>Not so Obvious</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>Lab notebooks, field notes</td>
</tr>
<tr>
<td>Spreadsheets</td>
<td>Questionnaires, survey instruments</td>
</tr>
<tr>
<td>Audiotapes, Videotapes</td>
<td>Codebooks, data dictionaries</td>
</tr>
<tr>
<td>Photographs, Films</td>
<td>Methodologies, workflows</td>
</tr>
<tr>
<td>Models, Scripts, Code</td>
<td>SOPs, protocols</td>
</tr>
</tbody>
</table>
What is **not** research data?

- Preliminary Analysis
- Drafts of papers
- Plans for future research
- Peer reviews
- Communication with colleagues
Research Data Lifecycle
What is research data management?

Research Data Management (RDM) is a broad concept that includes processes undertaken to create organized, documented, accessible, and reusable quality research data.
Why Manage Research Data?

**CARROTS**
- Save time
- Increase citations
- Enhance reproducibility
- Credibility
- Preserve data

**STICKS**
- Required sharing from funders and journals
- Required data management plans for funding
- Prevent retraction

Adapted from: "Keeping Up With... Research Data Management", American Library Association, April 17, 2018.
http://www.ala.org/acrl/publications/keeping_up_with/rdm (Accessed March 5, 2020)
Common data management issues

Not doing it at all

Waiting until the end of the project to start

Proprietary file formats

Incomplete and/or incoherent data

Lack of data documentation
Federal Funder Requirements

“The Administration is committed to ensuring that…the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data” (Holdren 2013).
1. **Types of data**, physical samples or collections, software, curriculum materials, and other materials

2. Describe the **standards** to be used for all the data types anticipated, including **data or file format and metadata**

3. **Roles and responsibilities** of all parties with respect to the management of the data

4. **Dissemination methods**, data access after the grant ends

5. **Policies for data sharing, public access and re-use**, including re-distribution by others and the production of derivatives. Where appropriate, include provisions for protection of privacy, confidentiality, security, **intellectual property rights** and other rights.

6. Plans for **archiving data**, samples, software, and other research products
NIH Data Sharing Plans

a brief paragraph immediately following the Research Plan Section of the application form

1. Description of data collected/created
2. Data formats and types
3. Plan for disseminating research data
4. Timeline for disseminating research data
5. Description of data documentation
6. Description of Data Sharing Agreement (if applicable)
Other NIH Considerations

Budget and Budget Justification Sections:
• Applicants may request funds in their application for data sharing. If funds are being sought, the applicant should address the financial issues in the budget and budget justification sections.

Background and Significance Section:
• If support is being sought to develop a large database that will serve as an important resource for the scientific community, the applicant may wish to make a statement about this in the significance section of the application.

Human Subjects Section:
• If the research involves human subjects and the data are intended to be shared, the application should discuss how the rights and confidentiality of participants would be protected. In the Human Subjects section of the application, the applicant should discuss the potential risks to research participants posed by data sharing and steps taken to address those risks.
U.S. Department of Energy

1. Data Types and Sources
2. Content and Format
3. Data Sharing and Preservation
4. Protection
5. Rationale
General Data Management Plan Requirements

• Types of Data/Materials Produced
  – Observational, experimental, simulation, compiled
  – Tabular, audio, visual, etc.

• File Formats
  – Data should always be made available in non-proprietary file formats (.txt, .csv, etc.)

• Metadata Standards
  – How are you documenting your data/procedures?
  – Which standards are you using to document?
### File Formats

<table>
<thead>
<tr>
<th>Category</th>
<th>Format 1</th>
<th>Format 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>MS Word</td>
<td>PDF, TXT, HTML</td>
</tr>
<tr>
<td>Images</td>
<td>Photoshop</td>
<td>TIFF</td>
</tr>
<tr>
<td>Video/Media</td>
<td>Quicktime</td>
<td>MPEG4</td>
</tr>
<tr>
<td>Database</td>
<td>MS Access</td>
<td>DBF</td>
</tr>
<tr>
<td>Tabular Data</td>
<td>MS Excel</td>
<td>CSV</td>
</tr>
<tr>
<td>Presentations</td>
<td>MS Powerpoint</td>
<td>PDF (unencrypted)</td>
</tr>
<tr>
<td>Sound/Music</td>
<td>Windows Media</td>
<td>WAV (uncompressed)</td>
</tr>
</tbody>
</table>

*Be conscious about the risks of compressing your files or migrating to a file format that has different affordances than the original. See more at [http://guides.library.cornell.edu/ecommons/formats](http://guides.library.cornell.edu/ecommons/formats) (Slide Source: Data Curation Network)*
General Data Management Plan Requirements

• Dissemination/Data Sharing/Public Access
  – How will you ensure the data meets funder data sharing and public access guidelines?

• Policies for Archiving/Preserving Data
  – How will the data be preserved for long-term access?
  – Which data repository will be used?
  – Who can access the data?

• Data privacy, copyright, and IP rights (if applicable)
  – Does the project require an IRB application?
Defining Data Documentation

Information about the dataset itself that allows researchers to understand data in detail and enables other researchers to find, use, and properly cite data.

Types of data documentation:
- Metadata
- README files
- Codebooks and Data Dictionaries
- Research Methods and Analytical Strategies
Data Repositories

- Curate and archive in a way that is FAIR:
  - Findable
  - Accessible
  - Interoperable
  - Reproducible

- Citation

- Linking and versioning
Data repositories

Discipline specific: re3data

Institutional repository: Diginole

National repository: Data.gov
Using DMPTool to build your DMP

• Tool developed by data librarians and other data professionals to guide you in writing your DMP
• Developed in collaboration with internal and external partners, including funding agencies
• Provides guidance for many funding agencies, directorates, and divisions, along with links to additional resources, services, and help.
Using DMPTool to build your DMP

Let’s do an example!
Example DMPs

- DMPTool Public Plans: https://dmptool.org/public_plans
- NSF Biological Sciences Directorate:
  https://dmptool.org/plans/35572/export.pdf
- NIH Data Sharing Plans:
- DOE example plan:
  https://docs.google.com/document/d/1Hy8gLbWBtZwyJP-yNfQqFctlSs0bbN9x4tRL8Py4-9o/edit?usp=sharing
- IES (Institute of Educational Sciences):
  https://docs.google.com/document/d/1RLfCXhzQ_he5ZZokMfFQBTGM7cmSbj-eWpBIWeQBpXQ/edit?usp=sharing
FSU Libraries Resources

Data management plan consultations
Data management [LibGuide]
Data intellectual property and copyright questions
Finding and selecting data repositories
Archiving datasets into Diginole
Help with data management tools
Contact Information

Dr. Nick Ruhs
STEM Data & Research Librarian
nruhs@fsu.edu

Dr. Jesse Klein
Social Sciences Research & Data Librarian
jrklein@fsu.edu
Questions/Discussion

Thank you