Data Management Plans and Data Sharing

Ryan Womack and Aletia Morgan
Rutgers University Libraries
Research Data Services

October 2015
Recall - The Data Lifecycle

**Data Life Cycle**

- **Project**
  - **Associated objects:** Design documents; Grant proposal; Hypothesis; Methodology; Grant award; IRB protocol

- **Data Collection Hierarchy**
  - **Associated objects:** Field notes; Lab notes; Data corrections; Instrument calibrations

- **Experiment**

- **Trial**

- **Data**
  - **Associated objects:** Code book; Software; Article, dissertation, etc.

- **Analysis**

- **Publication**

- **Re-use and Citation**
Data Management Plans

• Many sponsored research agencies now require a "Data Management Plan" (DMP) as a component of any proposal

• The DMP is a formal document that outlines what the PI will do with data during and after completion of a funded research project

• The specific requirements for the DMP vary by funder, and by research subject, with most emphasizing preservation and data sharing

• Data Management Best Practices are the foundation of a good DMP proposal and effective Data Sharing
Data Management Plans

• NSF has required a DMP in all grant applications, since 2011

• Each Directorate has its own guidelines, some more specific than others

• The DMP is not scored as part of the grant application, but can provide a boost in terms of demonstrating impact

• Keeping up with evolving disciplinary standards is important
What’s in a Data Management Plan?

Funder requirements vary, but the project DMP should generally address:

• **Data Attributes** The types of data, samples, physical collections, software, curriculum materials, and other research products;

• **Data Documentation** The standards to be used for data and metadata formats and content;

• **Security Policies** Plans for data access and sharing including protection of privacy, confidentiality, security, intellectual property, or other rights, including the right to embargo data for a specified time period to allow first publication and thorough use of data;

• **Use Policies** Provisions for fair re-use, re-distribution, and the production of derivatives; and

• **Preservation** Plans for archiving data, samples, and other research products, and for preservation of access to them.
Benefits of a Good DMP

• Improved competitiveness for grant programs; a clear and complete Data Management Plan will support the project’s evaluation plan

• Enhanced PI efficiency – writing a DMP encourages the creation of a structured plan for managing research data throughout the life of the project and beyond

• Long-term protection and preservation of Rutgers’ sponsored research data
Examples of Support for DMP Development

**RUL Data Resources** – offers information about services and experts to support your data management efforts

Operated by the California Digital Library, the **DMPTool** supports general DMP development with some school-specific guidance
How to Get Started With a DMP

- **Review Funder and Directorate DMP Requirements**  Guides at Columbia and Oregon State track federal agency responses to OSTP mandate
- **Check for Specific Solicitation Requirements** Some programs include specific details or requirements for the Data Management Plan which will be outlined in the Call for Proposals
- **Contact us at RUL** The Rutgers University Libraries offers faculty and staff with specific expertise in data management and compliance with recent funder requirements for data management. We are available to you to validate your initial data management plans, identify appropriate campus or external resources for data management, preservation, and sharing.
- **Think of Data Management throughout the planning of your project** The best time to consider how you will collect, protect and preserve your data is early in the proposal development process, and integrate it into the planning of your research workflow.
DMP Summary

- Most funders emphasize the availability of data to other researchers for a limited time, rather than long-term preservation.
- Funding for data management activities is not very clear – “at no more than incremental cost and within a reasonable time”
- DMPTool provides a template, but think through your own project and needs.
- DMP approach is spreading to other granting agencies.
- Many examples of Data Management Plans are available on the web. Talk to your librarian for assistance!
Discoverable Data

- Publicly available archives such as Dryad, Dataverse, ICPSR, and more allow other researchers to easily discover and reuse data.

- Metadata provide standardized terminology for searching and discovering data matching defined characteristics. Unlike a data upload to a website, metadata provides a well-structured way for computer indexing of the data.

- Your discipline may or may not have well-defined metadata standards. Check with your librarian if you are in doubt.

- re3data is a directory of research data repositories that can be used to discover possible locations for you to deposit and share your data.
Dryad

- **Dryad** is a publication-linked data repository linked primarily with environmental and life sciences.
- An NSF-funded project in itself, members include BioMed Central, American Genetics Association, PLOS, Wiley, OUP.
- There are currently 77 integrated journals.
- Institutional sponsors deposit for free, affiliated journal costs vary, and unaffiliated users are currently charged $90 (extra charge for more than 10 GB of data).
- One of the more widely adopted data repositories.
ICPSR

- **ICPSR** is the oldest social science data archive (founded 1962)
- Contains more than 500,000 files of research in the social sciences
- Hosts 16 specialized collections of data in education, aging, criminal justice, substance abuse, terrorism, and other fields
- Curated collections place emphasis on important data
- **OpenICPSR** is a new service available to all researchers. $600 self-deposit **free to Rutgers** as an ICPSR member institution.
Dataverse

- Dataverse is a project originated at Harvard
- Both a platform and an application
- The Dataverse software can be downloaded to implement a locally-hosted data repository
- The Dataverse hosted at Harvard's IQSS allows individual researchers or groups to create accounts and upload data (up to 1 TB free). Currently supported by Harvard without any fees, but long-term outlook is uncertain.
- No disciplinary restrictions – used for many different purposes.
Discoverable Data

- There are many other disciplinary repositories, some very prominent in their specialties.
- Consult re3data to discover these.
- The re3data project has absorbed Databib, an earlier directory project.
- Look for certifications such as the Data Seal of Approval and ISO 16363 as assurance that the repository is using best practices. Only larger repositories tend to undergo certification.
DOIs

- In order to be used by subsequent researchers, research data must be both findable, and accessible. It must be stored with descriptive information that will allow it to be found, and then with a web address that will always take the user to the dataset.

- By publicly sharing your data via an established institutional or disciplinary repository, you will typically get a DOI (Digital Object Identifier) or other persistent URL. This will serve as a permanent pointer to the data.

- DOI's generated by Datacite require metadata, which Datacite collects into a searchable database.

- Thomson's Data Citation Index is creating another cross-disciplinary searchable database of data.

- Structured metadata, data publication, and indexing make data more discoverable.
Citing Data

• You can cite your data, and others’ data, with DOI’s. This is more precise and easier for others to work with than a vague reference to the data by title, author, or even citing a paper that uses the data.

• Data citation increases the impact of your research of your research!