PREPARING DATA MANAGEMENT PLANS

UNIVERSITY OF GEORGIA WORKSHOP

Sherry Lake
Data Management Consulting Group
Research Data Services
University of Virginia Library

November 5, 2014

Roadmap

1. What do we mean by data management?

2. Why should you manage your data?

3. What is a data management plan, and why do you need one?

4. How do you create a data management plan?

5. Create a DMP and receive feedback.

What do we mean by …

Managing your Research

- Ensuring **physical integrity** of files and helping to preserve them
- Ensuring **safety of content** (data protection, ethics, morality, etc.)
- **Describing the data** (via metadata) and recording its history (**provenance**)
- Providing or enabling **appropriate access** at the right time, or restricting access, as appropriate
- **Transferring custody** at some point, and possibly destroying

Managing Data in the Data Life Cycle

- Choosing file formats
- File organization & naming conventions
- Version control
- Access control & security

- Backup & storage
- File format conversions
- Document all project/file details
- Sharing and preservation
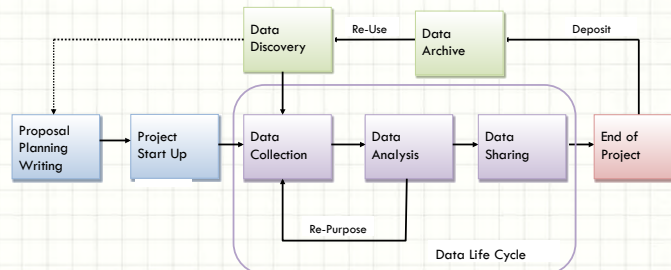
(Good) Data Management…

…helps research to be:

    Replicated and verified

    Preserved for future use

    Linked with other research products

    Shared and reused

…helps researchers:

    Meet funding requirements

    Increase visibility of research

    Save time and effort (avoid data loss)

    Deal with an ever-increasing amount of data

http://www.healthcare-informatics.com/article/guest-blog-data-management-challenge-unlocking-value-clinical-data-many-times-requires-enter
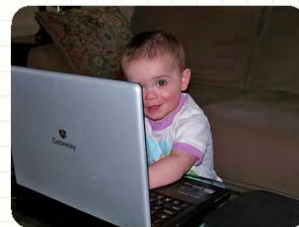
Who Cares about Data Management?

From Flickr by Redden-McAllister

Journal of Political Economy

JOURNAL OF PUBLIC ECONOMICS

Developmental Psychology

From Flickr by AJC1

www.rba.gov.au

What is a Data Management Plan?

- A comprehensive plan of how you will manage your research data throughout the lifecycle of your research project

AND

- Brief description of how you will comply with funder's data <u>sharing</u> policy
- Reviewed as part of a grant application

Types of Data Management Plans

- Document that is created to manage the data in you lab or project
- Document that is created at the start of a research project (required by funders or publishers)
  - Plan for data sharing
  - Plan based on funder specifications on how to manage your data

Who's Requiring Data Sharing or Data Management?

Require a Data Management Plan (DMP)	Require Sharing of Results – per a Data Policy
• National Science Foundation (NSF) • National Institutes of Health (NIH) • National Oceanographic and Atmospheric Research (NOAA) • Institute of Museum and Library Services (IMLS) • National Endowment of Humanities – office of digital humanities (NEH)	• Andrew W. Mellon • NASA • NEH – Preservation & Access • IES – Institute of Education Sciences • Wellcome Trust

This list is not inclusive.

Office of Science and Technology Policy

Increasing the Access to the Results of Federally Funded Scientific Research

Memo released February 22, 2013

To ensure that *"…direct **results** of federally funded scientific research are made available… Federal agencies investing in research and development (more than $100M in annual expenditures) must develop plans to support **increased public access to the results of research** …"*

http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Office of Science and Technology Policy

Update March 2014

- All required agencies have submitted plans
- OMB & OSTP reviewed and commented on plans
- Agencies working to identify infrastructure solutions to support their plans
- Interagency meetings soon to discuss solutions and strategies

http://www.whitehouse.gov/sites/default/files/microsites/ostp/OpenAccess_March-2014.pdf

What is in a Data Management Plan?

Parts of a (Generic) NSF Data Management Plan

I.   **Products of the Research:** The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project.

II. **Data Formats:** The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies).

III. **Access to Data and Data Sharing Practices and Policies:** Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

IV. **Policies for Re-Use, Re-Distribution, and Production of Derivatives.**

V.   **Archiving of Data:** Plans for archiving data, samples, and other research products, and for preservation of access to them.

Grant Proposal Guide (GPG) Chapter II.C.2.j
http://www.nsf.gov/pubs/policydocs/pappguide/nsf14001/gpg_2.jsp#dmp

Parts of a NEH-ODH Data Management Plan

Data Management Plans for NEH Office of Digital Humanities Proposals and Awards
http://www.neh.gov/files/grants/data_management_plans_2014.pdf

I.   **Roles and responsibilities:** The Data Management Plan should outline the rights and obligations of all parties with respect to their roles and responsibilities in the management and retention of research data.

II. **Expected data:** The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project. Describe the expected types of data to be retained.

III. **Period of data retention:** In their DMP applicants should address how timely access will be assured.

IV. **Data formats and dissemination:** The Data Management Plan should describe data formats, media, and dissemination approaches that will be used to make data and metadata available to others. Policies for public access and sharing should be described,.

V.   **Data storage and preservation of access:** The Data Management Plan should describe physical and cyber resources and facilities that will be used for the effective preservation and storage of research data.

How to Create a Data Management Plan?

Step-by-step wizard for generating DMP

Create | edit | re-use | share | save | generate

Open to community

Links to institutional resources

Directorate information & updates

http://dmptool.org

DMPTool Exercise

15 Minutes – Create a DMP

Work in teams

1. Identify a funding agency.
2. Use the funder's template to create a DMP.
3. Read through guidance provided by the DMPTool.
4. Record issues and questions for discussion.

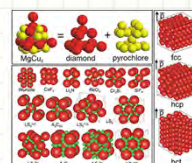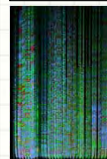# DMPTool Exercise

Discuss DMP Contents

Types of Data & Other Information

- **Types of data** produced

- Relationship to **existing data**

- How/when/where will the data be **captured or created**?

- How will the data be **processed**?

- **Quality** assurance & quality control measures

- **Security**: version control, backing up

- **Who will be responsible** for data management during/after project?

Data & Metadata Standards

- Identify the **formats of data files** created over the course of the project

- What **metadata are needed** to make the data meaningful?

- How will you **create or capture** these metadata?

- Why have you chosen **particular standards** and approaches for metadata?

Dublin Core® Metadata Initiative
Making it easier to find information.

TEI
< Text Encoding Initiative >

VRA CORE
a data standard for the description of images and works of art and culture

<ddi>
Data Documentation Initiative

Policies for Access & Sharing
Policies for Re-use & Re-distribution

- Are you under any **obligation to share** data?

- How, when, & where will you **make the data available**?

- What is the process for **gaining access** to the data?

- Who owns the **copyright** and/or **intellectual property**?

- Will you **retain rights** before opening data to wider use? How long?

- Are **permission restrictions** necessary?

- **Embargo periods** for political/commercial/patent reasons?

- **Ethical** and **privacy** issues?

- Who are the **foreseeable data users**?

- How should your data be **cited**?

Plans for Archiving & Preservation

- **What data** will be preserved for the long term? For how long?

- **Where** will data be preserved?

- What **data transformations** need to occur before preservation?

- What **metadata** or **documentation** will be submitted alongside the datasets?

- **Who will be responsible** for preparing data for preservation? Who will be the main contact person for the archived data?

Data Management Planning Resources

DMPTool
Build your Data Management Plan

http://dmptool.org – Helps you create a data management plan to meet grant requirements and identify UVA support resources and policies

Databib

http://databib.org – Helps you find an appropriate place to deposit your data

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

http://www.re3data.org/- Helps you find an appropriate place to deposit your data

QUESTIONS AND DISCUSSION?

Thank You

Email	shlake@virginia.edu
Twitter	@shLakeuva