AN INTRODUCTION TO DATA MANAGEMENT PLANS: An ACRL eLearning Webcast, May 27, 2014

Kiyomi D. Deards
University of Nebraska-Lincoln, kdeards2@unl.edu

DeeAnn Allison
University of Nebraska-Lincoln, dallison1@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/library_talks
Part of the Archival Science Commons, and the Scholarly Communication Commons

http://digitalcommons.unl.edu/library_talks/99

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Library Conference Presentations and Speeches by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
AN INTRODUCTION TO DATA MANAGEMENT PLANS

An ACRL eLearning Webcast, May 27, 2014

Kiyomi D. Deards, University of Nebraska-Lincoln

Dee Ann Allison, University of Nebraska-Lincoln
DEE ANN ALLISON

Professor

Director, Computer Operations and Research Services

Telephone: 402-472-3944

Email: dallison1@unl.edu
KIYOMI D. DEARDS  
Assistant Professor  
Liaison for Chemistry,  
Biochemistry, Physics and  
Astronomy  
Telephone: 402-472-2554  
Email: kdeards2@unl.edu
TODAY’S AGENDA

- Basics You Need to Know Before You Write
- Writing a Data Management Plan Checklist (For Consultants)
- Other Resources
- Most Common Errors When Writing a Data Management Plan
WHAT TYPE OF INSTITUTION ARE YOU FROM?

A. Academic
B. Special Library
C. Public
D. Library School Student
E. Other
“Therefore, the Office of Science and Technology Policy (OSTP) hereby directs each Federal agency that owns, maintains, or otherwise financially supports permanent scientific collections to develop a draft scientific-collections management and access policy within six months.

This policy applies to scientific collections, known in some disciplines as institutional collections, permanent collections, archival collections, museum collections, or voucher collections, which are assets with long-term scientific value.”
Check the funding agency’s requirements more than once during proposal preparation because requirements are changing constantly.

No one wants a grant denied because of an inadequate data management plan.
DATA MANAGEMENT – WHAT IT IS ALL ABOUT

Starts **before** project initiation and is followed throughout the life-cycle of the project

- During or upon project completion, data are stored (preserved) for:
  - Retrieval and re-use by the researcher or others
  - Sharing with a broader community (based on the funder’s requirements and the researcher’s data management plan)
WHAT IS IN A DATA MANAGEMENT PLAN?

Data management plans typically focus on the data management process:

- Data content and format
- Data collection, documentation and storage
- Data security and sharing (‘rights’)
- Data retention and preservation; includes task assignments for data life cycle
WHAT IS IN A DATA MANAGEMENT PLAN?

The disciplinary field and nature of the study will influence these decisions!
DATA FORMATS

Data exist in a variety of formats

• Ideally a format should be *nonproprietary* so you don’t need one particular software to read the data.

Select the format that will be most usable for the life cycle of the project!
DATA FORMATS

The PRONOM registry is a useful site about formats. (http://www.nationalarchives.gov.uk/PRONOM/Default.aspx#)

Consider data normalization so data can be interpreted correctly – this is particularly important for multi-institution research.

Carefully consider the pros and cons of file compression.
FILE NAMING CONVENTIONS

Use file naming conventions that support interoperability and re-use

• Use a period between the filename and file type extension.
• Use Universal Naming Convention when appropriate to designate a full path.
• Do not assume case sensitivity or use reserved characters in file names. These include < > : " / \ | ? *
FILE NAMING CONVENTIONS

• Do not use excessively long file names (some systems will truncate these).
• Use a file naming sequence that a person can read that is indicative of contents.
• File names should be unique, so files will not be accidently over-written.
DE-IDENTIFICATION OF DATA

- **Anonymization** is the removal of any identifiers, such as name, SSN, addresses and any other personal information that could be used to link data with an individual

- **Replace** sensitive information with machine generated codes
**DE-IDENTIFICATION OF DATA**

- **Aggregate** the data into related categories to provide analysis at a higher level than the level of the data collection. Census data are often aggregated.

- **Perturbation:** “Techniques for the release of data that change the data before the dissemination in such a way that the disclosure risk for the confidential data is decreased but the information content is retained as far as possible.” (source: [http://stats.oecd.org/glossary/detail.asp?ID=6950](http://stats.oecd.org/glossary/detail.asp?ID=6950))
METADATA: RECORDING DATA ABOUT DATA

- What are the relevant data and metadata standards, and controlled vocabulary in the field of research?
- How will descriptive metadata be captured (manually, automatically, both)?
METADATA: RECORDING DATA ABOUT DATA

- Who owns the data, can it be reused and how?
  (open access, FOIA requests, CC0, proprietary, etc.)
- How long will it be retained?
- How will the metadata and data be shared?
ACCOUNTING FOR DATA

- Local copy for researcher use (desktop, laptop)
- Local copy(s) for access and preservation
- Off-site copy for disaster recovery
- Access through a domain repository (for example: DRYAD)
DO YOU HAVE AN INSTITUTIONAL REPOSITORY?

A Yes
B No
Disciplinary repositories exist for many areas of study

http://oad.simmons.edu/oadwiki/Data_repositories

- **DataCite** - to find, access, and reuse data. Its listing, **DataBIB** (http://www.datacite.org/repolist), is a tool to identify and locate online repositories of research data. Users and bibliographers create and curate records that describe data repositories that users can search
DOMAIN REPOSITORIES

- **Digging into Data** - list of digital libraries, data archives and repositories
  (http://www.diggingintodata.org/Home/Repositories/tabid/167/Default.aspx)

- **re3data.org registry** - research data repositories from all academic disciplines
  (http://service.re3data.org/search/results?term)
ANY QUESTIONS SO FAR?
WRITING DMPS CHECKLIST

1. PANIC

2. Schedule a meeting to discuss the project and data management needs, allow at least an hour and a half.
CHECKLIST PT 2

3. If possible, everyone should have a copy of the grant proposal/project outline.

3.5 Ask if there are going to be any restrictions to access the data (copyright, patent, trademark, privacy, national security, Export Control, etc.).
4. Send the researcher some basic information on data management plans to review before your meeting.

5. Print out 3 different examples of data management plans so that you can refer to them during your consultations.
6. Go to the consultation.

6.5 If the researcher would like you to manage their data they should write your time into the grant.
7. Find an appropriate repository (if one exists). Registry of Research Data + Repositories: http://www.re3data.org/browse/

FigShare may be a viable alternative for those without a subject or institutional repository.

http://figshare.com/
CHECKLIST PT 6

8. Have a local repository?

Provide information on:

- policies
- requirements
- costs
- data deposit
  how to
9. Review the plan. If something confuses you it will confuse a reviewer.

Ensure that all samples, data file types, and methods of recording data are addressed.
10. After a successful consultation ask the researcher if their plan can be shared as a training tool within your organization.

11. Relax, you’ve done it!
Example of Boiler Plate Language:

To increase access to the published research that has been funded, the research collaborators will deposit peer-reviewed or pre-print manuscripts (with linked supporting data where possible) in UNL's open access repository, Digital Commons http://digitalcommons.unl.edu/. Other works, including presentations will also be made accessible via Digital Commons.

Automated Plan Generator: DMP Tool https://dmp.cdlib.org/
GOOD EXAMPLES FROM A DMP

1. No personal identifiers will be recorded or retained by the researchers in any form. There are no copyright or licensing issues associated with the data being submitted.
GOOD EXAMPLES FROM A DMP

2. The specified embargo period associated with the data being submitted extends from the projected conclusion date for initial research until six months after projected publication date for the findings.
GOOD EXAMPLES FROM A DMP

3. No data will reside on portable or laptop devices, and no other external media/format(s) will be used for data storage. Research data is backed up on a daily basis. The researchers are currently responsible for storage, maintenance and back-up of the data. The specific storage volume of the data being submitted will be not more than 1GB maximum.
BAD EXAMPLE FROM A DMP

Data will be available to anyone who desires access to our data. When possible, data will be made available online.

What is wrong with this example?
ANY QUESTIONS SO FAR?
Links to Other Resources

- University of Nebraska Lincoln, University Libraries Data Management Overview and Services
  http://libraries.unl.edu/datamanagement

- UNL’s Data Management LibGuide
  Detailed examples, explanations of terms, and links to resources on data management.
  http://unl.libguides.com/datamanagement

- O'Reilly Strata http://strata.oreilly.com/

- BigData by IEEE http://www.ieeebigdata.org/

- Data Conservancy http://www.dataconservancy.org/
## OTHER SOURCES OF INFORMATION

<table>
<thead>
<tr>
<th>Mailing Lists</th>
<th>Twitter Hashtags and Handles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Esip-preserve</td>
<td>#opendata</td>
</tr>
<tr>
<td>GEONET</td>
<td>#bigdata</td>
</tr>
<tr>
<td>PAMNET</td>
<td>#bigdatamgmt</td>
</tr>
<tr>
<td>Rdap</td>
<td>#bigdataprivacy</td>
</tr>
<tr>
<td>SLA-DST</td>
<td>Figshare</td>
</tr>
<tr>
<td></td>
<td>Research Data Summit</td>
</tr>
<tr>
<td></td>
<td>Open Science</td>
</tr>
<tr>
<td></td>
<td>Digital Science</td>
</tr>
</tbody>
</table>
OTHER INITIATIVES

COAR (Confederates of Open Access Repositories) has the goal of providing greater visibility and application of research through global networks of Open Access repositories

https://www.coar-repositories.org/

CHORUS (Clearinghouse for the open research of the United States) is a publisher initiated group for greater visibility and application of research through global networks of Open Access repositories

http://chorusaccess.org/
SHARE (SHared Access Research Ecosystem) is a higher education and research community initiative to ensure the preservation of, access to, and reuse of research outputs

https://sharewg.atlassian.net/wiki/pages/viewpage.action?pageId=8683527
MOST COMMON ERRORS

- Not maintaining 3 copies of data throughout the research process
- Not addressing secure storage during the research process to protect sensitive information
- Not addressing how all outputs, including publications or educational materials will be distributed/shared
- Forgetting to address a data or file type
- Not referring back to the body of the grant proposal when appropriate