

# DATA MANAGEMENT

# 101

---

ASERL Summertime Summit  
pre-conference webinar  
August 1, 2013

Jen Doty, Emory University  
Wendy Mann, George Mason University

# Outline

- What is Data Management?
- Motivation for providing services/support
- Service models: challenges for resources and staffing
- Preparing/training subject librarians
- Methods to raise awareness on your campus
- Institutional partnerships
- Assessments and analyses

# What is Data?

Wide variety across domains:

- Physical and life sciences—data are gathered or produced by researchers, such as by observations, experiments, or models.
- Social sciences—researchers may gather or produce their own data, or they may obtain data from other sources such as public records of economic activity.
- Humanities—data most often are drawn from records of human culture, whether archival materials, published documents, or artifacts.

# What is Data Management?

“**Data management** covers all aspects of handling, organising, documenting and enhancing research data, and enabling their sustainability and sharing.” ([UK Data Archive](#))

# What is Data Management?

“In the context of research and scholarship, **Data Management** refers to the storage, access and preservation of data produced from a given investigation.” ([Texas A&M guide](#))

# What is Data Curation?

“**Data curation** is critically important for a research institution because it provides two vital services:

(1) data are not merely stored, but are preserved in order to overcome technical obsolescence inherent in any storage system; and

(2) data are documented in such a way that they can be referenced in, and linked from, scientific publications and meet the requirements of funding agencies.” (UCSD – Research Cyberinfrastructure,

[Data Curation](#))

# Motivation

- Funding agency requirements for data management plans (DMP) and data sharing considerations
- Library as repository for scholarly output

# What is a Data Management Plan (DMP)?

- Outlines what the researcher will do with the data **during** and **after** the proposed research project, to ensure it is preserved and useful both now and in the future.

## **Description of Data**

The types of data to be produced in the course of the project include background demographic information, accuracy and reaction time for the proposed visual identification tasks, and the results from computational model simulations. The demographic data is needed for published reports to convey the characteristics of the subject population.

## **Standards for Data and Metadata Format and Content**

For preservation and long-term access, data collection will be accompanied with proper documentation and associated metadata. Files will include the data itself saved as Matlab (.mat) files, documentation files which a description of how the data was collected, and metadata in the form of the Matlab routines (Psychophysics toolbox) that administer each behavioral task.

## **Data Access, Sharing and Archiving**

The researchers associated with this study are not aware of any reasons that might prohibit the sharing of the data to be generated under this project for public use and potential secondary uses. The principal investigators retain the right for first use of the data. The principal investigators will work with their NSF Program Officer to identify public databases which would be appropriate to house data generated from this project as well as for long-term archiving of the materials. Until an appropriate public database is identified, access to the de-identified data will be provided by contacting the PI. For journal publications, summary forms of the data (means and standard deviations) will be published in tables, appendices, or online supplementary materials. Raw data will in principle be available for access and sharing as soon as is reasonably possible, normally not longer than one year after publication of the data. The raw data from the proposed research will be archived locally by the researchers, on regularly backed-up computers and will be preserved for at least three years beyond the award period, as required by NSF guidelines.

## **Data Confidentiality**

Research records will be kept confidential, and access will be limited to the PI and primary research team members. For each testing session, the recorded data will have any identifying information removed and will be relabeled with study code numbers. A database which relates study code numbers to consent forms and identifying information will be stored separately on password-protected computers in a secured, locked office. These computers are housed in research facilities in the Psychology Building at



# Why Prepare a DMP?

- Easier to preserve the data
- Prevents duplication of effort
- Can lead to new, unanticipated discoveries
- Increases visibility of research
- Makes research and data more relevant
- **Funding agency requirements:**

**NSF directorates  
and  
NEH Office of  
Digital Humanities:**  
1-2 page  
data management plan

**NIH  
(grants over \$500K):**  
1 paragraph  
data sharing plan

# Service models

Large research institution examples:

- Purdue, Cornell, UNC—coordinated and distributed services, working with research offices, central IT, data archives
- JHU, UCSD—include charge-back cost models for users of data management and curation services

# Service models

Medium and smaller levels of service:

- Library team approaches—Georgia State, Trinity University, James Madison
- Dedicated RDM position—Emory, Georgia Tech

# Preparing/training librarians

In-house training models:

- UVa - “re-tooling subject librarians” through brown-bag discussions and data interview participation
- Rutgers – internal course for librarians
- CU Boulder – one-day workshop for subject librarians

# Preparing/training librarians

External training options:

- ICPSR summer course – “Curating and Managing Research Data for Re-Use”
- online LIS credit courses (Wisconsin, SJSU)
- emerging data curation certification programs (UNC, Illinois, UNT)

# Resources and Links (1)

## Background and guidance:

- DataONE Best Practices Primer - <http://www.dataone.org/best-practices>
- UK Data Archive - <http://data-archive.ac.uk/create-manage>
- ICPSR guide - <http://www.icpsr.umich.edu/files/deposit/dataprep.pdf>
- MIT Libraries guide - <http://libraries.mit.edu/guides/subjects/data-management>
- ASERL/SURA step-by-step - [http://www.lib.ua.edu/wiki/sura/index.php/A\\_Step-By-Step\\_Guide\\_to\\_Data\\_Management](http://www.lib.ua.edu/wiki/sura/index.php/A_Step-By-Step_Guide_to_Data_Management)

## Tutorials and Training Guides:

- DataONE education modules - <http://www.dataone.org/education>
- UKDA training resources - <http://data-archive.ac.uk/create-manage/training-resources>
- MANTRA online training for PhD students - <http://datalib.edina.ac.uk/mantra/>
- University of Minnesota data management course for engineering graduate students - <http://z.umn.edu/datamgmt>

# Resources and Links (2)

Service models:

## *Large*

- Cornell University, RDM Service Group - <https://confluence.cornell.edu/display/rdmsgweb/About>
- Johns Hopkins University, Data Management Services - <http://dmp.data.jhu.edu/>
- Purdue University, Distributed Data Curation Center - <http://d2c2.lib.purdue.edu/>
- UCSD, Research Cyberinfrastructure - <http://rci.ucsd.edu/data-curation/>
- University of Virginia, Data Management Consulting Group - <http://dmconsult.library.virginia.edu/>

## *Medium & Small*

- Georgia State - <http://research.library.gsu.edu/datamgmt>
- James Madison - <http://guides.lib.jmu.edu/data> (see also article - DOI10.1080/19322909.2012.729394)
- Trinity - <http://libguides.trinity.edu/DataManage> (see also poster at [http://digitalcommons.trinity.edu/lib\\_faculty/27/](http://digitalcommons.trinity.edu/lib_faculty/27/) )

# Resources and Links (3)

Training models:

## *In-house*

- UVa - <http://www.slideshare.net/shlake/re-tooling-for-data-managementsupport>
- Rutgers – [http://iassistdata.org/downloads/2012/2012\\_e2\\_womack\\_etal.pdf](http://iassistdata.org/downloads/2012/2012_e2_womack_etal.pdf)
- CU Boulder – [http://www.slideshare.net/asist\\_org/poster-rdap13-dataday-participantdriven-training-for-library-research-data-services](http://www.slideshare.net/asist_org/poster-rdap13-dataday-participantdriven-training-for-library-research-data-services)

## *External*

- ICPSR summer course in data curation - <http://www.icpsr.umich.edu/icpsrweb/sumprog/courses/0149>
- online LIS credit courses
  - Wisconsin - <http://www.slis.wisc.edu/continueed-DataMgmt.htm>
  - SJSU -
- emerging data curation certification programs
  - UNC-Chapel Hill - <http://sils.unc.edu/programs/graduate/post-masters-certificates/data-curation>
  - Illinois - [http://www.lis.illinois.edu/academics/programs/specializations/data\\_curation](http://www.lis.illinois.edu/academics/programs/specializations/data_curation)
  - UNT - <http://icamp.unt.edu/icamp/content/project-abstract>



## **Part Two: Communication and Assessment**

- Outreach and Raising Awareness
- Institutional Partnerships
- Assessment and Analysis

# Outreach: Identifying Contacts

Who is involved in managing research?

- Office of Research, Vice President for Research
- Sponsored Programs
- Office of Research Integrity, Institutional Review Board (IRB)
- Information technology, research computing
- Others

Existing expertise: who is already involved in managing research data?

# Outreach: Toolkits & Resources

## Association of Research Libraries (ARL)

- E-Research. Outreach.  
<http://www.arl.org/focus-areas/e-research>
- Data Access, Management and Sharing  
<http://www.arl.org/focus-areas/e-research/data-access-management-and-sharing>

Association of College and Research Libraries (ACRL) report on research data services.

EDUCAUSE report on developing a research data management plan service.

# Outreach: Instruction

Holding information sessions or workshops advertises your services to the wider university community.

Information sessions on DMPs, DMPTool, repository services, etc.

Workshops on data management

# Outreach: Instruction Examples

## Cornell DMP Info Sessions

<https://confluence.cornell.edu/display/rdmsgweb/nsf-data-management-plan-info-session-slides>

## UNC at Chapel Hill Data Management Workshops

- Odum Institute + UNC

Libraries <http://www.irss.unc.edu/odum/contentSubpage.jsp?nodeid=667>

<http://guides.lib.unc.edu/content.php?pid=294213&sid=2874509>

## University of Virginia Libraries

<http://dmconsult.library.virginia.edu/training-sessions/>

# Outreach: Instruction Examples

## University of Oregon

- LIB 407/507: Data Management, 1-credit course, spring term 2013 offered at the University of Oregon.  
<http://library.uoregon.edu/node/3393>
- See ASERL “Two Institutions, Two Perspectives” webinar for more details: [http://www.aserl.org/wp-content/uploads/2013/06/U-Oregon\\_OSU\\_Digital\\_schol\\_Partnership.pdf](http://www.aserl.org/wp-content/uploads/2013/06/U-Oregon_OSU_Digital_schol_Partnership.pdf)

# Outreach: Institutional Buy-in

Once you identify stakeholders and start your outreach efforts buy-in should follow.

In addition to coordinating with relevant offices, knowledge of research policies at your institution is essential.

Buy-in fosters collaboration.

# Institutional Partnerships

Collaboration between the library and relevant offices is important for offering research data management support.

## Are there other partners?

- Selected academic departments, research centers or data archives.
- “Low hanging fruit” or natural partners
  - Cornell’s Research Data Management Services Group = Library + CISER + Advanced Computing



# Institutional Partnerships

## University Wide Partnership

- Research Data Stewardship at UNC [http://sils.unc.edu/sites/default/files/general/research/UNC\\_Research\\_Data\\_Stewardship\\_Report.pdf](http://sils.unc.edu/sites/default/files/general/research/UNC_Research_Data_Stewardship_Report.pdf)

## Regional Partnership

- AERL-SURA Model Language [http://www.aserl.org/wp-content/uploads/2013/01/ASERL-SURA\\_Model\\_Language\\_RDM\\_Policy\\_Language\\_FINAL.pdf](http://www.aserl.org/wp-content/uploads/2013/01/ASERL-SURA_Model_Language_RDM_Policy_Language_FINAL.pdf)

# Assessment and Analysis

When to assess and how

Consider an environmental scan part of the initial assessment process.

- Where have you identified gaps or partners?
- Anecdotal evidence

Formal assessment

- Survey instruments and interviews

# Formal Assessment: Examples

## Research Data Assessment

- Georgia Tech  
<http://www.library.gatech.edu/research-data/data-assessment>

## Data Interview

- University of Virginia Data Interview Initiative  
<http://dmconsult.library.virginia.edu/research-and-development-initiatives/>
- Emory Research Data Interviews.  
<http://guides.main.library.emory.edu/rdm-interviews>

# Formal Assessment: Examples

## Survey

- Cornell  
<http://escholarship.umassmed.edu/jeslib/vol1/iss2/1/>

## Content Analysis

- Georgia Tech analysis of submitted DMPs:  
[https://smartech.gatech.edu/bitstream/handle/1853/44391/NSF\\_DMP\\_Content\\_Analysis.pdf](https://smartech.gatech.edu/bitstream/handle/1853/44391/NSF_DMP_Content_Analysis.pdf)

# Assessment and Analysis

## Resources to guide you

- Data Curation Profiles  
<http://datacurationprofiles.org/>
- Participating in the E-Science Institute  
<http://duraspace.org/e-science-institute>
  - Time and financial commitment

What will you do with your results?

# Additional Resources and Links

Association of College and Research Libraries (ACRL). *Academic Libraries and Research Data Services: Current Practices and Plans for the Future*.

[http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir\\_Birch\\_Allard.pdf](http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf)

Charles W. Bailey. *Digital Curation and Preservation Bibliography: Digital Data*. <http://digital-scholarship.org/rdcb/rdcb.htm>

Digital Curation Centre. Data Asset Framework (DAF). <http://www.data-audit.eu>

EDUCAUSE. *Developing a Research Data Management Plan Service*.

<http://www.educause.edu/library/resources/developing-institutional-research-data-management-plan-service>

Sherry Lake. Conducting an Environmental Scan: Who's Important on Your Campus. DMPTool Webinar Series 4.

<http://www.slideshare.net/UC3/dmptool-webinar-4>

Susan Wells Parham, Jon Bodnar, and Sara Fuchs, "Supporting tomorrow's research: Assessing faculty data curation needs at Georgia Tech," *College and Research Libraries News*, 73 (2012):10-

13. <http://crln.acrl.org/content/73/1/10.full?sid=89c2612a-89d3-43c3-88a0-a9c23a056eff>

Graham Pryor, ed. *Managing Research Data*. London: Facet Publishing, 2012.

# Thank You!

- Contact information:
- Jen Doty
  - Emory University
  - [jennifer.doty@emory.edu](mailto:jennifer.doty@emory.edu)
- Wendy Mann
  - George Mason University
  - [wmann@gmu.edu](mailto:wmann@gmu.edu)