
JAKE CARLSON
BARRIE HAYES
SUSAN PARHAM
BRIAN WESTRA

IASSIST ANNUAL CONFERENCE
JUNE 02, 2011
NSF Requirements

- the types of data and other materials to be produced in the course of the project;
- the standards to be used for data and metadata format and content;
- policies for access and sharing;
- policies and provisions for re-use, re-distribution, and derivatives;
- plans for archiving and preservation of access
UNC Libraries: Building Data Expertise

- Spring 2008 – UNC science librarians visit Purdue, University of Chicago, Cornell
- June 2008 – Local E-Science Symposium
- Fall 2008 – University Library hires Data Services Librarian
- Spring 2009 – Science/data/GIS Librarians self organize into Data Management Working Group
- July 2010 – Libraries Data Management Committee charged
UNC Libraries Data Management Committee

City & Regional Planning Librn (Doty)
HSL Bioinformatics Librn (Hayes)
Data Services Librn (Hayslett)
GIS Librn (Henley)
Head of Science Libraries (Kamarei)
Geological Sciences Librn (Kennard)
Head, Special Formats & Metadata (McBride)
Electronic Records Archivist (O’Meara)
HSL Catalog and Metadata Coordinator (Richardson)
Biology Librn (Romito)
UNC-Chapel Hill Research:
FY2010 Funding Sources

- NIH: 52.72%
- NSF: 4.54%
- Other Federal: 16.17%
- Other (Non-profit, state, etc): 26.6%
<table>
<thead>
<tr>
<th>Department</th>
<th>School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthropology</td>
<td>Institute Of Marine Sciences</td>
</tr>
<tr>
<td>Biochemistry And Biophysics</td>
<td>Kenan-flagler Business School</td>
</tr>
<tr>
<td>Biology</td>
<td>Marine Sciences</td>
</tr>
<tr>
<td>Biomedical Engineering</td>
<td>Mathematics</td>
</tr>
<tr>
<td>Biostatistics</td>
<td>Molecular Biology</td>
</tr>
<tr>
<td>Carolina Center For Genome Sciences</td>
<td>Philosophy</td>
</tr>
<tr>
<td>Carolina Population Center</td>
<td>Physics And Astronomy</td>
</tr>
<tr>
<td>Center For Developmental Science</td>
<td>Political Science</td>
</tr>
<tr>
<td>Chemistry</td>
<td>Psychology</td>
</tr>
<tr>
<td>Computer Science</td>
<td>Renaissance Computing Inst</td>
</tr>
<tr>
<td>Dice Center</td>
<td>School Of Information And Library Science</td>
</tr>
<tr>
<td>Environmental Sciences And Engineering</td>
<td>School Of Pharmacy</td>
</tr>
<tr>
<td>Frank Porter Graham Child Dev Center</td>
<td>Sociology</td>
</tr>
<tr>
<td>Geography</td>
<td>Statistics And Operations Research</td>
</tr>
<tr>
<td>Geological Sciences</td>
<td>Unc Institute For The Environment</td>
</tr>
<tr>
<td>Graduate School</td>
<td>Urban Studies</td>
</tr>
</tbody>
</table>
NSF DMP Requirement: Opportunities

Opportunities for data librarians / library

- Raise awareness of data management issues and concerns
- Connect with researchers
- Build awareness of library and campus resources and services for data management
- Lead campus partners in coordinating existing resources and services for data management
- Apply, demonstrate and expand library expertise in data arena
Awareness Building among Researchers

Data management information sessions
- Getting the word out
  - Direct email to UNC NSF grantees
  - Broadcast via key partners and communication channels, e.g.
    - Department heads and research center directors (CFE)
    - Odum Institute list of social science researchers
    - Research support newsletter
    - Research administrators listserv
DMP Session Content

- Review NSF Data Management Plan Requirements
- Key Terms - data, data life cycle, metadata, repository
- Issues and Considerations
- Resources for Researchers
- Discussion and Questions
# Data Management Session Attendance

<table>
<thead>
<tr>
<th>Information Session</th>
<th>In class</th>
<th>Online</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding Data Management Plans &amp; Funder Requirements (1/9/10)</td>
<td>30</td>
<td>72</td>
<td>102</td>
</tr>
<tr>
<td>UNC Resources: Dataverse &amp; Carolina Digital Repository (11/16/10)</td>
<td>20</td>
<td>29</td>
<td>49</td>
</tr>
<tr>
<td>Data Management Plans for the Life Sciences and Dryad (1/31/2011)</td>
<td>17</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Considerations and Strategies for Handling Sensitive Research Data (3/29/11)</td>
<td>7</td>
<td>40</td>
<td>47</td>
</tr>
<tr>
<td>Considerations and Strategies for Handling Sensitive Spatial Data (5/10/11)</td>
<td>16</td>
<td>15</td>
<td>31</td>
</tr>
<tr>
<td>Attendance Totals Across 5 Sessions</td>
<td>90</td>
<td>156</td>
<td>192</td>
</tr>
</tbody>
</table>
Awareness Building among Library Staff

- Presented DMP session content at staff meetings
  - University Library public services+ collection dev librarians
  - Health Sciences Library liaison librarians

- Staff Guide: “Data Management for Subject Librarians”
  - Includes links to NSF and NIH data policies
  - Many resources for further reading
Outreach: New Resource for Researchers

Research Data Toolkit
Resources for Writing a Data Management Plan

Links
- Toolkit Home
- About the Toolkit
- Campus Resources
- Confidential Data
- Contact Us
- Courses
- FAQs
- Metadata Standards/Schema
- Needs Assessment

General Information

On the NSF requirement:
The (very general) guidelines for data management plans are presented in Section 1: Special Information and Supplementary Documentation of the Grant Proposal Guide, Chapter I - Proposal Preparation Instructions. They are the second bullet item in the linked section.

Data management requirements and plans specific to the Directorate, Office, Division, Program, or other NSF unit, related to a proposal are available at http://www.nsf.gov/dias/sgp/dmp.jsp. If guidance specific to the program is not available, then the general requirements listed at the first web site (i.e., in Section 1, linked above) apply.
Research Data Toolkit

Site resources:

- Templates for formulating DMPs
- Example language for DMPs
- Profiles of campus resources
- DM guides and links
- Contact listserv for consultation about plans
1334 visits, 3566 page views since 1/9/11

Max Visits
131, 1/30/11 – 2/5/11
Research Data Toolkit Usage

- 603 unique visitors, max = 93 the week of 1/16/11
- 2.67 pages/visit (avg); 2.15 minutes on site (avg)
- Returning visitors = 54.7% of visits (New = 45.3%)
- 58.6% bounce rate
- Top viewed pages besides toolkit home page
  - Courses page
  - Templates (for DMPs) – 50% selected Example Language for DMPs link
  - Example DMP language page
  - About the Toolkit
  - Campus Resources
  - Metadata Standards/Schema
  - Repositories
  - Confidential Data
  - FAQs
  - Questions to Assess Data Management Needs
Challenges for data librarians / library

- Communicating importance of research data curation, sharing, plans more widely
- Measuring impact and effectiveness of outreach efforts
- Providing, scaling data services
- Building a coordinated research data stewardship effort
New Opportunities

- Collaboration with Provost’s Task Force on the Stewardship of Digital Research Data (data needs assessment)
- DMP consultation requests
- Build on new connections with key units and personnel – opportunity to coordinate research support services and resources broadly
Consultation and Instruction

Brian Westra

University of Oregon
Consultation Opportunities

1. Implement curation-friendly practices and tools
2. More competitive grant proposals
3. Build institutional capacity for data curation
4. Leverage DMP requirements for conversations about data curation, resources, tools, pilot cases, roles
5. Build relationships within the team and with researchers and other support units
Phased Approach

1. Create the team
2. Develop web content for the elements of a DMP
3. Outreach - announce DMP requirement, support materials and contacts, and introductory meetings
4. Introductory sessions/basic instruction
5. Consultations
# DMP Support Team

<table>
<thead>
<tr>
<th>Person</th>
<th>Title &amp; organization</th>
<th>Support scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karen Estlund</td>
<td>Head, Digital Library Services</td>
<td>Metadata, file formats, IR</td>
</tr>
<tr>
<td>Kira Homo</td>
<td>Electronic Records Archivist</td>
<td>File formats, metadata</td>
</tr>
<tr>
<td>Brian Westra</td>
<td>Science Data Services Librarian</td>
<td>Coordination, metadata, file formats, tools/systems</td>
</tr>
<tr>
<td>Sean Sharp</td>
<td>Research and Instructional Technology, Campus Information Services</td>
<td>File storage and backup, pre-grant IT consults</td>
</tr>
<tr>
<td>Cathleen Leue</td>
<td>College of Arts and Sciences Information Technology</td>
<td>File storage and backup</td>
</tr>
<tr>
<td>Chuck Williams</td>
<td>Technology Transfer</td>
<td>Technology transfer</td>
</tr>
<tr>
<td>Meg Rowles/Sherry Lealess</td>
<td>Office for Responsible Conduct of Research</td>
<td>RCR education</td>
</tr>
</tbody>
</table>
1. Applicable to DMP, and informative for general data management needs
2. Based on review of other institutional sites
3. Team members authored or contributed to subtopics, i.e., file formats, metadata standards, tech transfer
Outreach

Deans and Center & Institute Directors

VP for Research

Departmental Grant Administrators

Via the support team and their contacts and distribution lists/newsletters/emails
Web pages link to ticketing system (RT) email address (grantconsult@ithelp.uoregon.edu)

- Allows the two leads (Brian and Sean) to cover respective issues
- Hand off questions
- Know what has already been covered
- Track all responses; “open” and “close” tickets
Consultations - Outcomes

Consulted on 13 completed or in-progress plans
Leads to questions about post-consultation services and infrastructure
Backup and storage issues are being addressed
Investigating data-specific options for the IR
Consultations - Challenges

Unit and inter-unit politics and scope creep (individual – unit – team – NSF)
Sub-optimal solutions and/or conflicting advice
Impersonal and clunky ticketing system
Questions for which we currently have no clear answer
Consultations - Opportunities

Leads to questions about post-consultation services and infrastructure

- Changes in technology, standards
- Annual reports – annual review of plan?
- Implementation support?
  - Metadata
  - Tools/software
  - Preservation/archiving
  - Research team membership
- Iteration of web site
RCR is required by NIH, NSF

NIH: all trainees, fellows, participants and scholars receiving support... must receive instruction in responsible conduct of research.

NSF: all undergraduate students, graduate students and postdoctoral researchers... receive training in the responsible conduct of research.

http://orcr.uoregon.edu/index.cfm?action=rcr
RCR - Outcomes

Successful class on data management with Psychology department

Other departments will implement the modules next academic year

- Biology
- Chemistry
- Human Physiology
Good entre for working with graduate students

- illuminates common issues
- defines topics for conversations with Office of Research, faculty (i.e., institutional policies and guidelines)

Stronger relationship with more of the research (grant) support and oversight staff
Panel Questions

- What awareness building strategies have you found to be most successful?
- What are the most common issues you are encountering in your consultations?
- How are you planning to follow up with researchers post consultation?
Partnerships

SUSAN PARHAM

GEORGIA TECH UNIVERSITY
Georgia Tech Example: Partnerships

Opportunity: build partnerships within the library and across campus (and beyond) to develop a comprehensive response to mandated data management and sharing requirements.

Outcomes & Challenges:
1) Internal Library Partnerships
2) External Partnerships
# Internal Library Partnerships

| Research Data Project Team | • Research data assessment survey (including questions re: DMPs)  
|                           | • Data management guide & best practices |
| Repository Team           | • Repository policies for research data  
|                           | • DMP text for repository access & preservation |
| Subject Liaisons          | • Outreach and feedback for data assessment & NSF DMP guidelines & workshops |
| User Engagement Dept      | • Print brochures & web graphics for outreach |
Internal Library Partnerships: Challenges

• Working across organizational boundaries
• Articulating roles & responsibilities of various teams & team members
• Keeping multiple stakeholders informed (e.g., Outreach Council, Scholarly Communication Council, Repository Team)
• Building consensus on the importance of research data curation without top-down incentive for staff participation (“results without authority”)  
• Finding the resources to devote to outreach, consultation, research, strategic planning, etc.
## Campus Partnerships

| Campus Information Technology | • IT infrastructure  
|                              | • Institutional policy & procedures  
|                              | • Support services  
| Office of the Executive Vice President of Research | • Institutional policy & procedures  
|                                                | • Support services  
| Office of Sponsored Programs | • Institutional policy & procedures  
|                                   | • Support services  
| Library                            | • IT infrastructure  
|                                   | • Support services (including managing metadata)  
| Additional Partners               | • Institutional policy & procedures  
|                                   | • Support services  

*Creating a Data Management Framework*
Australian National Data Service
## Campus Information Technology

<table>
<thead>
<tr>
<th>Individual College IT</th>
<th>Office of Information Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>• College-level technology support for research projects</td>
<td>• Enterprise services</td>
</tr>
<tr>
<td>• Data storage during research</td>
<td>• Campus networking &amp; infrastructure</td>
</tr>
<tr>
<td>• Data backup during research</td>
<td>• High Performance Computing</td>
</tr>
<tr>
<td></td>
<td>• Data storage (no digital preservation services)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IT Governance</strong></td>
<td></td>
</tr>
<tr>
<td>• Strategic planning for institute-level IT projects</td>
<td></td>
</tr>
<tr>
<td>• Resource allocation</td>
<td></td>
</tr>
<tr>
<td>• Campus-wide stakeholder engagement</td>
<td></td>
</tr>
</tbody>
</table>
Central administration for all research, economic development, and related support units at GT

- Georgia Tech Research Institute
- Enterprise Innovation Institute
- Interdisciplinary Research Centers
- Georgia Tech Research Corporation

Georgia Tech does not currently have widely published, up-to-date research data policies.
Office of Sponsored Programs

Educational, informational and technological support services for research administration.

Library collaborations include:

• Research Admin Buzz (research administrators)
• Outreach for research data assessment survey
• NSF DMP workshops
• RCR courses (data stewardship)
• Research News (articles in OSP newsletter)
## Additional Campus Partners

| Graduate Research Ethics Center | • Responsible Conduct of Research courses (ethical data management)  
| | • Information sharing and current events  
| | • IP issues  

| Office of Institutional Research and Planning | • Support for strategic planning & policy-making  
| | • Strategic planning for institute wide data stewardship  

| Office of Legal Affairs | • Intellectual property issues & data policies  

---

[Georgia Tech Library logo]
Campus Partnerships: Challenges

- Lack of awareness among faculty and administrators on the importance of sharing and preserving research data, and of developing data management plans
- Effectively communicating the need for data curation to the Institute's broad, diverse audience
- Difficulty in coordinating the relevant campus units to implement data curation services and infrastructure
- Shortage of dedicated personnel hours for outreach and consultation regarding research data management, sharing, and preservation
- Lack of funding to support new services, including data repository management and development
Opportunity

**Strategic Technology Investment Collaboration**
- Institute-level IT strategic planning
- Data Curation Virtual Community Task Force
- 1 of 4 pilot projects
- Library: idea champion & team leader
- Data stewardship white paper

Means to achieve:
- Stakeholder engagement
- Resource allocation
Tools & Infrastructure

JAKE CARLSON

PURDUE UNIVERSITY
Background on Purdue

2004: Purdue Interdisciplinary Research Initiative revealed many data needs on campus

2006: Founded D2C2 to further investigations, organize research and leverage collaborations
2010: The Data Curation Profiles Toolkit released

Who’s willing to share their research data, with whom, when, and under what conditions?

The Data Curation Profile is a means to determine:
• Information about a particular data set
• What a researcher is doing to manage / curate the data set
• What a researcher would like to do with the data.
Initial response: DMP steering committee

Jim Mullins  
Dean of Libraries

Richard Buckias  
Vice-President of Research

Gerry McCartney  
Chief Information Officer
The Data Curation Profile is not designed to produce a Data Management Plan, however it could be used as a foundation to develop a more specific tool.
Background Research

- “Unpacking” the NSF requirements
- Review of the content of existing data management plans
- Review of existing guides on creating a DMP
- Review of the information gathered from our Data Curation Profiles work, and other faculty-librarian collaborations.
- Examination of DCP questions in light of DMP requirements
Interviews

- Working with OVPR, four proposals were selected:
  - Engineering Education
  - Agronomy
  - Physics / Electrical & Computer Engineering
  - Pharmacy

- Interviews are conducted:
  - Multiple faculty / Multiple interviews
  - Sponsored Programs personnel and Subject Librarians also attend interviews
Metadata Standards
Metadata is “structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource”\(^2\) (such as a data set). “A metadata record is a file of information which captures the basic characteristics of a data or information resource. It represents the who, what, when, where, why and how of the resource.”\(^3\)

8. How will metadata be generated and captured for each of your data sets?
9. Are you aware of any metadata standards specific to your field that could be used for your data sets (e.g., Dublin Core [DC], Resource Description Format [RDF], Federal Geographic Data Committee [FGDC], Directory Interchange Format [DIF], Ecological Metadata Language [EML], Minimum Information About a Proteomics Experiment [MIAPE], and the Data Documentation Initiative [DDI])?
10. If there is not a metadata standard, what metadata will you need to generate so that others in your field will be able to find, understand, and make use of your data?
11. Who are you research team will be responsible for ensuring metadata standards are followed?

http://purl.lib.purdue.edu/d2c2/dmp_saq
Guides

Data Management Plan (DMP) - Overview

What should be included in a Data Inventory of Your Project?

- Section 1a - Conducting a Data Inventory of Your Project
- Section 1b - Selection and Appraisal Policies
- Section 2a - Data Standards
- Section 2b - Metadata Standards
- Section 2c - File and Directory Naming Conventions
- Section 3a - Components of Access & Sharing policies
- Section 3b - Privacy and Confidentiality / Policies
- Section 3c - Security Issues / Policies
- Section 3d - Intellectual Property & Copyright Issues / Policies
- Section 4a - Attribution, Acknowledgement, and Citing Data
- Section 5a - Preservation Issues and Strategies
- Section 5b - Selection and Appraisal for Preservation

DMP - Section 2a

2a. Data Standards

Data Standards are typically the file format in which your data is generated. (Examples of standard data formats include: XML, ASCII, CSV, MySQL, netCDF, etc.)

Agreeing to particular data standards will not only allow you to share data with all project personnel, but will also allow others to view or use your data once it is shared. The more common the data standard, the more universal use and acceptance your data is likely to have. Using a standard data format will also decrease the likelihood of your data becoming obsolete because it can no longer be read and will help others archive your data for long-term preservation if necessary.

Conforming to data standards might be as simple as making sure your project personnel are not saving spreadsheets from three different versions of Microsoft Excel. It might also mean having to convert data taken from one piece of software and converting it to another data format so it can be easily shared with others (such as converting a word processing file into a more standardized ASCII or XML data standard).

Possible Questions to Consider:

- Will your data conform to any standard formats already agreed upon by those in your field or discipline?
  - Whether you will be using them or not, consider what data standards are currently accepted within your field. It will be easier to compare and share data with others if your data is similar to others in your discipline.

- How accessible are your chosen data standards for others to read and/or re-use your data?
  - Consider whether the data standards you have chosen are Open Source or may require specific software or instrument to read it.

- Who on the project will be responsible for properly applying these data standards?
  - Those responsible should ensure the agreed upon data standards are properly and consistently applied. Following-up to make sure the data is properly formatted for later dissemination and re-use by others is also an important factor.

Resources:

Data Formats Table (data formats currently recommended by the UK Data Archive for long-term preservation of research data)
http://www.data-archive.ac.uk/create-manage/format/format-tables

Open format - listing of open (non-proprietary) file formats from Wikipedia.
Results

- Faculty, SPS personnel and librarians drafted 3 data management plans
- Experience in interviewing faculty is used to inform the development of a NSF DMP workshop
  - sponsored by OVPR but featuring the Libraries
- Additional faculty reach out to the Libraries for help crafting DMPs
Building a Data Repository

What Does My NSF Data Management Plan Need to Address?

There are many things that need to be included in your data management plan. Watch our video tutorials, read the step by step instructions, or view a completed data management plan to help you get started. You can also click the link below to start now.

Start Your Data Plan

- **Read the DMP Requirements**
  Read the detailed NSF requirements for your data management plan. The requirements are very specific and should help to clear up any confusion regarding your own data management plan. [Get Started](#)

- **Create a Group for Your Project**
  Groups are an easy way to share content and conversation, either privately or with the world. Create your new project group as a collaborative environment for sharing information and as a portal for your project team. [Get Started](#)

- **Upload Your Project Data**
  Become a contributor and share your work with your team and the community! Contributing content is easy. Our step-by-step forms will guide you through the process. [Get Started](#)

Featured Questions

- **How long does my NSF DMP need to be?**
  On average your NSF data management plan should be no more than two pages, but it will vary for each proposal. [Answered by Michael J. McLeman](#)

- **What do I do with IRB data?**
  Not all data gathered with your project needs to be made available. You should only published distilled, de-identified data on the hub. [Answered by Samuel P. Wilson](#)
Challenges & Opportunities

- Developing the Service
  - HUB technology
  - Policies
  - Providing training
  - Costs and Business Models
  - Assessment

- Developing Ongoing Roles & Relationships
  - Communication
  - Expansion to other units (IRB)
Panel Questions

- Which partnerships have been particularly important?
- Have your relationship-building efforts led to unexpected outcomes?
- What infrastructure/tools are needed for data curation at your institution?
- Does your institution have or is it developing data management policies (such as data retention)?

JAKE CARLSON – jrcarlso@purdue.edu

BARRIE HAYES - bhayes@email.unc.edu

SUSAN PARHAM- susan.parham@library.gatech.edu

BRIAN WESTRA - bwestra@uoregon.edu

IASSIST ANNUAL CONFERENCE
JUNE 02, 2011