DEFINING DATA SERVICES

A case study on process at a large research university library

Computers in Libraries 2012: Dealing with Data
Rebecca Reznik-Zellen
University of Massachusetts Amherst
Overview

- UMass Amherst Libraries and the Data Working Group
- Defining Data Services at UMass Amherst Libraries
  - Data support service levels
- UMass Amherst Libraries’ approach to integrating data services
Data Is Here

- 2003 Atkins Report "Revolutionizing Science and Engineering Through Cyberinfrastructure"
- 2009 NAS Report "Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age"
- Nature (08,09), Science (10), Economist (10), special issues
The Library’s Role


- “Roles for librarians in digital data curation will fall into one or more of three tiers:
  - National infrastructure;
  - Campus infrastructure;
  - Professional development and education.”

http://works.bepress.com/agold01/9/
University of Massachusetts Amherst

Stats
• +21,000 undergraduates
• +6,000 graduate students
• 1,174 full-time instructional faculty
• 60 doctoral and 78 master’s degree programs in 8 schools and colleges

FY 2011
• $143.5 million of sponsored research
• 1,094 awards from 1,233 proposals
• 624 federal awards (38% NSF)

http://www.flickr.com/photos/41700051@N06/5652224230/
UMass Amherst Libraries

<table>
<thead>
<tr>
<th>Humanities, Social Sciences, Management</th>
<th>Enrollment (primary majors and grad students)</th>
<th>Degrees Awarded (BA, MA, PhD)</th>
<th>Liaison Librarians</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10,965</td>
<td>3994</td>
<td>14</td>
</tr>
<tr>
<td>Science, Engineering, Health Sciences</td>
<td>11,272</td>
<td>2519</td>
<td>4</td>
</tr>
</tbody>
</table>

3/23/12  Computers in Libraries 2012: Dealing with Data

http://www.flickr.com/photos/41700051@N06/5652224230/
Determine if the University Libraries should accept broad responsibility for curating research data and, if so, how that should be done, what would be expected, and who would be involved (2010).
Data Working Group (DWG)

Self-education and environmental scanning

Open Access Week events and workshops for the campus

Interaction with campus entities and area institutions

Vision Statement

2010
DWG Vision

“DWG envisions the library as a full partner in the campus-wide research enterprise by offering, independently and in collaboration with other campus entities where appropriate, the full spectrum of data management services to our community and by building in-house expertise in data management.” (2011)

Process
1. Put national trends into local context
2. Identify (current and) desired services
3. Develop services that match capacity, mirror trends, and meet needs
UMass Amherst Context

Interviews and Focus Group
- 28 Faculty (7 pre- and 19 post-mandate)
- 17 Graduate Students
- 19 Disciplines
UMass Amherst Context

Faculty Questions

• Tell us about the data you’re creating/collecting. How is it generated? Are you building on existing or established data?
• How is the data stored and what file formats are you using?
• How large and what is it’s growth rate?
• How do you use the data? Do you expect to reuse or repurpose it in the future? Is the data used by others?
• What was your experience like in preparing a data management plan?

• What was the most difficult aspect of preparing a data management plan?
• Did you get any feedback on your plan?
• Is there any information that you would like to have had at your disposal when preparing the data management plan?
• What would you do differently in preparing your next data management plan?
• Are you aware of services for data management that are available through the Libraries?
• What services would you like to see the Libraries provide regarding data management?
UMass Amherst Context

Graduate Student Questions

• Tell us a little bit about your data management practices, particularly in terms of what you think works well for you or your project/department.
• Are there aspects of managing your research data that you’re struggling with?
• What is your data management role in your project or department? (Managing own data only, responsible for shared data, responsible for an advisor’s data…)
• What kind of training, documentation, or other support related to data management have you been given? (in the project, department, from advisor)
• [If it hasn’t come up yet: Is using/reusing data from other sources? How big a part of the research is this data? How did you find it?]
• [If there’s time: How do you feel about sharing your data publicly or on a case-by-case basis?]

• What resources or services have you found on campus to help you manage your data? How did you learn about these services?
• What services related to data management would you find useful? (If this doesn’t generate ideas, follow up on concerns that were stated and/or follow up with examples such as training, storage.)
## UMass Amherst Context

<table>
<thead>
<tr>
<th>Data Management Issues by Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data storage &amp; infrastructure</strong></td>
</tr>
<tr>
<td>• Size</td>
</tr>
<tr>
<td>• Backup/storage</td>
</tr>
<tr>
<td>• Computational power</td>
</tr>
<tr>
<td>• Preservation of proprietary programs/file formats</td>
</tr>
<tr>
<td>• Various infrastructure ownership scenarios</td>
</tr>
<tr>
<td><strong>Procedures &amp; training</strong></td>
</tr>
<tr>
<td>• Workflow routinization</td>
</tr>
<tr>
<td>• Knowledge transfer</td>
</tr>
<tr>
<td><strong>Documentation &amp; metadata</strong></td>
</tr>
<tr>
<td>• Project and discipline-specific practices</td>
</tr>
<tr>
<td>• Impact of external requirement on practices</td>
</tr>
<tr>
<td><strong>Data Reuse &amp; sharing</strong></td>
</tr>
<tr>
<td>• Versioning</td>
</tr>
<tr>
<td>• Making data public vs. making data useful</td>
</tr>
<tr>
<td>• Collaboration for publication vs. collaborative projects</td>
</tr>
<tr>
<td>• Occasional need to reproduce research</td>
</tr>
<tr>
<td><strong>IP &amp; data sensitivity</strong></td>
</tr>
<tr>
<td>• Privacy/IRB constraints</td>
</tr>
<tr>
<td>• USPTO rules for confidential exposure</td>
</tr>
<tr>
<td>• Emulation and post-publication sharing</td>
</tr>
</tbody>
</table>
UMass Amherst Context

Data management strategies vary widely up until the point of publication and do not typically support preservation or sharing.

Overarching Data Needs
  • Training in best practice for data management
  • Consultation services – particularly for data management plans
  • Infrastructure for data storage and backup
Identifying Services

Web Audit
• 18 institutions (10 peers, 8 models)
• 4 Categories

<table>
<thead>
<tr>
<th>Infrastructure</th>
<th>Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization</td>
<td>Marketing</td>
</tr>
</tbody>
</table>
Identifying Services

Generalized findings

- Service levels emerged based on degrees of librarian involvement with researchers and their data.
  - Education: Providing information resources to faculty
  - Consultation: Interacting one-on-one with faculty
  - Infrastructure: Taking stewardship of faculty research data

- Level of service is commensurate with inferred degrees of organizational and infrastructure support.
- Service levels add granularity to Gold’s campus-based data curation tier.

Education

Libraries educate their communities about data management

- Low-investment strategy with some opportunities for engagement with campus community.
- Education is a traditional role for libraries; data management expands the repertoire by targeting researchers at an earlier stage of the research process.

Media
- Web pages
- LibGuides
- Tutorials
- Workshops

Content
- Contextual information
- Mandate information and resources
- Practical, “how-to” information
- Metadata and standards
- Disciplinary repositories
- On-campus services
Education

University of Connecticut

Data Management Plans for Grant Funded Projects (NSF, NIH)

Writing a Data Management Plan
- Sample Data Management Plans
- Template for Data Management Plans
- Further information
- Data Management Workshop
- Contact Us

Since 1996, all federal granting agencies have required that funded research data be made available to the public. NSF Office of Science requirements for the creation, sharing, and storage of data.

The National Institutes of Health have required data sharing since 2003. Their Data Sharing Policy page provides the policy.

Starting in January, 2011, the National Science Foundation will require the submission of a two page data management plan. A comprehensive approach to data policy. Guidance from the NSF is general, leaving the specifics to be determined by the "community of practice". The NSF has updated their Grant.gov Application Guide to include information about data management plans.

Some subject areas, such as engineering or mathematical and physical sciences, may have specific guidelines addressing unique factors of data or division website for additional guidelines. Existing disciplinary requirements are posted at the Data Sharing and Interoperability section of the NSF.gov website.

NSF also provides an FAQ on data sharing which will answer some questions about the requirements and process.

Funding for Data Sharing

The NSF guidelines state that:

- NSF recognizes that it takes time and money to prepare data for sharing. Thus, applicants can request funds for data sharing and data management in the initial design of the study. A more realistic and economical way to establish deliberate data management practices.

- The NSF Social Sciences Directorate says funding: Any costs should be explained in the Budget Justification pages.

Managing Your Data

Digital data is growing at an exponential rate: from the digital family photos on your home computer to the terabytes of data generated by researchers in the various disciplines across the university. How do we as individuals and scholars in the digital research environment keep up with our growing data management needs?

The University Libraries are here to assist you with research data management issues through best practices, training, and awareness of data preservation issues. This site examines the research data life-cycle and offers tools and solutions for creation, storage, analysis, dissemination, and preservation of your data.

Creating a data management plan:

The Libraries can help you create an data management plan. We are interested in working with individuals to consult on the best ways to share, disseminate, and make accessible their research data. Here are some next steps you can take toward creating your plan:

- Take one of our data management workshops or watch our online tutorial on best practices.
- Learn about the various funding agency requirements and recommendations (i.e. NSF Data Sharing Policy).
- View a list of subject-specific data repositories to determine the best place to share your data.
- Get access to research computing resources on campus and include them in your data management plan.
- Consult with a librarian by contacting us with your questions.

Why is data management important?
Education

Frameworks for a Data Management Curriculum (2012):

- A robust program for graduate student data management education

http://library.umassmed.edu/imls_grant
Consultation

Libraries consult with faculty and researchers on a variety of issues relevant to the management of research data.

• Mid-range investment strategy requiring commitment of librarians and library administration.
• Expands battery of library services in a relevant manner.

Activities
• Review data management plans
• Identify data repositories and assist with deposition of material
• Consult on metadata standards, intellectual property, data sharing
• Referrals for storage options
Consultation

University of Nebraska-Lincoln

Workshops & Consultations

Help With Your Data Management Plan
The Libraries partners with Information Services and the Office of Research and Economic Development to offer support to UNL researchers who need help creating a data management plan.

This is what we can help you with:
- Conduct a 1.5 hour workshop for your department, laboratory, graduate students, or project team,
- Conduct 10-60 minute workshops for your department, laboratory, students, or project team,
- Review data management plans prior to submission,

Workshops may cover a combination of the following Data Management topics:
- Overview of what you need in order to effectively manage data,
- File formats and handling best practices,
- Where to get help for data storage, metadata, and archiving.

Schedule Workshops and Consultations

To schedule workshops:
- Contact us two weeks prior to the date you would like the workshop held.

To review your data management plan:
- Contact us two weeks prior to your submission deadline.
- Provide us with a copy of your narrative or proposal.

Contact Us

Please contact us at data-management@unl.edu to ask questions or to schedule a workshop or consultation.

University of Wisconsin-Madison

Research Data Services

Our Services
All our services are free to UW-Madison faculty, researchers, staff, and graduate students.

Data management plan help
- We can help draft a plan to meet requirements from NSF and other funders.
- We can also review your plan and suggest improvements.

Consultations
- Data workflow and process improvement in your department, research unit, or laboratory.
- File format and metadata standards that fit your research and your community.
- Digital preservation and archival concepts, to help you avoid losing your work.
- Advice on data sharing and reuse rights, to maximize your influence and credit.
- Database design advice and data modeling suggestions to get the most from your data.

Training and education
- We will train your trainers in data-management best practices.
- We also train you and your lab, customizing our approach to what you want to accomplish.
- We come to research-methods courses to train the next generation of researchers.
- We bring our expertise to your symposium, brown-bag, or meeting.

Referrals
- Storage and backup solutions, on campus and off.
- Data-security experts, particularly in the Office of Campus Information Security.

See something you need? Fill out the form below and we’ll be in touch.
Infrastructure

Libraries provide infrastructure for data management and data curation to their campus communities.

- High-investment, long-term strategy that requires the support of not only the Library but other campus entities.
- Opportunity to expand the library’s role on campus, but requires extensive scoping.

Examples
- Data staging platforms
- Institutional data repositories

Activities
- Relationship building
- Developing expertise
Infrastructure

RUcore - Rutgers University Community Repository

RUresearch Data Portal

The RUresearch Data Portal provides a place for Rutgers researchers to share their research data with the global scholarly community. RUresearch leverages all the capabilities of RUcore and adds additional tools and services specific to research data. More...

Learn More

RUresearch Services
- Free assistance and advice to design your data strategy and prepare the data Management component of a grant proposal.
- A customized search and retrieval portal for your data.
- Ongoing management and support for your data.
- Is there a fee for placing my data in RUresearch?
- Guidance for the NSF Data Management Grant Requirement

What does RUresearch Provide
- Robust preservation and long term access
- Metadata customized to your project's needs or community practice (e.g., DDI, Darwin Core, etc.)
- Access control, including flexible embargo periods and statements of use
- All associated information, presented in Context, such as codebooks, lab notes, images, software, related publications, etc.
- Contact the RUresearch Data Team

Specific RUresearch collections
- Video Mosaic Collaborative
- Equine Science Center

Data Management and RUresearch Presentation

About the RUresearch Data Team

Rutgers
Education

- Advertise notices about funder mandates
- Develop online resources with boilerplate text, descriptions of metadata, information about controlled vocabularies, links to funding agency requirements, supply a DMP template
- Facilitate workshops and create tutorials
- Participate in hands-on data management training.

Consultation

- Implement/coordinate comprehensive consultation program to address data management plans as well as general data management practices
- Conduct discipline-specific consultations with the input of departmental liaisons
- Recommend and assist with data deposition into institutional, disciplinary, or third party repositories for faculty
- Serve as a formal liaison to the Office of Research, the Graduate School, and other institutional entities invested in the research enterprise

Infrastructure

- Establish or develop relationships with information technology office, office of research, and other entities invested in the research enterprise with the goal of developing institutionally-endorsed platforms and policies for data management on campus
- Mediate data deposition into institutional, disciplinary, or third party repositories for faculty
- Utilize existing or create new repository for research data
Defining Services

A Rubric for planning services

- Activity one
- Activity two
- Activity three
- ...

Tiers of Service

- Education
- Consultation
- Infrastructure

Library Functions

Resources

Time (years)
## Defining Services

### DWG Vision Statement Objectives

<table>
<thead>
<tr>
<th>Education</th>
<th>Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Improve existing educational resources (ie: research toolkits);</td>
<td>DWG + formalized permanent committee and/or dedicated staff time and professional development support</td>
</tr>
<tr>
<td>2. Create high-value workshops;</td>
<td>DWG + 1 FTE (librarian) and professional development support</td>
</tr>
<tr>
<td>3. Participate in hands-on data management training.</td>
<td>Cross-institutional cooperation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consultation</th>
<th>Time (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Establish relationships with OIT, the Office of Research, and other entities on campus invested in the research enterprise with the goal of developing institutionally-endorsed platforms and policies for data management on campus;</td>
<td>1 yr.</td>
</tr>
<tr>
<td>2. Mediate data deposition into institutional, disciplinary, or third party repositories for faculty;</td>
<td></td>
</tr>
<tr>
<td>3. Understand high capacity and/or cloud storage, networking, and security issues and recommend cost efficient solutions to faculty researchers.</td>
<td>2 yrs.</td>
</tr>
<tr>
<td>4. Develop in-house expertise in data management;</td>
<td></td>
</tr>
<tr>
<td>5. Coordinate a comprehensive consultation program, extending beyond data management plans to include all the services that would enable best practices for data management on campus.</td>
<td>2 yrs.</td>
</tr>
<tr>
<td>6. Coordinate an ongoing training program on data management for the campus community;</td>
<td>6 mos.</td>
</tr>
<tr>
<td>7. Serve as a formal liaison to the Office of Research, the Graduate School, and other institutional entities invested in the research enterprise;</td>
<td></td>
</tr>
<tr>
<td>8. Recommend and assist with data deposition into institutional, disciplinary, or third party repositories for faculty;</td>
<td></td>
</tr>
<tr>
<td>9. Develop in-house expertise in data management;</td>
<td></td>
</tr>
<tr>
<td>10. Coordinate a comprehensive consultation program, extending beyond data management plans to include all the services that would enable best practices for data management on campus.</td>
<td></td>
</tr>
</tbody>
</table>

### Library Functions

- Coordinate an ongoing training program on data management for the campus community;
- Serve as a formal liaison to the Office of Research, the Graduate School, and other institutional entities invested in the research enterprise;
- Recommend and assist with data deposition into institutional, disciplinary, or third party repositories for faculty;
- Develop in-house expertise in data management;
- Coordinate a comprehensive consultation program, extending beyond data management plans to include all the services that would enable best practices for data management on campus.

- Defining Services
- DWG Vision Statement Objectives
- Library Functions
- Resources
Integrating Services

Digital Strategies Group oversees coordination of digital activities, establishes priorities and provides guidance for digital projects.
Integrating Services

Digital Strategies Group

Metadata

Digital Creation and Preservation

Data

Strategic Plan
Going Forward

Current Services
• Web pages
• DMP consultation
• Workshops
• IR for small, static, public data sets

Where We’re Headed
• Liaison education
• Subject-specific workshops for graduate students
• Mediated data deposition
• Consultation on standards and metadata
• Referrals for infrastructure
• Relationships building toward infrastructure
Thank You!

Questions?

Contact
• datamanagement@library.umass.edu

http://www.flickr.com/photos/41700051@N06/5652224230/