Managing Data Management: building an RDM service at MIT Libraries

Christine Malinowski & Phoebe Ayers

eScience Symposium | April 6, 2017
MIT by the numbers...

4,524 undergraduates

6,852 graduate students

~12,000 staff & faculty *

* half academic
MIT Libraries by the numbers...

~170 staff

5 libraries
+ Institute Archives & Special Collections

Research-level collection in most areas
Data Management Services
@ MIT Libraries

Program Head, DMS
Currently Interim
(100% time)

Research Data Librarian
Christine Malinowski
(75-100% time)

EE&CS Liaison Librarian
Phoebe Ayers
(30% time)

Biosciences Liaison Librarian
(30% time)

Digital Humanities Librarian
(10-20% time)

Repository and MIT Publications Manager
(10% time)

Application Developer / Analyst
Data Management Services @ MIT Libraries

- 2004: Data Services Reference Librarian began providing informal DM support
- 2008: Cross-library Data/GIS Group & RDM web guide created
- 2009: Held first DM workshop
- 2010: Formal service launched as Research Data Management Team
- 2012: Refined, rebranded as Research Data Services Working Group
- 2014: 1st full-time staff (Library Fellow for RDM)
- 2015: Rebranded as DMS program; 2nd full-time position (Program Head)
- 2016: Library Fellow made permanent staff position
Data Management Services
@ MIT Libraries

is NOT:

● GIS services
● Data reference or collection development support
● Data visualization support (*yet*)
● Embedded in grants/research projects
● Systematic review service
● IT or Archives
Data Management Services
@ MIT Libraries

PROVIDES:

- assistance with all questions about storing and organizing research data
- assistance with data management plans and DMPTool
We ...

- answer **direct questions** and **referrals**
- give **individual consults** by email or in person
- offer research **group consults** on file organization, etc.
- teach **classes**
Wednesday August 10, 8:37pm

To: data-management@mit.edu

“We are asked to provide a data management plan for our grant proposal that has been accepted for award negotiations by the DOE. We put together a draft according to the guidelines. It is to be submitted by Friday ...

It would be very helpful to have your feedback on the draft. Do you think the language looks good? Would you have any suggestions about how to improve it?”
**Tools for managing consults**

**Coordinating**
- Contact address that is a list
- Referrals from other service points & librarians
- Monthly email screening schedule

**Delivering**
- Consultation pair ups
- Review DMPs via DMPTool feedback mechanism

**Tracking / Follow up**
- Researcher contact log on wiki (what, who, duration, status...)
Case study

Data Management Plan
15-30 consults / yr
includes brief 1-2 hrs and multi-week/visit consults

~15 workshops / yr
to individual departments and whole campus
DMS Projects
behind the scenes...

- Reviewing and updating public website & knowledgebase
DMS Projects
behind the scenes...

- Reviewing and updating public website & knowledgebase
- Developing IR procedures for student-generated data
- Supporting LabArchives
- Interviewing faculty about their data needs
- Reviewing agency guidelines for DMPs & updating our documentation
- ...
Case study

DMS internal project: Knowledgebase

Search Knowledgebase

General/Getting Started
• Data Management Services procedures
• Outreach
• Questions to ask in a consultation
• Points to make in a consultation
• Consultation Worksheets
• Services Across the Data Lifecycle
• Knowledgebase maintenance
• RDM Professional Resources

File Organization
• File naming best practices
• Version Control and Collaboration
• Questions to ask in a consultation: file organization

Data Storage during the Active Phase of Research
• Campus data storage resources
• Backups
• File format recommendations
• Digitalization and format conversion
• Electronic Lab Notebooks (ELNs)

Metadata
• Metadata: What is it?
• Metadata: Creating it
• Science dataset read me example: txt version | PDF version
• Social science dataset read me example: txt version | PDF version

Data Security & Confidentiality
• Data protection and security
• Confidentiality of data

Intellectual Property
• Data licensing
• Data and copyright

Data Sharing and Long-term Storage
• Data repository resources
• Repository recommendations provided, use cases
• Questions to ask in a consultation: repository selection criteria
• Points to make in a consultation: long-term data storage
• Data citation and credit
• Data Preservation Resources
• Citing and preserving software/code

Referrals or Other Resources
• Referrals
• Putting data in DSpace
Case study

DMS internal project: Knowledgebase

Comparing and selecting among repositories

Repository comparison tools

Comparison Chart Template (Word document) - Mikkie is reviewing/updates doc (March/April 2017):
- Tool that can be filled in for relevant repositories and presented to a user to help them understand differences among repositories
- Lists criteria for evaluating different types of data repositories
- Includes baseline information about DSpace, Dataverse, and Dryad

DCC How-to Guide: Where to keep research data: checklist for evaluating data repositories
- Describes different types of repositories, their pros/cons, and special cases such as confidential data
- Focuses on 5 key questions:
  1. Is a reputable repository available?
  2. Will the data you want to deposit?
  3. Will it be safe in legal terms?
  4. Will the repository sustain the data value?
  5. Will it support analysis and track data usage?
- Outlines 3 tiers of service capability levels for each criterion
- Includes bibliography of key good practice recommendations

rediata.org: Registry of Research Data Repositories
- Can search or browse for repositories by criteria such as subject, content type, or country
- Features of repositories are described in a standardized way

Options to present to the user

At minimum, consider presenting the following options to the user (guided by what you already know about their needs):

1. DSpace@MIT:
   - Policies: About DSpace@MIT, Policies, scroll down to Guidelines for Research Datasets
   - Make sure to note for user also the Format Support policy on About DSpace@MIT, Policies, which indicates what level of support/preservation is provided depending upon the file format: bottom-line, if they are planning on depositing files in the "known" (i.e., proprietary) format, encourage them to also make a copy in a "supported" format as this will receive more full treatment long-term.
   - To begin the deposit process: see putting data in DSpace

2. Harvard Dataverse Network, a service of the Harvard-MIT Data Center
   - Policies: Harvard Dataverse Network Guidelines
   - To begin the deposit process: go to the Harvard Dataverse Network > create an account > create a Dataverse > then begin to add data
   - Documentation: User guide covers: account creation/management, Dataverse management, dataset and file management, and use features

3. Other (suggest open access deposits, institutional repositories, and large research data organizations)
**Knowledgebase Content Tracking**

Added by Christine N Malinowski, last edited by Christine N Malinowski on Mar 16, 2017 12:27 (view change) show comment

**Related Pages**
- Data Management Consulting Knowledgebase
- Knowledgebase maintenance

**Content Review Tracking Table**

Instructions for Content Creators: Add pages to the table as they are created.

Instructions for Content Reviewers: Review the content on the pages for which you are assigned. Mark the review status as:
- In progress when the review is underway
- Needs input when the review requires additional input from the group (provide notes on what the content in question is in review notes column)
- Complete when all needed revisions have been completed
- Complete, edited when the review is complete and content edits have been made

**2017 Review**

- Pages assigned to individual reviewers (Mikki and Ece will self assign individual pages)
- Reviewers check pages, following instructions for reviewers above; for minor changes (e.g., broken links), reviewers can go ahead and make changes
- Review team discusses review notes, determines updates (or issues to take to larger DMS team)
- Residual edits are made to pages

<table>
<thead>
<tr>
<th>General / Getting Started</th>
<th>2017 Assigned Reviewer</th>
<th>2017 Review Notes</th>
<th>2017 Review Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMS procedures</td>
<td>ooa</td>
<td>minor edit: added A. Chassanoff name to CLIR post-doc wish list</td>
<td>complete, edited</td>
</tr>
<tr>
<td>Outreach</td>
<td>ooa</td>
<td></td>
<td>complete</td>
</tr>
<tr>
<td>Questions to ask in a consultation</td>
<td>ooa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Points to make in a consultation</td>
<td>ooa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledgebase maintenance</td>
<td>ooa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Services Across the Data Lifecycle</td>
<td>ooa</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>covered by other reviewers below</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>covered by other reviewers below</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>complete</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>complete</td>
<td></td>
</tr>
</tbody>
</table>

**External Requirements**

<table>
<thead>
<tr>
<th>Funder &amp; Publisher Requirements</th>
<th>Assigned Reviewer</th>
<th>2017 Review Notes</th>
<th>Review Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMPs: What are they?</td>
<td>ooa</td>
<td>minor update to U Vic site link</td>
<td>complete, edited</td>
</tr>
<tr>
<td>DMPs: Answering DMP review queries</td>
<td>ooa</td>
<td>minor edit to WHouse 2013 mamo link</td>
<td>complete, edited</td>
</tr>
<tr>
<td>DMPs: DMPTool &amp; other drafting resources</td>
<td>ooa</td>
<td>question: have we not had a DMP review request since 10/2016?</td>
<td>complete</td>
</tr>
<tr>
<td>File Organization</td>
<td>Assigned Reviewer</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DMS internal project:**

Knowledgebase
Tools for project management

Individual project management

- Wiki pages for projects
- Small teams for each project assignments, including project lead, based on member capacity

Project portfolio management

- Monthly, in-person team meetings
- Tableau visualization
# DMS Project & Activity Plan, FY17

**Updated: 03.07.2017**

### Project Category 1: Screening
- Instruction
- Blog Writing (Group sessions)
- Project Planning, FY18
- RDM Biweekly updates
- Web, DMPTool, & KB Review
- What's up with DMS brown bag

### Project Category 2: DMS Development
- New Program Head
- Create a base calendar of DMS services and activities

### Project Category 3: Work with MIT: Services
- DMP followup
- Develop File Organization
- Consultation Materials

### Project Category 4: Work with Others: Outside Libraries
- Campus RDM Network: Pre-work
- Campus RDM Network: Phase I

### Project Category 5: Tools/Technology
- Dispace Student Research Data
- Data Management Infrastructure: Discussion
- Data Management Infrastructure: Project Prioritization & Submission

### Project Category 6: Other
- Consider DMPTool Article
- Write/Submit DMPTool article

### Projects to Potentially Add to FY17
- LabArchives article
- Other
- Map project open data metadata to metadata standards

### Projects by Commitment Level
- Data Management Infrastructure: Discussion
- Dispace Student Research Data
- Campus RDM Network: Pre-work
- Develop File Organization
- Consultation Materials
- New Program Head
- Create a base calendar of DMS services and activities
Tools for keeping up with RDM

- professional development opportunities
- internal documentation (via knowledgebase)
- *RDM Updates* biweekly digest of external RDM information

DMS is hard; there’s never just one “right answer”
RDM Updates

Your biweekly RDM Update is here! The next update will be distributed on March 17th. The content within this email will also be available via this wiki archive page.

CALL FOR PARTICIPATION

New DLF Interest Group: Government Records Transparency and Accountability
...With this blog post we officially announce the formation of the DLF Interest Group on Government Records Transparency and Accountability. A few emails have been exchanged, and an initial conference call has been convened...

RDA Data Versioning Interest Group: Call for policy examples
The demand for reproducibility of research results is growing, therefore it will become increasingly important for a researcher to be able to cite the exact extract of the dataset that was used to underpin their research publication. It is widely acknowledged that published data need to be versioned but no guidelines exist to date.

RESOURCES, TOOLS, ETC.

RISE, a self-start tool for research data management service review
DCC releases RISE, a new framework aiming to help review RDM services.

New rdatoolkit.org website
We are excited to announce that we will roll out a new rdatoolkit.org site next week! Just to clarify the rdatoolkit.org is the information and communication site that serves as a kind of entrance way to the RDA Toolkit proper. Our goal is to make the site easier to navigate and more informative. The updated site will include a more dynamic home page with images and headlines that will make valuable news more conspicuous and accessible. Learn more about the new rdatoolkit.org website.

Full Text Book: “Curating Research Data” (Open Access Version Available)
An open access edition of Curating Research Data (two volumes) published by ACRL (Association of College and Research Libraries) are linked below. Bound volumes are also available for sale.

ARTICLES

Nature News: Peer-review activists push psychology journals towards open data
Editor asked to resign from journal for saying he'll review only papers whose data he can see.

Nature, Editorial: Don't let useful data go to waste
Researchers must seek out others' deposited biological sequences in community databases, urges Franziska Denk.

Library Workshops in Support of Data-Driven Research in Top NIH- and NSF-Funded Universities
In response to an urgent need for computationally-savvy researchers, leading university libraries are developing and offering participatory workshops to introduce their patrons to data-driven research methods and resources. In doing so, libraries help to improve the competitive advantage of their universities relative to other schools that compete for federal funds. We conducted a qualitative study to explore the nature of library workshops offered in top NIH- and NSF-funded universities. To guide strategic planning, we present a catalog of the workshops offered in fall 2015, two indexes for each workshop's context, and a data-driven case study of the overall significance of the workshops to the library's mission.
Case study

Research group: Environmental data

Scenario
- Decade-long, multi-site, multi-million dollar research project
- Multi-terabyte data and software with several versions used worldwide
- Hosted several places, including sites maintained by now-departed people

Request
- Wanted secure place to store latest versions & documentation
Tools for promoting our services

**Internal**
- Library staff *brown bags*
- **Regular interactions** with IT, Office of Research, OGC, etc.

**External**
- Public *website* & regular *blog posts*
- **Open classes** offered as part of library instruction
- DMS *promoted by liaisons* in orientations & to new faculty
DMS
looking ahead

What do we want [to do]:

○ more community outreach and promotion
○ more research into & public documentation on tricky topics (file organization, software, etc.)
○ better local data storage support
○ clearer relationships with liaisons, archives, IT ...
DMS
looking ahead

Future projects:

○ engage with campus partners to develop RDM network
○ develop a research data repository
Discussion & Questions

● What aspects of managing RDM have proven difficult?
● Do you have tips to share?
● What kinds of questions do you get most often?
● Questions for us?
Thanks!

Any questions?

http://libraries.mit.edu/data-management | data-management@mit.edu

cmalin@mit.edu & psayers@mit.edu
Credits

Special thanks to SlidesCarnival who made and released the presentation template used (‘Fidele’ with modifications) in this slide deck.