Establishing Services & Resources to Support Faculty Writing Data Management Plans: Lessons Learned from an Engineering Pilot

Jen Green & Natsuko Nicholls

Practice Paper Co-authors
Leena Lalwani, Paul Grochowski & Sara Samuel

IDCC14, San Francisco, CA
24-27 Feb 2014
CONTEXT
Data@UM

- Strongly decentralized campus
- Grassroots works: deploying multiple data initiatives & projects
- Capitalizing on liaison relationships
- Defining new roles for librarians
College of Engineering

- ~16 departments, ~500 faculty, ~10,000 students
- Strong liaison relationship
- New Responsible Conduct of Research program
- Responsive leadership
PROGRESS

Engineering Data Management Support Pilot
NSF (Total) Funding Rate, Annual Trend, FY 2004-2013
U of Michigan vs. All Institutions

Mean value: UM=29%; All institutions=25%
NSF (ENG) Funding Rate, Annual Trend, FY 2004-2013
U of Michigan vs. All Institutions

U of Michigan vs. Average of all institutions

Mean value: UM=23%; All institutions=19%
FACULTY SURVEY
Familiarity with DMP Requirements

- Not at all familiar: 12%
- Slightly familiar: 20%
- Moderately familiar: 40%
- Very familiar: 20%
- N/A: 8%
Have you written a DMP?

- Yes: 72%
- No: 28%
DMP ANALYSIS
DMP Analysis: Criteria

- Roles and responsibilities for data management
- Types of data produced
- Data and metadata standards
- Data storage options
- Plans for data sharing and redistribution
- Plans for archiving and preservation
Level of Detail in DMPs

- **High**: 38%
- **Med**: 32%
- **Low**: 21%
- **N/A**: 9%

Quality of DMPs

- **Fair**: 65%
- **Poor**: 7%
- **Good**: 19%
- **N/A**: 9%
Roles & Responsibilities for Data Management

- Individuals (PIs, Co-PIs, students): 41%
- Project Team: 8%
- Institutions: 8%
- N/A: 9%
- No info: 34%
Volume of Data Generated

- No info: 74%
- Detailed info: 8%
- Some info: 9%
- N/A: 9%
Period of Data Retention

- No info: 30%
- 1-3 years: 25%
- 4-5 years: 15%
- 6-10 years: 6%
- > 10 years: 15%
- N/A: 9%
Means of data dissemination

- Publication: 42%
- Faculty/project website: 36%
- Conference presentation: 11%
- "Upon request": 11%

Total: 100%
Preservation

- ~20% indicated Deep Blue (UM’s IR) as a location for preservation of access to data.
- DR was not mentioned.
RESOURCES & SUPPORT

Research Guides & Faculty Workshops
Web-based NSF DMP Guide and Resources

http://guides.lib.umich.edu/engin-dmp
Please note that these DMP excerpts are copyrighted by their respective authors.

**Preferred:**

“These data will provide a detailed experimental look at the mechanical regulation of mesenchymal stem cell osteogenesis. The data will further delineate the functional role of the cytoskeleton-focal adhesion-extracellular matrix signaling axis in the mechanoresponsive mesenchymal stem cell osteogenesis, as described in the main body of the proposal. As such, they will be of interest to the tissue engineering and regenerative medicine communities.”

This is a thorough description of who would be interested in the data.

**Less Developed:**

“No, or very little, derivative use of the data is expected since the research is aimed at developing machining process models and the data to be gathered is directed toward this end.”

A belief that others won’t be interested in your data is an insufficient reason for declining to share it.

---

Please note that these DMP excerpts are copyrighted by their respective authors.

**Preferred:**

“As previously mentioned, data and codes developed during this research will be stored on both the PIs’ groups' local servers and on the campus servers. The campus servers are automatically backed up every day. Data generated at the University of Michigan will be stored in a repository called Deep Blue, while data generated at the University ______ will be stored in a repository called ______ providing 100Gb (expandable).”

Mentioning both short-term and long-term storage is good practice, and these are reliable storage options with automatic back-ups.

**Less Developed:**

“All data will be available at request immediately after the scientific results are published and will be stored at least another five years on magnetic and optical storage devices (hard disks, CDs, DVDs). Optical storage devices will serve the purpose of the Disaster Recovery Plan.”

Optical storage media are not sufficiently reliable for long-term archival storage.
Discussing a Data Management Plan Example

using National Science Foundation Engineering Directorate Guidance

Data Management Plan

1. Responsibility for Data Management
   - The PI will be responsible for enforcing the data management plan. He will carry out a monthly check to ensure that the plan is being followed by all concerned.

2. Types of data to be produced in the course of the project
   - Matlab (or equivalent software) codes, figures and simulations
   - Frequency response measurements (frequency data, time domain data and coherence)
   - Modal analysis files
   - Time domain vibration data
   - System parameters, measurement conditions, set up photos and videos
   - The estimated size of all the data is 100 GB per year

All the data will be collected in the PI's lab or at the facilities of the industrial collaborators. All data will be saved on password-protected computers at the PI's lab. A backup of the data will be stored on Cloud storage providers by U-M College of Engineering. A backup will be...
OTHER RESEARCH

Graduate Student Interviews about Data
Q. How much data did you generate in this research project?

Note: N=109; Three students answered “Don’t know.”
Q. “What percentage of your data would it take to replicate this research?”

![Bar chart showing student counts across different percentage ranges.]

<table>
<thead>
<tr>
<th></th>
<th>0%</th>
<th>1 to 10%</th>
<th>11 to 20%</th>
<th>21 to 30%</th>
<th>31 to 40%</th>
<th>41 to 50%</th>
<th>51 to 60%</th>
<th>61 to 70%</th>
<th>71 to 80%</th>
<th>81 to 90%</th>
<th>91 to 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Series1</td>
<td>20</td>
<td>30</td>
<td>4</td>
<td>7</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>27</td>
</tr>
</tbody>
</table>

Note: N=103
NEXT STEPS
Lessons Learned & Next Steps

Lessons:
- Engage stakeholders
- Time

Next steps:
- Integration into key required components
- Increased graduate student offerings
- DM Review Service development
- Iterated pilot with LS&A
Questions?

• Jen Green: Director of Research Data Services  
greenjen@umich.edu

• Natsuko Nicholls: CLIR/DLF Data Curation Fellow  
hayashin@umich.edu