The library’s role in promoting the sharing of scientific research data

Katherine Akers
Biomedical Research/Research Data Specialist
Shiffman Medical Library
Wayne State University
Funding agency requirements

Starting in 2003, applications requesting $500k/year in direct costs are required to include a data sharing plan (1-2 paragraphs) or state why data sharing is not possible.

Starting in 2011, all applications are required to include data management plan 1-2 pages) describing how data will be disseminated and shared.
NSF Data Management Plan (DMP): Biological Sciences Directorate

- Describe data that will be collected and metadata standards that will be used.
- Describe the resources/facilities that will be used to store and preserve the data after the grant ends.
- Describe dissemination methods that will be used to make data available to others after the grant ends.
- Describe policies for data sharing and public access (e.g., protection of privacy, intellectual property rights)
- Describe roles and responsibilities of all parties with respect to data management after the grant ends.
Library research data services:
Born out of the NSF DMP requirement

ARL SPEC Kit 334: Research Data Management Services, 2013
DMP-related services

• Providing information about DMPs (libguides, etc.)
• Holding workshops on DMPs for faculty members
• Assisting with the formulation and/or writing of DMPs
U-M faculty with grants requiring DMPs

- NSF, 9%
- Large NIH, 6%
- NASA, 0.6%
- NOAA, 0.1%
- More than one funding agency, 1%
- None, 83%

“... a thin thread on which to hang an entire service.”
–Dorothea Salo
Consider the entire research data lifecycle

1. **Plan**
   - DMP creation
2. **Collect**
   - Locate existing datasets
3. **Process**
   - Storage and back-up
   - File organization
4. **Analyze**
   - Statistical analysis
   - Data visualization
5. **Preserve**
   - Preservation file formats
6. **Provide access**
   - Data repositories
   - Data journals
7. **Re-use**
   - Data documentation

**Appraise and select**
Data sharing and management snafu in 3 short acts

http://youtu.be/N2zK3sAtr-4
Data repositories

Institutional repositories
Data repositories

Institutional repositories

Disciplinary repositories

- GenBank
- SRA
- dbGaP
- UniProt
- RCSB PDB
- FlyBase
- DRYAD
- THE CANCER IMAGING ARCHIVE
- OpenfMRI
- NDAR
- MG1
- Influenza Research Database
Dryad: scientific data repository

1. Deposit Data
2. Get Permanent Identifier
3. Watch Your Citations Grow!
4. Relax, Your Data are Discoverable and Secure

Submit data now
How and why?

Search for data

Browse for data

Recently Published Data

Biro PA, Adriaenssens B (2013) Data from: Predictability as a personality trait: consistent differences in intraindividual behavioral variation. The American Naturalist
http://dx.doi.org/10.5061/dryad.h8c10

Sangster G (2014) Data from: Declining rates of species described per taxonomist: slowdown of progress or a side-effect of improved quality in taxonomy? Systematic Biology
http://dx.doi.org/10.5061/dryad.m4g2n

Partnerships between data repositories and academic libraries
Partnerships between data repositories and academic libraries

1. Membership

• Pay annual membership fee
• Attend annual Dryad membership meeting
• Vote on Dryad Board of Director members, bylaws, and budget
• Potential to influence the development of Dryad
Partnerships between data repositories and academic libraries

2. Provide financial assistance for data deposit

<table>
<thead>
<tr>
<th>Payment Plan</th>
<th>Member</th>
<th>Non-member</th>
<th>Minimum purchase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Voucher Plan</td>
<td>$65 per data package</td>
<td>$70 per data package</td>
<td>25 vouchers</td>
</tr>
<tr>
<td>2. Deferred Payment Plan</td>
<td>$70 per data package</td>
<td>$75 per data package</td>
<td>1 yr contract</td>
</tr>
<tr>
<td>3. Subscription Plan</td>
<td>annual fee based on $25 per published research article</td>
<td>annual fee based on $30 per published research article</td>
<td>2 yr contract</td>
</tr>
</tbody>
</table>

For individuals: Pay on submission

Excellent! 1 data package
Partnerships between data repositories and academic libraries

3. Local data curator
Barriers to research data dissemination

- Data sharing takes time
- Lack of incentive
Journals that publish “data papers”
Featured Data Descriptor

A Southern Indian Ocean database of hydrographic profiles obtained with instrumented elephant seals
Roquet et al.
Data Descriptor | 2nd September 2014

Our understanding of Southern Ocean currents is limited by a lack of data, particularly in seasons and locations that are hard for ships to access. By affixing sensors to wild seals, these researchers have built a large database of temperature and salinity profiles, extending our knowledge of this key component of our planet's oceans.

Also this week: Our first Data Descriptor Collection, from the National Evolutionary Synthesis Center (NESCen)

Latest content

Data Descriptor | 62 September 2014
Artificial selection on anther exsertion in wild radish, *Raphanus raphanistrum*
Jeffrey K. Conner, Cynthia J. Mills  […]  Keith Keroly

Data Descriptor | 62 September 2014
A Southern Indian Ocean database of hydrographic profiles obtained with instrumented elephant seals
Fabien Roquet, Guy Williams  […]  Mike Fedak

About Scientific Data

Scientific Data is an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the Data Descriptor, is designed to make your data more discoverable, interpretable and reusable.
Nature Publishing Group’s Scientific Data

http://youtu.be/hrHM3bUym3g
Example “Data Descriptor”

- **Background**: motivation for data collection
- **Methods**: data collection and processing procedures
- **Data Records**: data location, description of data file contents
- **Technical Validation**: quality-control measures
- **Usage Notes**: suggestions for data use
Data journals complement use of data repositories

Traditional journal articles

Data papers

Data repositories
Data training for librarians
Data training for librarians

RDM Rose

MANTRA
Research Data Management Training

UNC
School of Information and Library Science

Graduate School of Library and Information Science
The iSchool at Illinois
Tiered approach to data training for librarians

1. Research Data Concepts (basic training)

Working with data
- File formats, naming, versioning
- Database design
- Metadata
- File organization
- Data storage
- Data security

Sharing and preserving data
- Social barriers to data sharing
- Data repositories
- Data journals
- Data citation
- Data management planning
Tiered approach to data training for librarians

1. Research Data Concepts (basic training)

2. Deep Dives into Data (advanced training)
   - Ecology (example science discipline)
   - Psychology (example social science discipline)
   - Health Sciences
   - Arts & Humanities
Deep Dives into Data (for specific subject areas)

What are the data sharing policies of the subject area’s major funding agencies, academic societies, and journals?

What subject-specific data repositories are available?

What is the culture of data sharing in the subject area?
Tiered approach to data training for librarians

1. Research Data Concepts (basic training)

2. Deep Dives into Data (advanced training)

3. Data-related seminars (continuing education)
   - Evaluating DMPs
   - Data storage options on campus
   - Text mining
   - Data interviews with faculty members
At Wayne State University . . .

• Connect and collaborate with other research units across campus

• Develop workshops on research data management, scientific publishing, and research impact.

• Build research-oriented libguides for specific populations of campus researchers.

• Plan “Science Boot Camp” for librarians in the Midwest

• Establish formal Research Data Services program within the library system