Research Data Services: New Roles for Academic Libraries?

Carol Tenopir
University of Tennessee
cntenopir@utk.edu
Visiting Research Professor
University of New South Wales
(through April)

Science is more collaborative, interdisciplinary, computational, and data-intensive
Open Science Movement…

...governments and funding agencies are requiring data management

U.S. Office of the the President

To that end, I have issued a memorandum today (pdf) to Federal agencies that directs those with more than $100 million in research and development expenditures to develop plans to make the results of federally-funded research publicly available free of charge within 12 months after original publication.

...the memorandum requires that agencies start to address the need to improve upon the management and sharing of scientific data produced with Federal funding.


Data loss

- Natural disaster
- Facilities infrastructure failure
- Storage failure
- Server hardware/software failure
- Application software failure
- External dependencies
- Format obsolescence
- Legal encumbrance
- Human error
- Malicious attack by human or automated agents
- Loss of staffing competencies
- Loss of institutional commitment
- Loss of financial stability
- Changes in user expectations and requirements
NSF Sustainable Digital Data Preservation and Access Network Partners (DataNet)

Will create exemplar partners to address “...one of the major challenges of this scientific generation: how to develop the new methods, management structures and technologies to manage the diversity, size, and complexity of current and future data sets and data streams”.

Need new types of organizations that will...

- integrate library and archival sciences, cyberinfrastructure, computer & information sciences, and domain science expertise to:
  - provide reliable digital preservation, access, integration, and analysis capabilities for science and/or engineering data over a decades-long timeline
The DataONE Vision and Approach:

Providing universal access to data about life on earth and the environment that sustains it, as well as the tools needed by researchers.

1. Building community
2. Developing sustainable data discovery and interoperability solutions
3. Enabling science through tools and services
DataONE is Cyberinfrastructure

Three major components for a flexible, scalable, sustainable network

**Coordinating Nodes**
- retain complete metadata catalog
- indexing for search
- network-wide services
- ensure content availability (preservation)
- replication services
DataONE is Cyberinfrastructure

Three major components for a flexible, scalable, sustainable network

<table>
<thead>
<tr>
<th>Coordinating Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Member Nodes</td>
</tr>
<tr>
<td>- diverse institutions</td>
</tr>
<tr>
<td>- serve local community</td>
</tr>
<tr>
<td>- provide resources for managing their data</td>
</tr>
<tr>
<td>- retain copies of data</td>
</tr>
</tbody>
</table>

Member Nodes
- TheCornellLab
- UC3Merritt
- PISCO
- USGS
- ESA
- ONEShare

Investigator Toolkit
- command line interface
- python
- Java
- ONEDrive
- DMP Tool
- Mercury
- DATA
- Kepler
- ONE

DataONE is Cyberinfrastructure

Three major components for a flexible, scalable, sustainable network

<table>
<thead>
<tr>
<th>Coordinating Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Member Nodes</td>
</tr>
<tr>
<td>- diverse institutions</td>
</tr>
<tr>
<td>- serve local community</td>
</tr>
<tr>
<td>- provide resources for managing their data</td>
</tr>
<tr>
<td>- retain copies of data</td>
</tr>
</tbody>
</table>

Member Nodes
- TheCornellLab
- UC3Merritt
- PISCO
- USGS
- ESA
- ONEShare

Investigator Toolkit
- command line interface
- python
- Java
- ONEDrive
- DMP Tool
- Mercury
- DATA
- Kepler
- ONE
DataONE Principles

1. Data should be part of the permanent scholarly record and requires long-term stewardship.

2. Sharing and reuse maximize the value of data to environmental science.

3. Science is best served by an open and inclusive global community.

4. The data environment is dynamic and requires evidence-based decision-making about practice and governance.
Alison Specht, ACEAS Program Manager, Associate Professor, University of Queensland

Australian Centre for Ecological Analysis and Synthesis

Libraries…

• facilitate interdisciplinary work, data access, and data knowledge through collections and services
• Can take a leadership role in a variety of research data services
Research Data Services are…

• …services that address the full data lifecycle, including the data management plan, digital curation (selection, preservation, maintenance, and archiving), and metadata creation and conversion.

• Services can be hands-on or consultative.

Technical/Hands-On RDS

• Providing tech support for RDS systems/data repositories
• Deselecting datasets for removal from a repository
• Preparing datasets for deposit
• Creating or transforming metadata
• Identifying datasets for deposit in repositories
Informational/Consulting RDS

- Consulting on Data Management Plans
- Consulting on data and metadata standards
- Outreach/collaboration with other RDS providers on campus
- Reference for finding and citing data
- Creating web guides and finding aids for data

More Informational/Consulting RDS

- Directly participating with researchers on a project
- Discussing RDS with others on campus on a (semi-) regular basis
- Training co-workers on data issues
Researchers might…
Researcher Challenges

Organizational Support & Services, Training?

Plan

Collect

Assure

Describe

Preserve

Integrate

Discover

Collect

Where?

Data Sharing?

Plan

Analyze

Integrate

Discover

Preserve

Librarian & Libraries

Level of my knowledge & skills?

Plan

Level of participation with data?

Are RDS priority?

Role in partnering with researcher?

Collect

Level of involvement with metadata?

Integrate

Are there an agency repository that accepts data?

Describe

Stewardship role (select & deselect)?

Preserve

Role of the librarian to help preservation?

Discover

Role of librarian discovering data?
Assessment of Stakeholders

Data Managers  Public Officials  Publishers

Scientists  Students & Teachers  Citizen-scientists  Libraries & Librarians

Results of *scientists* survey shows opportunities for libraries…
What metadata do you currently use to describe your data?

- DIF: 12
- DwC: 21
- DC: 26
- EML: 95
- FGDC: 95
- Open GIS: 96
- ISO: 97
- My Lab: 266
- NONE: 676

Gap Between Willingness to Share and Accessibility

- Place at least some of my data into a central data repository: 78%
- I share my data: 75%
- Place all of my data into a central data repository: 41%
- Others can access my data easily: 36%
Academic Librarians

- Unit of analysis = individual librarian
- 302 librarians at 111 ARL libraries who were most likely to be involved in RDS
- Opinion about personal preparation and opportunities for continuing education
Academic Libraries

- Unit of analysis = library
- 223 ACRL libraries
- From stratified panel of directors from all types of academic libraries
- Current RDS and educational offerings
- Plans for next 24 months

*librarians* (in academic research libraries)…
ARL librarians strongly or somewhat agree…

- RDS are as important as other services (82% of integral; 68% of occasional; 32% of no)
- RDS are a priority at my library (67% integral; 40% of occasional; 19% of no)
I have the skills, knowledge, and training necessary to provide RDS. (n = 159)

I have sufficient subject expertise to provide RDS to my patrons. (n = 159)

My library provides opportunities to develop skills related to RDS. (n = 159)

My library supports me to attend conferences/workshops on RDS. (n = 157)
For Those Not Currently Involved in RDS, Potential Motivations for Involvement

<table>
<thead>
<tr>
<th>Rank</th>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>If my patrons request RDS</td>
</tr>
<tr>
<td>#2</td>
<td>If RDS became a responsibility in my job</td>
</tr>
<tr>
<td>#3</td>
<td>If my institution becomes more involved with RDS</td>
</tr>
<tr>
<td>#4</td>
<td>If my institution develops an IR that accepts data</td>
</tr>
<tr>
<td>#5</td>
<td>If external funding agencies require RDS</td>
</tr>
<tr>
<td>#6</td>
<td>If RDS becomes important to subject disciplines I support</td>
</tr>
<tr>
<td>#7</td>
<td>If I learn more about RDS</td>
</tr>
</tbody>
</table>

Next *libraries* (in all types of academic libraries)…
Informational / Consulting RDS Currently Offered or Planned

- No, and no plans
- No, but > 24 months
- No, but 13-24 months
- No, but < 12 months
- Yes, currently offered

Informational / Consulting RDS Currently Offered or Planned

- No, and no plans
- No, but > 24 months
- No, but 13-24 months
- No, but < 12 months
- Yes, currently offered
Conclusions

- Most respondents do not yet do RDS
- Some have the knowledge, skills, and opportunities to provide RDS
- They believe RDS are important and consistent with library mission and role
- Libraries are at an early point in transition to RDS—requiring resetting of priorities, realignment of responsibilities, and opportunities to develop skills
Data Management services are part of transforming research libraries and librarians…

**This is next-generation librarianship.** The curation of research data is an activity that has gained traction in the wake of library and information science programs offering concentrations in data curation and institutes in digital curation, promising a cohort of librarians qualified to meet the challenges of managing data.

http://www.arl.org/rtl/erresearch/escien/nsf/leadershiproles.shtml
Authors: Patricia Hswe and Ann Holt. Part of ARL’s transforming research libraries series.