Research Data Management in U.S. Universities and Academic Libraries

Perspectives from Florida State University Libraries
Overview

Defining research data and research data management

Issues and trends in the management of research data (U.S.)

U.S. university and library responses to research data management policies

Data Management at Florida State University
What is data?

“Data are outputs of research and inputs to scholarly publications and inputs to subsequent sharing and learning” (Borgman 2007)
What is data?

“...The recorded factual material commonly accepted in the scientific community as necessary to validate research findings.” (2 CFR 200.315(3))
# Examples of Research Data

<table>
<thead>
<tr>
<th>Obvious</th>
<th>Not so Obvious</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>Lab notebooks, field notes</td>
</tr>
<tr>
<td>Spreadsheets</td>
<td>Questionnaires, survey instruments</td>
</tr>
<tr>
<td>Audiotapes, Videotapes</td>
<td>Codebooks, data dictionaries</td>
</tr>
<tr>
<td>Photographs, Films</td>
<td>Methodologies, workflows</td>
</tr>
<tr>
<td>Models, Scripts, Code</td>
<td>SOPs, protocols</td>
</tr>
</tbody>
</table>
What is research data management?

"Research data management concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information."

Whyte, A., Tedds, J. (2011)
"And yet, research data is the currency of science, even if publications are still the currency of tenure. To be able to exchange data, communicate it, mine it, reuse it, and review it is essential to scientific productivity, collaboration, and to discovery itself" (Gold 2007)
Common Data Management Issues

Obtaining original file set
Lack of data documentation
Proprietary file formats
Incomplete and/or incoherent data
The Problem of Reproducibility

2012, Amgen (pharmaceutical company) scientists report tried to reproduce 53 ‘landmark’ papers but were only able to verify the results of 6. (Nature 2012)

2015, Open Science Collaboration tries to reproduce 100 psychological studies and is only able to replicate 36 of them.
“The Administration is committed to ensuring that...the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data” (Holdren 2013).
Data Sharing Policy: Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing.

Data Management Plan Requirements: As of January 18, 2011, a separate data management plan (DMP) limited to two pages is required for all proposals submitted to NSF. The DMP should include specific details of data standards, accessibility, electronic dissemination, and sustainability.

Each directorate from the NSF has its own data management requirements.
Components of a DMP

Data types and formats
Data description (Metadata)
Data ownership and copyright
Plans for data security
Plans for preservation
2011 ACRL Survey

Libraries in large or PhD granting schools more likely to offer hand on RDS as well as consultative services

Only 20% of libraries offering support for data management plans (DMPs)

Over half of respondents had no plans to offer DMP support

Few academic libraries responsible for developing research data policies

Reassigning existing library staff most common tactic for offering data services
Survey of 73 Association of Research (ARL) Libraries in 2013

74% of respondents offer some type of research data management (RDM) support

Most developed their service between 2010 - 2012

Very few before 2009

Only 3% had no plans to offer any type of RDM support

Most important factor in creating RDM support services were NSF’s data management plan requirement in 2011
Survey of 80 Association of Research (ARL) Libraries in May 2017

51/80 respondents are currently providing data curation services at their library (~64%)

Additional 13% plan to provide data curation support in the future

Interest in data curation services has not translated into strong staff levels to provide these services
Florida State University

Research 1 Institution

$190.1 million in contracts and grants brought in by faculty in 2016/2017

~1,900 full time faculty

FSU Libraries ~ 60 librarians
RDM Activities at FSU

Data Management Plan (DMP)
Consultations
Instruction and Workshops
Tools/Resources
Collaborations/Partnerships
FSU: Initial Steps

Data Librarian hired in April 2014

Planning team: Scholarly Communication and Data Librarian

Self training

Workshops and instruction sessions
FSU: Strategic Partnerships

Office of Research
  Office of Proposal Development
  Office of Research Compliance

Graduate School

Research Computing Center
Assisted researchers at FSU with 48 Data Management Plans for Federally Funded Research

At least 11 proposals received funding

Department Of Energy Frontiers proposal: $10 million

Draft campus-wide Data Management, Authorship, and Public Access policies
Plans for future growth

Shifting approach in research data management support:

Focus advocacy rather than federal regulations

Increased confidence in skills of team and established partnerships

Working collaboratively with other offices and centers on campus

Expanded data management instruction to facilitate student success

Enhancements to DigiNole for data ingestion
Where to start?

Assess campus needs and demands

Inventory existing staff skills and capacity

Identify free/open-source resources and tools

Self education

Build partnerships
Questions/Discussion

Thank You
Works Cited


