Getting Started with Research Data Services

Jake Carlson
Director of Research Data Services
University of Michigan Library

@jrcarlso
Why are we Talking About Data?

• Push for access to results of tax-payer funded research.
  – Data management / sharing plan requirement

• Researcher’s recognizing benefits of sharing data.
  – Scholarly societies issuing statements supporting data sharing

• Trust / reproducibility issues in research.
  – Publishers requiring access to data
Why Librarians?

• Libraries connect people with information to address their needs.
  – data are an important source of information

• Libraries understand how information is best communicated to specific audiences.
  – data need structure and context to be found, understood and used
Why Librarians?

• Libraries see across communities and make connections between them.
  – data could be used by others to fuel new ideas and innovations

• Libraries think in the long term.
  – data need to be preserved to be of value to scholarship
Data Management Plans

• “The NIH expects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers.”

• “[The NSF] expects investigators to share with other researchers...within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work.”
Data Management Plans (NSF)

**Data** - samples, physical collections, software, curriculum materials, and other materials;

**Standards** - for data and metadata formats and content;

Policies for **access and sharing** – incl. IP, protection of privacy/confidentiality, security, etc.;

Policies for **re-use** – including provisions for re-distribution, and the production of derivatives;

**Archiving** - data, samples, and other research products, and for preservation of access.
Data Sharing Plan (NIH)

Varies based on data and how it will be shared

May include:
- expected schedule for data sharing
- format
- documentation to be provided
- whether or not analytic tools will be provided
- whether a data-sharing agreement will be required
- the mode of data sharing
Data Management Planning Tool
Create, review, and share data management plans that meet institutional and funder requirements.

Public DMPS
List of sample data management plans provided by DMPTool users.

Public DMPTool
Latest information about data management and the DMPTool.

Help
Overview of how to use the tool, plus resources and guidance on data management.

- Understanding the role of physician integration within nursing homes in post-acute care outcomes
- A Political Ecology of Value: A Cohort-Based Ethnography of the Environmental Turn in Nicaraguan Urban Social Policy
- A unified approach to preserving cultural software objects and their development histories
- On the right track(s) – DCC release dra...
- RDA-DMP moving and shakings
- Active, actionable DMPs
- DMPTool and RDM consultants support hum...
- Roadmap retrospective: 2016

- Frequently Asked Questions
- Create a DMP
- Administer the DMPTool
- Data management guidance
- Community resources

View All
More News
View All
Research Data Services

Your research data are important. Research Data Services is a network of services throughout the Library to assist you during all phases of the research data lifecycle. For questions about research data or to schedule a consultation, please get in touch with your subject librarian or email us.

We provide or are planning to provide services in the following areas:

- **Data Management Planning**: helping plan for managing, sharing and curating data and develop Data Management Plans (DMPs) that meet funder requirements.

- **Discovery & Access**: assisting in discovering, accessing, and acquiring different types of research materials, including data.

- **Data Organization & Management**: helping researchers to understand, develop and apply strategies for organizing and managing their data.

- **Metadata & Documentation**: locating standards for documentation that capture the details of generating, processing and analyzing data so it can be discovered, understood and reused.

- **Data Sharing & Publication**: helping disseminate research data for discovery, access and reuse in ways that enable researchers to receive credit for their work.

- **Preservation**: taking action to sustain the accessibility and scholarly value of data over time.

- **Data Visualization**: a rich and diverse set of practices, methodologies and tools from hand drawn charts to interactive web maps to immersive 3-D environments.

Learn more about Deep Blue Data
Deep Blue Data is a repository offered by the University of Michigan Library that provides access and preservation services for digital research data that were developed or used in the support of research activities at U-M.

Featured Works

- Unused Pharmaceutical...
  Deposited by: skerlos@umich.edu

- Evans Old Field Plant...
  Deposited by: cwdick@umich.edu

- The Bird's Ear View of...
  Deposited by: thomasz@umich.edu
Going Beyond Reviewing DMP

• Treating Data as a First Class Information Resource

• Learning the Data Eco-Systems of the Disciplines you Serve

• Understanding the Data Needs of your Researchers

• Using DMPs to Inform Library Services
Data as a First Class Info Resource

Include data in your catalog

In your teaching

And in your reference work

Ask a Librarian: https://flic.kr/p/9k55sb
Teaching: https://www.slideshare.net/laurenpressley/teaching-teaching-the-improbable-success-of-a-class-for-librarian-teachers
Data Eco-Systems

In a particular field of study:

• What data are important and why?
• How are they discovered & shared?
• Are there standard tools, formats, metadata, etc.?
• How are data discussed?

https://www.flex.io/blog/what-we-have-learned-from-exploring-the-data-ecosystem-part-one/
Deep Data Dive Workflow - UM

I. Data Requirements of Stakeholders
   a. Journals
   b. Funders
   c. Societies
   d. Institutes / Local Organizations

II. Data Repositories
   a. Identify Relevant Data Repositories
   b. Describe Relevant Data Repositories

III. Metadata Standards

IV. Subject-Specific Data Literature

V. Subject-Specific Culture of Data Sharing

Local Data Needs

A means to learn:
• Information about a particular data set
• What a researcher is doing to manage / curate the data set
• What a researcher would like to do with the data.

http://www.datacurationprofiles.org
A Data Curation Profile is a resource for Library and Information Science professionals, Archivists, IT professionals, Data Managers, and others who want information about the specific data generated and used in research areas and sub-disciplines that may be published, shared, and preserved for re-use. Data Curation Profiles capture requirements for a specific data set generated by a single scientist or lab, based on their reported needs and preferences for the data. More information about the Data Curation Profiles Project is available at http://datacuratoprofiles.org

The Data Curation Profiles Directory provides a stable, citable home for completed Data Curation Profiles. Each completed profile is subjected to review before being formatted, published with a DOI, and distributed through the information supply chain. More information about this process can be found in the "About" section of this site.
Using DMPs to Inform Services

DATA MANAGEMENT

knowledge

capabilities

practices

needs

Research Data Services

https://osf.io/kh2y6/wiki/home/
## The DART Rubric

### Section 2: Standards for data and metadata

The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies)

<table>
<thead>
<tr>
<th>Performance Criteria</th>
<th>Complete/detailed</th>
<th>Addressed issue, but incomplete</th>
<th>Did not address</th>
<th>Directorates w/this req.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Identifies metadata standards and/or metadata formats that will used for the proposed project</td>
<td>The metadata standard that will be followed is clearly stated and described. If no disciplinary standard exists, a project-specific approach is clearly described. “Data will be described using Darwin Core Archive metadata, and accompanied by readme.txt files providing information on field methods and procedures.”</td>
<td>The metadata standard that will be followed is vaguely stated. If no disciplinary standard exists, a project-specific approach is vaguely described. “We will observe common data format and content protocols. For behavioral and the neuroimaging data, the participant log/response files include metadata embedded within the initial data collection. In addition, the behavioral and neuroimaging protocol/stimuli delivery files discussed above document the data and metadata format and content.”</td>
<td>The metadata standard that will be followed is not stated and no project-specific approach is described.</td>
<td>All</td>
</tr>
<tr>
<td>2.2 Describes data formats created or used during project</td>
<td>Clearly describes file format standard(s) for the data. “Data will be collected in plain text, Excel, SPSS, R, Matlab, Access DB, ESRI Shapefile, TIFF, JPEG, WAV, MP4, XML, HTML, or other software-specific file formats.”</td>
<td>Describes some but not all file formats, or file format standards for the data. Where standards do not exist, does not propose how this will be addressed. “We produce data in the form of photon-counting events in a given time interval. The data is transferred to a laboratory server and stored. We also collect digital images of optical spectra from a spectrometer and store them on the server. We also collect digital samples of voltages monitoring various optical signals and store them on the server. The data will be available as raw data files.”</td>
<td>Does not include information about data format standards.</td>
<td>All</td>
</tr>
<tr>
<td>2.3 Identifies data formats that will be used for storing data</td>
<td>Clearly describes data formats that will be used for storing data and explains rationale or complicating factors. “NMR data will be saved in proprietary format to preserve embedded information, and converted to JCAMP files for ease of access and in case the proprietary systems fail or become unavailable.”</td>
<td>Only partially describes data formats that will be used for storing data and/or the rationale or complicating factors. “Complex data analysis will be done using either Origin 8.0 or Matlab software and will be saved in a format that is easily transferable.”</td>
<td>Does not describe data formats that will be used for storing data and does not explain rationale or complicating factors.</td>
<td>GEO AGS, MPS AST, MPS CHE</td>
</tr>
</tbody>
</table>
## DMP Analysis - UM

### Description of Metadata

<table>
<thead>
<tr>
<th></th>
<th>Score LSA</th>
<th>% LSA</th>
<th>Score ENG</th>
<th>% ENG</th>
<th>Score All</th>
<th>% All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete / detailed</td>
<td>13</td>
<td>26%</td>
<td>6</td>
<td>12%</td>
<td>19</td>
<td>20%</td>
</tr>
<tr>
<td>Addressed issue, but incomplete</td>
<td>10</td>
<td>20%</td>
<td>19</td>
<td>38%</td>
<td>29</td>
<td>31%</td>
</tr>
<tr>
<td>Did not address the issue</td>
<td>22</td>
<td>44%</td>
<td>24</td>
<td>48%</td>
<td>46</td>
<td>49%</td>
</tr>
<tr>
<td>(No Data)</td>
<td>5</td>
<td>10%</td>
<td>1</td>
<td>2%</td>
<td>6</td>
<td>6%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>50</strong></td>
<td><strong>100%</strong></td>
<td><strong>50</strong></td>
<td><strong>100%</strong></td>
<td><strong>100</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Thank You!
Questions?

Jake Carlson
Director of Research Data Services
jakecar@umich.edu
@jrcarlso