Data Management
Effective Data Management

JHU researchers: contact a data management consultant for help specific to your data management plan. The NSF now requires that all proposals submitted must include a data management plan. Successful proposals will be expected to execute that data management plan. Effective data management can significantly impact how projects operate and deliver in research environments.

The JHU Data Management Services (JHU DMS), offered by the Sheridan Libraries, has been set up to help alleviate this burden from the researcher, freeing up valuable time and resources so you can focus on your research. JHU DMS offers direct support for data management planning and assists with the generation of Data Management plans. JHU DMS, backed by the Data Conservancy, provides a state-of-the-art digital archive for both short-term and long-term data archiving. For information specific to your grant proposal and NSF directorate, contact your Data Management Consultant.

Explore the links below for specifics about the NSF data management requirements and the services available to you as JHU researchers:

What is data management?
- What does data management planning mean for you in terms of NSF grants?
- What kinds of information does a data management plan cover?
- Where can I archive my data?

What services are available?
- Overview of people involved in data management who can help me
- About the JHU Data Management Services

What are the implications?
- Legal and ethical issues
Lifecycle Data Management Planning

Why manage research data? Data are information assets that greatly increase in use value through active management. By planning the management of your research data at proposal time it is possible to:

- Secure funding specifically for research data management
- Improve the impact and visibility of your research
- Improve and standardize data management practice and policy in your lab
- Facilitate collaboration, increase research efficiency, and make new discoveries
- Assure the greatest return on investment by adapting a value chain model


Says Who?

Lifecycle management of data is becoming increasingly important to funding agencies and many agencies now encourage data management, re-use, and sharing plans. Some major funders even require a formal "Data Management Plan" (DMP) in order to be competitively considered for funding. Below is a list of agencies that have encouraged or mandated that grant proposals include a Data Management Plan:

<table>
<thead>
<tr>
<th>US Federal Funding Agency</th>
<th>Policy and Guideline Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Department of Energy (DOE)</td>
<td>DOE's CIO has primary responsibility to ensure that Information Technology (IT) is acquired and information resources are managed in a manner consistent with statutory, regulatory, and Departmental requirements and priorities. With this responsibility, the CIO provides information resources management advice and assistance to the Secretary of Energy and to other senior managers.</td>
</tr>
<tr>
<td>Environmental Guidelines</td>
<td>DOE Policies</td>
</tr>
<tr>
<td></td>
<td>Standard Research Terms and Conditions</td>
</tr>
<tr>
<td></td>
<td>ARM Data Sharing and Distribution Policy</td>
</tr>
<tr>
<td></td>
<td>Developing Data Management Policy and Guidance Documents for your NARSTO Program</td>
</tr>
<tr>
<td></td>
<td>Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection</td>
</tr>
</tbody>
</table>

\[10/5/11 11:51:13 AM\]
<table>
<thead>
<tr>
<th>Agency</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental Protection Agency (EPA)</td>
<td>EPA does not have an official data management and sharing policy.</td>
</tr>
<tr>
<td>Institute of Museum and Library Services (IMLS)</td>
<td>IMLS does not have an official data management and sharing policy.</td>
</tr>
<tr>
<td>National Aeronautics and Space Administration (NASA)</td>
<td>&quot;NASA promotes the full and open sharing of all data with the research and applications communities, private industry, academia, and the general public.&quot;</td>
</tr>
<tr>
<td>National Oceanic and Atmospheric Administration (NOAA)</td>
<td>&quot;The Federal Ocean Data Policy requires that appropriate ocean data and related information collected under federal sponsorship be submitted to and archived by designated national data centers.&quot;</td>
</tr>
<tr>
<td>National Endowment for the Humanities (NEH)</td>
<td>NEH does not have an official data management and sharing policy.</td>
</tr>
<tr>
<td>National Institutes of Health (NIH)</td>
<td>&quot;The NIH expects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers. ...[I]nvestigators submitting an NIH application seeking $500,000 or more in direct costs in any single year are expected to include a plan for data sharing or state why data sharing is not possible.&quot;</td>
</tr>
<tr>
<td>National Science Foundation (NSF)</td>
<td>&quot;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.&quot;</td>
</tr>
<tr>
<td></td>
<td>&quot;Proposals submitted to NSF must include a supplementary document of no more than two pages labeled ‘Data Management Plan’. This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results.&quot;</td>
</tr>
</tbody>
</table>
Plan for it!

Throughout this guide you will find tips, recommendations, and prompts that will help you write a Data Management Plan for your research project. These recommendations are meant to address the primary concerns of grant funding agencies, but they will likely inform practice and policy in your own research environment whether or not you are writing for a grant. This guide is an attempt to highlight and allow researchers to answer the questions raised in the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council’s report “Harnessing the Power of Digital Data for Science and Society” (January 2009) by mapping such concerns to the Data Documentation Initiative (DDI) Version 3.0 Combined Lifecycle Model. This guide is intended for MSU students, faculty, administrators and colleagues.

Your Data Management Plan will need to cater to your research, domain, and available resources. However, any project which generates a significant amount of digital research data should include a Data Management Plan.

General DMP Template:

- A general description of the data
  - type
  - size
- A claim that expresses the value and impact of these data
  - value
  - impact
- A specific description of the content and format of data
  - identification (filenameing)
  - data format
  - data fidelity
  - metadata standards
  - metadata quality
  - methodology
  - processing
- Any provisions for protection of data
  - intellectual property
  - ethics
- Any restrictions on access
  - levels of access
  - sharing
  - cite
- A specific description of the preservation environment
  - short term storage
  - long term storage
- Provisions for transfer of responsibility
  - succession planning
Welcome to the Data Management, Curation and Archiving Research Guide. This guide is intended primarily for researchers and data librarians to facilitate the management, sharing and archiving of their data. This guide covers a rapidly changing large and complex topic. We will regularly update, correct and add information and tools to the guide. Please let us know if you do not find what you need, feel that some information or concepts are incorrect, or want to add information from your field of research. Also, you may want to check back periodically for updates.

As research data librarians, we are here to help you:

- Create data management plans for grant proposals
- Manage, curate and archive your data.
- Maximize the usefulness of your data.
- Increase the lifespan of your data.
- Manage sharing of your data.
- Prepare your data for archiving.
- Collaborate with you in creating innovative new ways to share your data.

This Research Guide should give you the necessary background and tools to better manage your data. We also hope it will help you in understanding what supporting information (metadata) will help increase the usability, understandability and longevity of your data.

Please do not hesitate to contact the research data librarian specializing in your field of research for more information or assistance.

Why Manage Your Data?

The rapid increase in the use of digital formats for data has allowed data to be shared more easily among researchers. Often this sharing is between close collaborators, but data is also increasingly being made open for anyone to use. There are many benefits to sharing data. Shared data can...

- Be combined to answer otherwise unanswerable questions
- Recent advancements in Alzheimer’s research made possible by making data open for use. (NY Times article)
- Be reused in unexpected ways
- Be used to plan new studies
- Help avoid repeating studies
- Help avoid repeating mistakes
- Allow data to be audited

Although there are many benefits to digital data, there are also potential problems.

- First, digital formats can quickly become obsolete, corrupted or otherwise lost (Digital Dark Age).
- Average life expectancy for born digital data has been estimated to be as short as five years.
- We are already losing vast amounts of data.
- While digital data is easily shared, finding data without firsthand knowledge of its existence is difficult.
- Understanding data without the participation of its creator is often impossible.
- Researchers are often poorly positioned to manage and curate their data.
- Researchers are focused on using data for publishing results.
- Researchers usually have no background in digital preservation or data curation.