Mastering Data Management Plans

Karl Nilsen, Research Data Librarian
Robin Dasler, Engineering/Research Data Librarian
Alex Carroll, Agriculture and Natural Resources Librarian
Jeremy Garritano, Life Sciences/Research Data Librarian

Updated 2014-10-09
Research Data Services

We help researchers manage, curate, publish, and preserve data and other research products

lib.umd.edu/data
lib-research-data@umd.edu

Our writing guide: lib.umd.edu/data/dmp
Data management plans

Describe what data you will collect or generate and how you will format, store, document, share, and preserve them

Essentials:

- Manage and share
- During and after your project
Data sharing/availability plans

Requested by journals and certain funding agencies

Similar to data management plan, but primary focus on data sharing
Do you have to write a plan?

Almost certainly

Can say “no data collected”, but consult agency or journal instructions

Collaborative proposals or subaward proposals: contact the program officer
Layers of requirements

A solicitation may have requirements in several places. Follow the requirements in this order:

1. Solicitation
2. Directorate, division, program, or office instructions
3. Agency general proposal requirements
Agency motives

Support reproducibility, replication, and validation

Stimulate new research, analysis

Assemble new datasets, models
What counts as data?

Data, images, software, models, publications, samples, physical collections, etc.


- Consult agency or journal instructions
What counts as data?

Focus on

1. Material necessary to reproduce your results

2. Material that is broadly useful to other researchers in your field and beyond
What does “share” mean?

Make your data, code, and other materials available to other researchers at some point

- Consult agency or journal instructions

Ideally in public-access repositories, archives, databases, or data centers

PI or team website is acceptable (but potentially less stable in the long run)
Special situations affect sharing

- Confidential or sensitive data
- Data that will support commercialization
- Massive volumes of data
- Data already available elsewhere
- No data produced or generated

Explain your situation
What to share?

Focus on

1. Material necessary to reproduce your results

2. Material that is broadly useful to other researchers in your field and beyond
When to share data

- Consult agency or journal instructions

Sharing data when you publish an associated article is normal, but customs vary by field

Long delays require justification
Who reviews your plan?

Depends on the agency or journal

For funding agencies, plan should be convincing to the peer reviewers
Impact on future proposals

Final report should address outcome of your data management plan

Some agencies request information about previous data management and sharing activities in “results of prior support”
Writing process

Organize your plan according to the agency instructions (solicitation, program, and agency requirements)

Consult our writing guide to help you fill in the sections: lib.umd.edu/data/dmp

Contact us for advice: lib-research-data@umd.edu
Basic data management plan I

1. Who will be responsible for managing and sharing data?

   • Basic problem is that risk of loss or error increases with team size

2. What data do you expect to collect or generate?

3. Of all the data you generate, which data will you keep and share with other researchers? In what formats?
Basic data management plan II

4. What documentation will you provide to help other researchers understand your data?

- General information about your project, inventory of data files, data collection methods, spatial and temporal extent, processing steps, meaning of column headers, meaning of codes or abbreviations, terms and conditions of use, software required, and so on.
5. How will other researchers get your data? When?

● Public repository, archive, data center, database
● PI or team website
● Multiple repositories sometimes necessary
Avoid

“Data will be available on request”

Use only as a last resort or if the volume, complexity, or security requirements of your data make it impossible to use a public archive or data center.

Contact the program officer or journal editor before writing this in your plan.
General purpose data repositories

DRUM (UMD Libraries) http://drum.lib.umd.edu

Zenodo (CERN) https://zenodo.org/

Dataverse (Harvard) http://thedata.harvard.edu/dvn/

Dryad (UNC-NESCent-NCSU) http://datadryad.org

figshare (Macmillan) http://figshare.com
General purpose code repositories

GitHub https://github.com (git)

Bitbucket https://bitbucket.org/ (git, mercurial)

SourceForge http://sourceforge.net

Can also use DRUM, Zenodo, Dataverse, etc. for code
Basic data management plan III

6. Are there any ethical, legal, regulatory, contractual, or technical issues that will prevent you from sharing data?

7. Are there any terms or conditions that will affect how other researchers can use your data? (e.g. intellectual property terms, licenses, or attribution and credit expectations)
8. Where will you archive data for long-term preservation? For how long?

Typically, the repository that you use to share data will provide long-term preservation. What is their policy on retention?

Funding agency or journal may specify minimum retention period, but consider future value of your data.
Budgeting

Depending on agency, you may be able to request funds for data curation and preservation in your proposal budget.

Key factors include volume of data, period of retention, and cost of storage and curation.
Getting started

Review solicitation, division/program instructions, agency proposal requirements, and FAQ

Identify a potential public-access archive, database, or data center before you start writing—it will give you content for your plan

Start early—and contact us early:

lib-research-data@umd.edu