Mastering Data Management Plans

Karl Nilsen, Research Data Librarian

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

Example plans, boilerplate language, and comprehensive writing guide: lib.umd.edu/data/dmp

Data management plans

Describe what data you will collect or generate and how you will format, store, secure, 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

Requirements are changing now

Most federal funding agencies are revising their requirements

New/updated requirements expected 2016-2018

Contact us before you start writing for guidance

Do you have to write a plan?

Almost certainly

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

Collaborative proposals, omnibus grants, or subaward proposals: contact the program officer

Layers of requirements

A grant 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

Describing data mgmt elsewhere

Some projects require extended data management documentation (e.g. spacecraft missions, data integration projects, database or data archive development, etc.)

When data management is described extensively in the project description, intellectual merit, or other supplemental sections, a “data management plan” is still required, but you can indicate that data management is addressed elsewhere and use the plan to describe materials associated with scholarly outputs such as publications and presentations.

Agency motives

Support reproducibility, replication, and validation

Stimulate new research, analysis

Assemble new datasets, models

Public access to taxpayer-funded research

What counts as data?

Agency instructions often vague

e.g. NIH: “The definition of digital scientific data includes data that are used to support a scientific publication as well as data from completed studies that might never be published”

e.g. NSF: “may include, but is not limited to: data, publications, samples, physical collections, software and models”

http://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf

http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp#1

What counts as data?

All your data? Raw? Converted? Derived? Subset? Summary? Supplemental?

- Usually not all
- Consult agency or journal instructions

e.g. NSF: “Rarely does NSF expect that retention of all data that are streamed from an instrument or created in the course of an experiment or survey will be required.”

http://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf

What counts as data?

Focus on

1. Material necessary to validate and reproduce your findings

- esp. data used to produce figures and summary statistics

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, or databases (some agencies run their own repositories) providing a citable unique identifier (e.g. DOI)

Avoid using PI or team website for ‘archiving’ or promising “data available on request”

What does “share” mean?

e.g. NIH: “NIH will expect funded researchers to deposit data in appropriate, existing, publicly accessible repositories before considering other means of making data available.”

http://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf

Special situations affect sharing

- Confidential or sensitive data (human subjects, endangered species, BSAT, ITAR, classified, etc.)
- IRB conditions
- Data that will support commercialization (SBIR, STTR, patents, etc.)
- 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 validate and reproduce your findings

- esp. data used to produce figures and summary statistics

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

What to share?

NASA: “ we can only recommend that proposers use their best judgment and the standards of their community in deciding what should be archived to allow others in your community to really understand what you have done.”

http://science.nasa.gov/researchers/sara/faqs/dmp-faq-roses/

When to share data

- Consult agency or journal instructions

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

Some fields have standard embargo periods

Long delays require justification

When to share data

NASA: “Lets say that your paper is based on a big dataset that you generated and this is just the first of a series of articles. In that case we certainly won't require you to release the entire dataset along with the first publication. On the other hand, you do have to release the data that was represented in that paper, and we expect you to release the full dataset at the end of the award.”

NOAA: “data must be shared no later than 2 years after the time of observation”

http://science.nasa.gov/researchers/sara/faqs/dmp-faq-roses/
https://geo-ide.noaa.gov/wiki/index.php?title=FAQ_for_Data_Sharing_for_NOAA_Grants_PD

When to share data

DOD: “Data underlying the conclusions of peer-reviewed journal publications will be made freely available at the time of initial publication.”

NIH: “expects that the data will be shared at the time of acceptance for publication.”

http://www.dtic.mil/dtic/pdf/dod_public_access_plan_feb2015.pdf
http://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf

Who reviews your plan?

Depends on the agency or journal

For funding agencies, plan should be convincing to the peer reviewers

Program officer has some discretion

Bad plans don’t sink applications (yet), but you may be asked to re-write it

Impact on future proposals

Compliance will affect future funding (and continuing funds)

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

Final report should address outcome of your data management plan

- Your plan can change—but describe changes in interim/final reports

Demonstrating compliance

1. Deposit data in public-access repositories, archives, data centers, or databases

2. Provide a citation with the unique URL (e.g. DOI) generated by the repository

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?

    - Risk of loss or error increases with team size and/or student inexperience

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, protocols, data collection methods, spatial and temporal extent, calibrations, processing steps, meaning of column headers, meaning of codes or abbreviations, terms and conditions of use, software required, etc.

Basic data management plan III

5. How will other researchers get your data? When?

- Public repository, archive, data center, database
- Multiple repositories sometimes necessary
- Some funding agencies (NIH, NASA, NOAA) operate their own repositories and may stipulate where to archive data, code, etc.

e.g. “NOAA facilities that archive data and make the data openly available should be considered first for the disposition of the data.”

https://nosc.noaa.gov/EDMC/documents/EDMC-PD-DSP.pdf

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

Avoid

“Data will be available on the PI/team’s website”

Use only as a last resort

Funding agencies prefer repositories backed by stable organizations or institutions

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

Open Science Framework (Center for Open Science) https://osf.io/

General purpose code repositories

GitHub https://github.com (git)

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

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, data-use agreements, licenses, or attribution and credit expectations)

Basic data management plan IV

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, journal, or gov regulations may specify minimum retention period (e.g. HIPAA), 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