Complete Guide to Writing Data Management Plans

University of Maryland Libraries

v2.1
Last updated November 3, 2017.

Contents

Introduction

This guide outlines a writing strategy for creating a data management plan based on requirements common to many funding agencies. Some of the advice in this guide also applies to data sharing plans or data availability statements required by journals and certain funding organizations. Several of the resources provided in this guide are intended for members of University of Maryland, College Park campus community, but the advice and guidelines apply beyond the campus.

Before you start writing, please review carefully the agency or journal's requirements for official instructions. This guide assumes that you have read and understood the agency or journal's instructions and have determined which requirements apply to your proposal.

1. Getting started

1.1 What is a data management plan?

A data management plan usually describes the data, code, and other research products you will produce and how you will format, document, store, and preserve them. A plan also describes what data you will share with other researchers and how you will distribute those materials.

Funding agencies and journals increasingly expect that you will, as much as possible, share data, code, and other products with other researchers. The agencies and journals believe that data sharing supports reproducibility and replication, which are essential to the integrity and progress of science and scholarly inquiry. For this reason, your plan should describe data management and sharing during your research and, importantly, after your research is complete.

> **Note:** *If ethical, legal, contractual, or technical conditions prevent you from sharing or distributing data, you may still have to submit a data management or sharing plan. Consult the agency or journal's requirements for official guidance. In this situation, your plan could explain why you cannot share data*

Many funding agencies and sponsors require a data management plan with each proposal, but any researcher or team will benefit from developing a data management plan at the beginning of a project. Developing a plan is an excellent way to identify useful and important records, optimize your data handling process, and anticipate issues that may arise in publishing, archiving, and preservation.

1.2 Using this guide

This guide covers topics that frequently appear in data management and data sharing plans, but each funding agency or journal has its own guidelines. Include only the information requested by the agency or sponsor in its official instructions.

Sections 2-4 of this guide provide general information and guidelines for data management plans, whereas sections 5-8 provide guidance on how to develop content for a typical data management plan. Sections 5-8 are divided into a series of basic questions; your answers will provide the content for your plan. Not all questions will be relevant to your research.

Typical data management plan topics include:

- Roles and responsibilities
- Types of data, and what data will be kept
- Format and documentation
- Access and sharing
- Re-use and re-distribution
- Long-term archiving and preservation, and budgeting for these activities

2. General advice

2.1 Identify all relevant requirements
A grant solicitation may have data management requirements in several places. Check for requirements in this order:

- Solicitation
- Directorate, division, office, or program
- Agency general proposal instructions

2.2 Use the norms and standards in your field
In most cases, your data management strategy will reflect the unique needs of your project and the prevailing norms and practices in your field. You should make reference to those norms and practices whenever appropriate.

It can be helpful to identify any relevant rules, standards, or codes of practice that will affect how you manage and share data. This demonstrates that your data management practices are consistent with the standards in your field.

2.3 Amount of detail and required length
The amount of detail in your plan will depend on the agency or journal requirements and the characteristics of your research project. Consult the agency or journal for directions. While data management plans are typically short, you should provide more detail whenever your plan describes unique, special, or especially complex situations.

2.4 Criteria for data retention
Before you start writing your plan, you should consider the usefulness and long-term value of your research products. The following kinds of data usually have high value and should be managed and retained accordingly:

- Data necessary to understand your work and validate, replicate, or reproduce your findings
- Unique data not easily or cheaply recreated, or data that are impossible to recreate
- Data that are broadly useful in your discipline and beyond (e.g. social or environmental observations)
- Data that you or your students may re-analyze in the future
- Data that support property claims such as patents
- Data that you are compelled to retain for regulatory, legal, ethical, institutional, or contractual reasons

Generally, funding agencies and journals do not expect you to retain all your data. However, your plan should explain why you will retain certain data and discard others. Consult your funding agency or journal's requirements for official guidance.

3. Advice on data sharing

3.1 Determining what data to share
Assuming there are no conditions that prevent you from sharing data, consider sharing at least the following data:

- Data necessary to understand your work and validate, replicate, or reproduce your findings
- Unique data not easily or cheaply recreated, or data that are impossible to recreate
- Data that are broadly useful in your discipline and beyond (e.g. social or environmental observations)
- Data that are not otherwise available from public repositories, data centers, or archives

3.2 Dealing with multiple investigators, institutions, and sponsors
If your research involves multiple investigators or teams, either domestic or international, your plan should describe how you will harmonize and synchronize data management and post-project data sharing. At a minimum, indicate who is responsible for data management and sharing.

If your research involves multiple funding sources or partnerships, your plan should describe how you will accommodate and balance the data management expectations of the different sponsors or partners.

3.3 If your work will not produce data
In most cases, you have to submit a plan with every proposal, but it is acceptable to state that your project will not produce data. [Here's a good example](#).

However, if you are working on a capacity-building project or educational program, explain how you will manage the products of that work. For example, if you record videos of a workshop, how will you manage and share them?

4. Resources for managing and archiving data

4.1 Public-access repositories, archives, and databases

Ideally, you should use a public-access repository, archive, data center, or database to share and preserve data. If possible, identify a potential repository and review its submission requirements before you start drafting a plan. The submission requirements will shape your data management strategy and provide material for your plan. In some cases, you may have to use different repositories for different types of data. The University Libraries can help you locate a repository.

> **Tip:** *If you cannot find an appropriate data repository or archive, contact the program officer or editor for direction?*

A funding agency, funding organization, or journal may operate or sponsor a data repository. Contact the program officer or editor for recommendations. Peers in your field may maintain a repository. Inquire on any mailing lists or forums. Ask a librarian, they can do some research for you, too. If there is no central repository for your field, you may need to investigate general purpose repositories. Here are a few suggestions:

- [Digital Repository at the University of Maryland](#) (DRUM) is managed and maintained by the University of Maryland Libraries and is available for faculty and affiliates.
- [Zenodo](#) is maintained by CERN.
- [Dataverse](#) is maintained by IQSS at Harvard.
- [Dryad](#) is maintained by UNC-NESCent-NCSU.
- [Open Science Framework](#) is maintained by the Center for Open Science.

For software code, we recommend:

- [GitHub](#) (for use with Git version control)
- [BitBucket](#) (for use with Git or Mercurial version control)

Any of the data repositories listed above will also accept software code.

4.2 Useful services at UMD

Data sharing and preservation	[Digital Repository at the University of Maryland](#) (DRUM) **email:** lib-research-data@umd.edu
Data storage and backup	[Division of Information Technology](#) **email:** storage-help@umd.edu
Grant proposals	[Office of Research Administration](#)
Copyright and intellectual property	[Maryland Intellectual Property Legal Resource Center](#)

| Commercialization and patent applications | [Office of Technology Commercialization](#) |

5. Typical DMP content: Roles and responsibilities

Identify the individuals who will collect, organize, process, analyze, and share data. Outline their basic responsibilities.

> **Question:** *Who will manage data during your project?*

- If various collaborators and students will be managing data, how will you monitor their work?
- If you will be collaborating with researchers at other institutions, how will you harmonize and synchronize data management?
- If a PI or co-PI leaves the institution, how will you ensure that data and documentation are not lost?
- How will you transfer responsibility to another member of the team?

6. Typical DMP content: Types and formats of data

6.1 Types of data, and what data will be kept

Describe the data, code, and other research products produced. Use this section to outline how you will store and manage data during the project. You should also identify what data you will retain and share after the project is complete.

Funding agencies and journals have different definitions of 'data', so consult the official instructions to determine what materials count as data.

Question: *What types of data will you produce, how much, and for how long?*

– Consider the volume and variety of data.

Question: *What are the data sources?*

– Are you collecting data yourself or using publicly available data from open-access repositories or data centers?

Question: *What instruments or software are involved?*

Question: *What is your plan for data storage, security, and backup during your project?*

– If you have IRB approval for your project, you may be able to adapt this information from your IRB content.

Question: *Of all the data, you will collect or produce, what data will you retain after the project is finished, and why?*

– Consider what data are necessary for replication and what data may stimulate new research in your field and beyond (See 2.4 Criteria for data retention for additional considerations).
– If possible, reinforce your decision to retain data with reference to potential user communities. Who could use your data?
– If you choose to discard certain data, explain why.

6.2 Format and documentation

For this topic, describe how you will format and document your data. Data formats refer to the data structures and file formats that you use to save, transfer, and share

data. Documentation refers to all the information about your data that another researcher would need to understand and use your data for replication, reference, or new research.

Depending on how you document your research, you may already collect this information in a codebook, data dictionary, readme file, metadata file, or lab notebook. Typically, the documentation will include general information about your project, data collection methods, data processing, meaning of any codes or abbreviations, terms and conditions of use, software required, inventory of data files, and so on.

Note:	*You do not have to include any documentation in your plan, only a description of your method of documentation.*

Tip:	*If you identify a potential public-access repository, data center, or archive before you start writing your plan, the data managers can often direct you to specific data formats and documentation standards. Please refer to "Public-access repositories, archives, and databases" for suggestions and recommendations.*

Question:	*What data structures and file formats will you use to capture and store data during your project?*

Question:	*What data structures and file formats will you use to share data after your project is finished?*

- Many commercial software formats and instrument formats are not suitable for public access and long-term preservation because they can only be opened and manipulated by the software or instrument that created them.
- See our [format recommendations](#) for platform-independent alternatives.

Question:	*How will you record documentation for your data? Will you use a standard form of metadata?*

- In some fields, metadata is highly standardized and requires specific information. If this applies to you, identify the standard.
- If there is no commonly used metadata standard appropriate to your situation, state that fact and describe how you will document your data.

Question:	*Where will you store and backup your documentation?*

7. Typical DMP content: Access and re-use

7.1 Access and sharing
This is the most important part of your plan. Describe how you will share data with reviewers and other researchers. Funding agencies and journals have different expectations for data sharing, so consult the agency or journal's official instructions.

Question: *What data will you share with other researchers?*

– If you use data from a public-access repository, you may be able to refer people to the original data rather than distribute it yourself. In this case, you should provide links to these data in any documentation and publications.

Question: *Who will have access to your data?*

– Common audiences to consider in this situation are other researchers (in your field and beyond) and the general public.
– In some cases, depending on the nature of your project, you can share data with both groups without restriction.
– In other cases, you may be able to share data with other researchers but not the general public. Explain any such conditions.

Question: *How will other researchers find and obtain your data after your project is complete?*

– Please refer to [4.1 Public-access repositories, archives, and databases](#) for data repositories, including the [Digital Repository at the University of Maryland](#) (DRUM).
– If you have identified a public-access repository, data center, or archive for your data, you will have to comply with their policies and requirements for access and sharing. Note any restrictions on access.
– *Avoid using your personal website, or your team's website, to share data.* Funding agencies increasingly view this method as potentially unstable in the long run. It's preferred to deposit your materials in a dedicated repository or archive and then use the links provided by the archive to share data (on your site or wherever).
– Depending on the nature of your data or the availability of public-access repositories, you may have to stipulate that your data "will be available on request." However, you should avoid this method if possible. Funding agencies and journals are increasingly dissatisfied with this method, viewing it as a barrier to efficient public access. If you are compelled to take this approach, contact the program officer or editor in advance for guidance.
– If you cannot find an appropriate data repository or archive, contact the program officer or editor for direction.

Question: *How soon will other researchers or the public have access to your data?*

- − Consult the funding agency or journal's requirements for guidelines.
- − If there is no explicit length of time in the official instructions, answer this item with reference to the customary practices in your field.
- − Making your data available when you publish associated findings is typical, but norms vary by field. Delays that exceed customary practices will require more substantial justification.

Question: *If you produce confidential or sensitive data, how will the measures you take to protect subjects affect public access?*

Question: *If you are working under the terms of an IRB, how will they affect public access?*

Question: *Are there any additional federal, institutional, professional, or sponsor regulations that will affect public access?*

Question: *Will you have any special security provisions or data use agreements?*

Question: *Are there any intellectual property issues, such as ownership, copyright, or potential commercialization, that will affect public access?*

- − For research products generated under federal agency awards, all intellectual property developed by researchers and students and all intellectual property rights therein shall belong to the University unless an exception or waiver is granted.
- − In many cases, this will not prevent you from sharing data and other materials with researchers or the general public, but conditions apply when your activities involve materials transfer, inventions, patents, royalties from inventions, third-party contracts, and other special circumstances.
- − Contact the [Office of Research Administration](#) for guidance on your situation.

7.2 Re-use and re-distribution
For this topic, describe any terms or conditions of use, including reproduction, distribution, or creation of derivatives.

Question: *Are there any intellectual property issues that will affect re-use and re-distribution?*

- If your project uses data, software, or materials that belong to another individual, group, organization, or institution, you must comply with any terms, conditions, permissions, licenses, or agreements specified by the data owner. Note any conditions that affect re-use, such as restrictions on data sharing.
- If you have identified a repository or archive for your data, you will have to comply with their policies and requirements for re-use and re-distribution. Note any conditions that affect re-use.

Question: *Will you make your data available with specific terms and conditions, licenses, restrictions, or disclaimers?*

- If there are no legal, ethical, or contractual issues that will affect re-use or re-distribution, the simplest option is to refrain from adding any terms or conditions.
- However, you may wish to insist upon attribution, citation, or another form of credit whenever someone uses your data.

8. Typical DMP content: Long-term preservation

For this topic, describe the long-term disposition of your data, code, and other research products. If you plan to deposit your data at a repository, data center, or archive, your response for this section may overlap with information in [7.1 Access and sharing](#).

Question:	*Will you submit your data to a repository for long-term archiving and preservation? Which one?*

- In some cases, you may have to use different repositories for different data types.
- See our suggestions above for data repositories.
- If you cannot find an appropriate data repository, contact the program officer or editor for direction.

Question:	*For how long will you (or a repository) preserve your data?*

- This depends on wide variety of factors. Consult the funding agency or journal's requirements for official guidelines.
- All data produced with federal grants should be retained for a minimum of three years ([OMB](#)).
- UMD's retention policy for most research records is a minimum of seven years after the completion of research. Different terms apply to investigational new drugs and investigational devices ([UMD Records Schedule](#), Item 84).
- If you conduct research under HIPAA regulations, you should plan to retain data for a minimum of six years.
- Data related to patents should be retained for the life of the patent.
- In addition, consider the potential value of your data in temporal terms: will the value increase, decrease, or remain constant over time? For example, social and environmental observations that cannot be recreated may increase in value.

Question:	*If there are costs associated with long-term archiving and preservation, such as deposit fees at a repository, how will you cover them?*

- You may be able to request funds in your proposal budget. Consult the funding agency or journal's requirements for official guidelines.