

Preparing data management plans for NSF grant applications

A guide to the NSF [policy](#) for data management plans



Science Libraries
University of California, Berkeley



Contents

The importance of a plan	2
The purpose of a plan	3
Identify NSF requirements	3
Learn data management basics	5
Examine data management activities and find ideas, tools, and services for your plan	6
Understand your data	7
Organize and describe data	8
Determine the approach to sharing	9
Find tools for sharing	11
Store data	12
Find inspiration from sample plans	13
Write your plan with a template	13



February 2011

1. The importance of a plan

A requirement

NSF grant proposals now require a supplementary document known as a "data management plan."

In this plan, you describe how your project will conform to the NSF [policy](#) for disseminating and sharing research results. The plan will undergo peer review and is an integral part of the proposal.

If no such plan is included, NSF will return your proposal without review.

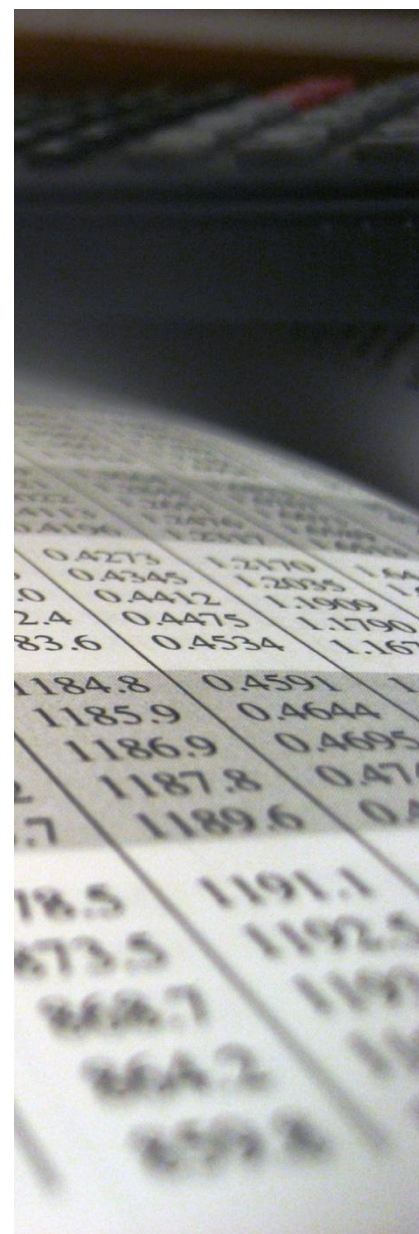
An NSF FAQ responds to your questions about [what](#) research results will be shared, [when](#) to share them, and sharing information that is [sensitive](#) or has potential [commercial value](#).

The value

Besides fulfilling the NSF requirement, a data management plan may help you organize your data and cultivate quality and efficiency in your data work. In the long run this may save you time and resources. Furthermore, data planning may help with preserving and sharing your data.

Sharing research products is increasingly important ([Nature](#), 2009; [Science](#), 2011).

- National funding agencies have data sharing [policies](#), and scientific journals like [Nature](#), [Science](#), and [PLoS](#) have sharing requirements.
- Distributing your research results may enhance your work because of increased citations ([Piwowar et al.](#), 2007) and the public record of your research.
- A shared, common data set may help researchers collaborate and accelerate discoveries in complex fields ([NY Times](#), 2010).



2. The purpose of a plan

In general, data management plans address the following topics.

Nature of the research products	The characteristics of the “data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project” (NSF , 2011)
Sharing	The data and research products that will be available to others
Access	The ways that researchers will be able obtain your data and research products
Archiving	The long term storage and organization of your data and research products



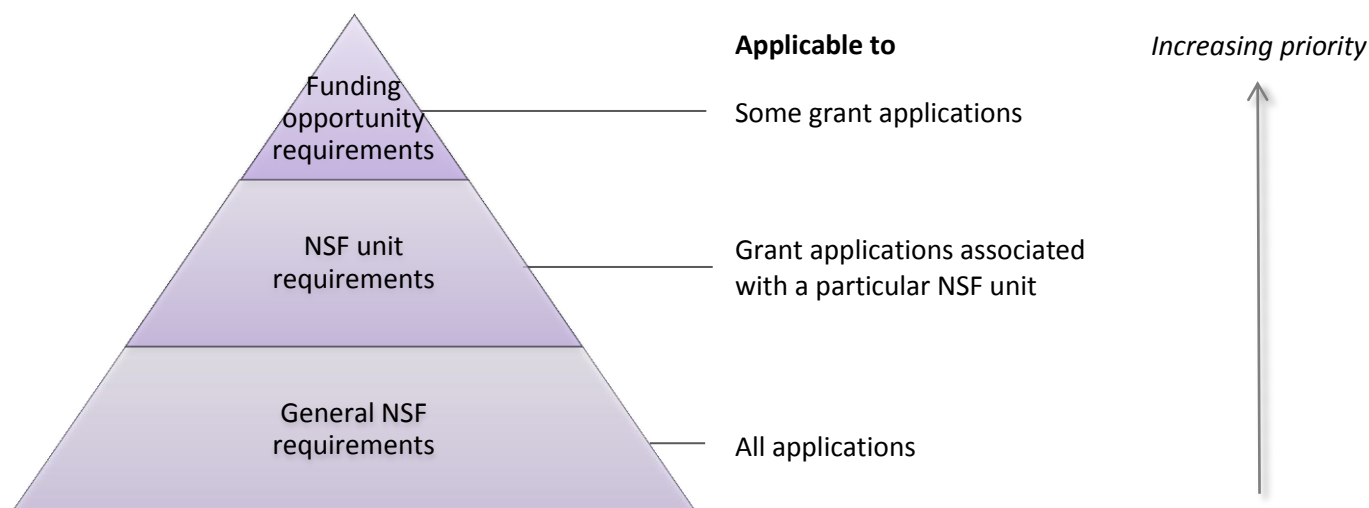
3. Identify NSF requirements

Length

The data management plan is a supplementary document to your NSF proposal. There is a 2-page limit.

Specific guidance and requirements

Guidance and requirements for your data management plan are provided at three levels.



The three levels of NSF requirements in detail.

Funding opportunity requirements If your funding opportunity synopsis or solicitation has specific guidance or requirements, follow them as your top priority.

NSF unit requirements Your funding opportunity may be associated with an NSF unit that has specific guidance or requirements. These instructions are listed under "[Requirements by Directorate ...](#)"

If your NSF unit is not listed, then follow the general requirements.

General NSF requirements Your plan needs to describe how your proposal conforms to the [NSF policy](#) on the dissemination and sharing of research results.

In general, the NSF is looking for discussion on the following topics:

1. Products of the research
2. Data and metadata formats (Metadata are information that describes the context and content of your data – for example, date of creation, subject, author name, version, and more.)
3. Policies and practices for data access and sharing
4. Policies and provisions for re-use, redistribution, and the production of derivatives
5. Plans for archiving your research products and preserving access



Tips

Exclusion of a plan

You can indicate that no detailed plan is needed, but you will need to provide a clear justification that will undergo peer review and program management.

For comprehensive instructions, read the NSF [Grant Proposal Guide](#).

4. Learn data management basics

What is data?

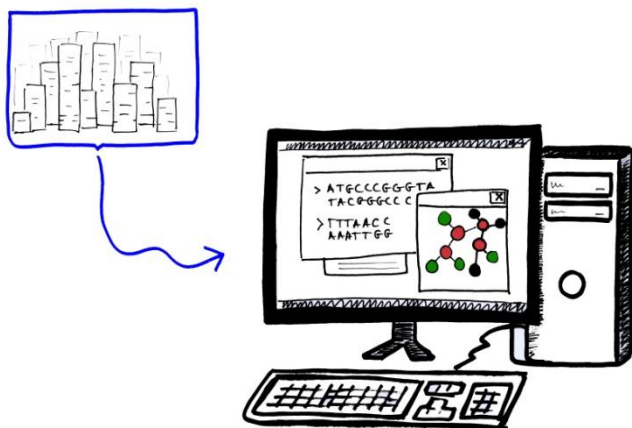
According to the NSF, what constitutes data and research products will be determined by the community of interest through peer review and program management ([NSF](#), 2011). Research products include:

- data
- samples
- physical collections
- software
- curriculum materials
- [other materials produced](#)

For examples of research data and products and what are typically excluded from sharing, here is a guide ([University of Oregon](#), 2011).

Other data definitions

Wikipedia on [data management](#)
Data glossary ([Oregon](#), 2011)



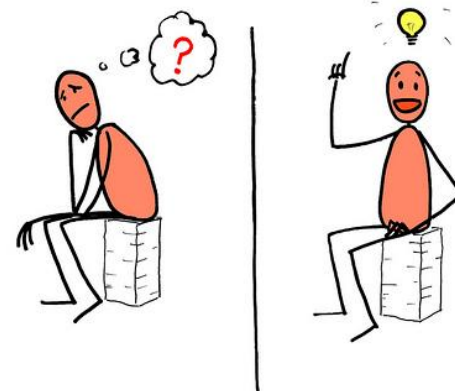
For a comprehensive overview of data management

These guides are quite comprehensive. Skim them now, and refer to them when you are writing your plan.

- Data management and publishing subject guide ([MIT Libraries](#), 2010)
- Managing and sharing data: A best practice guide for researchers, 2nd edition ([UK Data Archive](#), 2009)
- ANU data management manual: Managing digital research data at the Australian National University, 2nd edition ([ANU](#), 2010)
- Digital Curation 101 materials ([Digital Curation Centre](#), 2010)
- Guide to social science data preparation and archiving: Best practice throughout the data life cycle, 4th edition ([ICPSR](#), 2009)

5. Examine data management activities and find ideas, tools, and services for your plan

Here is an outline of data activities. Examining them will help you find tools and services for data management and also help you write your data plan.



Examining this data activity helps you address this NSF data plan topic (NSF , 2011)
Understand your data	Products of the research The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project
Organize and describe data	Data format 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
Determine the approach to sharing	Access to data and data sharing practices and policies Policies for access and sharing including provisions for the appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements
Find tools for sharing	Policies for re-use, re-distribution, and production of derivatives Policies and provisions for others to use your research products
Store data	Archiving of data Plans for archiving data, samples, and other research products, and for preservation of access to them

5.1. Understand your data



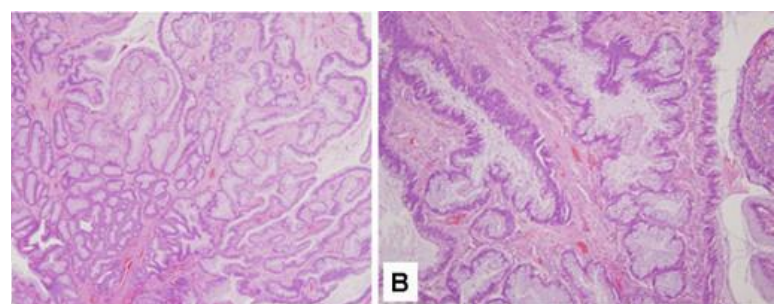
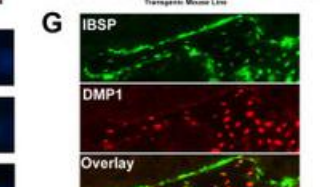
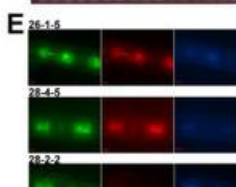
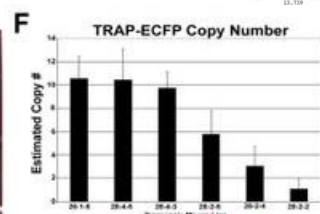
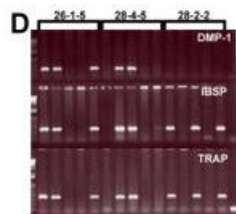
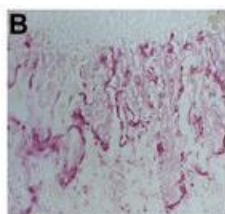
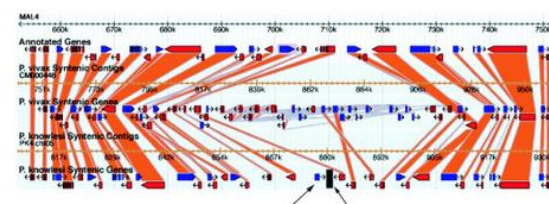
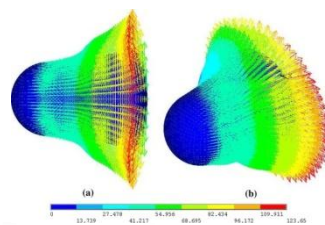
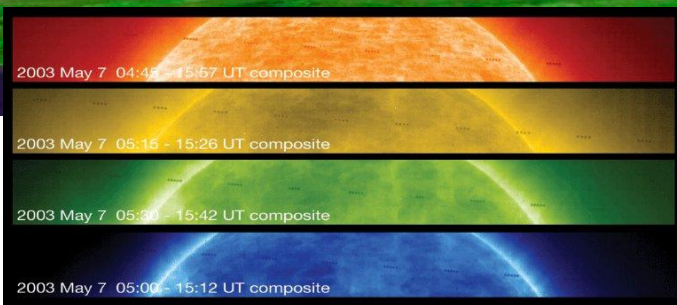
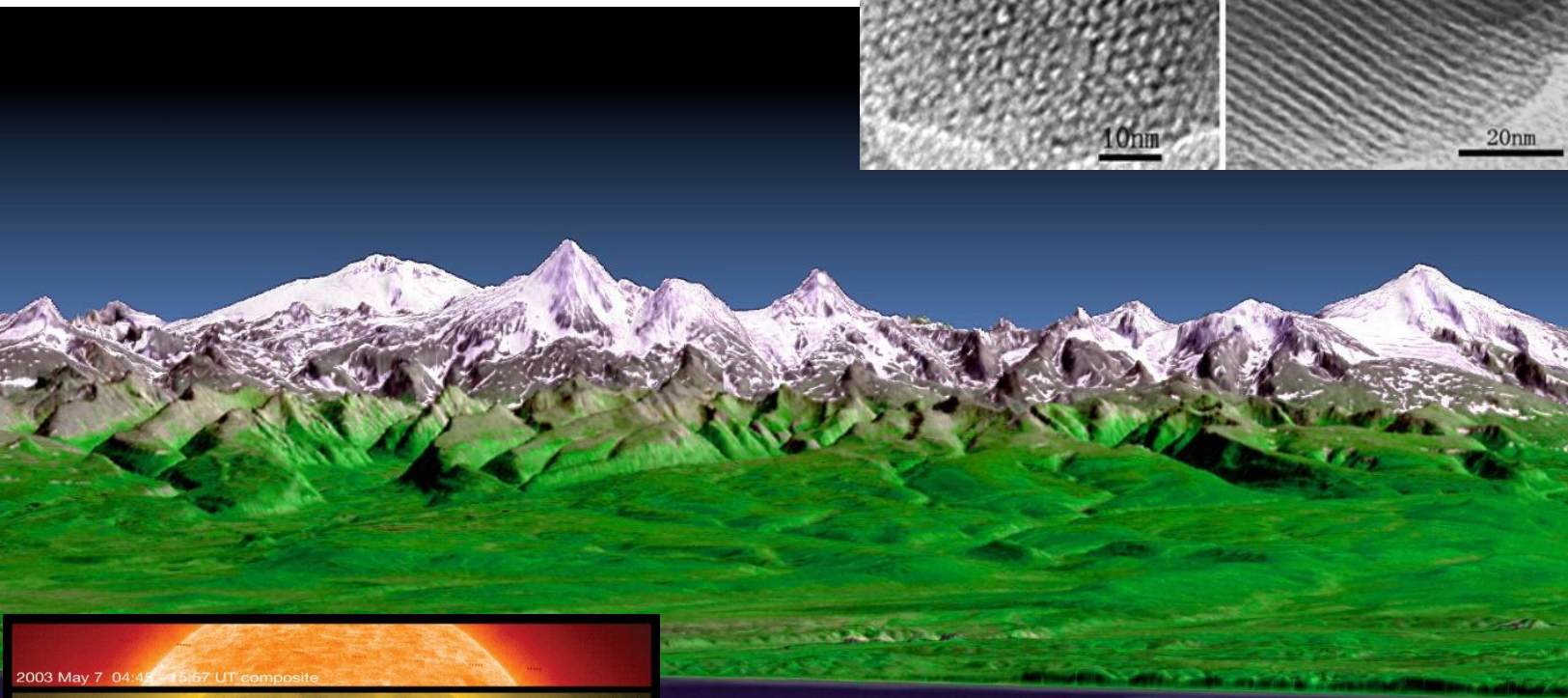
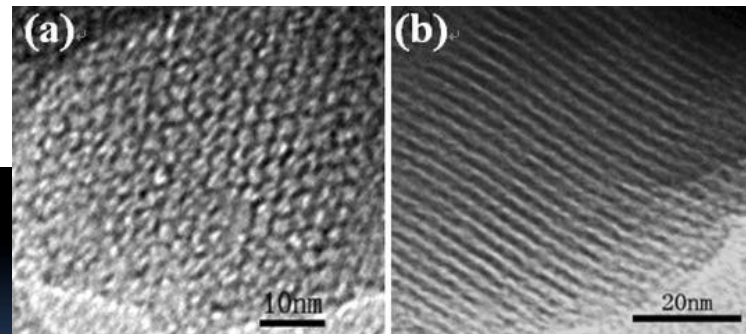
Questions to ask

1. What types of data will be produced for your project?
2. How much data will be produced?



Issue to consider

Review the [data life cycle](#), and think about your data needs and activities for each stage.



5.2. Organize and describe data



Questions to ask

1. How will you document your data and project?
2. What file formats will be produced for your project?
3. How will you organize your files into directories, and what naming conventions will you apply?
4. What project and data [identifiers](#) will be assigned?
5. How often will the data change or be updated, and will versions need to be tracked?



Issues to consider

Metadata are the description and documentation of your data to keep them organized and prepared for storage and sharing. For example, the date of creation, subject, author name, and more. This description will help researchers find, understand, use, and cite your data ([MIT Libraries](#), 2010).

Metadata can be used for documenting the data contents of files, for describing the technical characteristics of these files, or for expressing the relationships between files ([CDL](#), 2011).

Your NSF data management plan needs to address metadata.

- There are [different standards](#) for assigning metadata to your data files. Use the standard recommended by your funding opportunity specifications or by the associated NSF unit.
- If there are no existing standards or if you find them inadequate for your needs, explain this situation and propose solutions or remedies ([NSF](#), 2011).
- At the very least, write a text document that describes your data. Here are [elements to address](#). Save this document as a “readme.txt” file and store with your data. Alternatively, your data can reference a published article that has this description.



5.3. Determine the approach to sharing



Questions to ask

1. Who is responsible for managing and controlling the data?
2. Does project funding require your data to be shared or publicly accessible?
3. For what or for whom are the data intended?
4. How long must the data be retained?
5. When and where do you intend to publish or distribute your data?
6. Are there issues with privacy or intellectual property (for example, personal, high-security, or commercially sensitive data)?



Issues to consider

Different ways to share

When sharing your data and research products, follow the requirements of your NSF funding opportunity and the standards of your discipline.

If there are no existing requirements or standards for sharing, choose from different options ([Vision et al., 2011](#)).

1. **Share-upon-request**
Simple, but inefficient for wide distribution.
2. **Self-archives such as a lab website**
Customizable to your needs and interests, but long term availability is at the risk of website updates and institutional affiliation changes.
3. **Publish in a journal as supplementary online materials**
Convenient, but journals may have limits on file size and format. Additionally, expensive journal subscription costs may be a barrier for other researchers to access your data.
4. **Institutional or public data archives and repositories**
They provide reliable long-term preservation – especially when they are developed and maintained by federal agencies (for example, GenBank at NCBI, NIH). Additionally, these online storage sites may provide tools that help others find your data – for example, services for assigning metadata. Here's a tip: help others find your data by citing the data repository in your publications.

Least open



*Facilitating
discovery,
re-use, and
preservation*



Responsible sharing

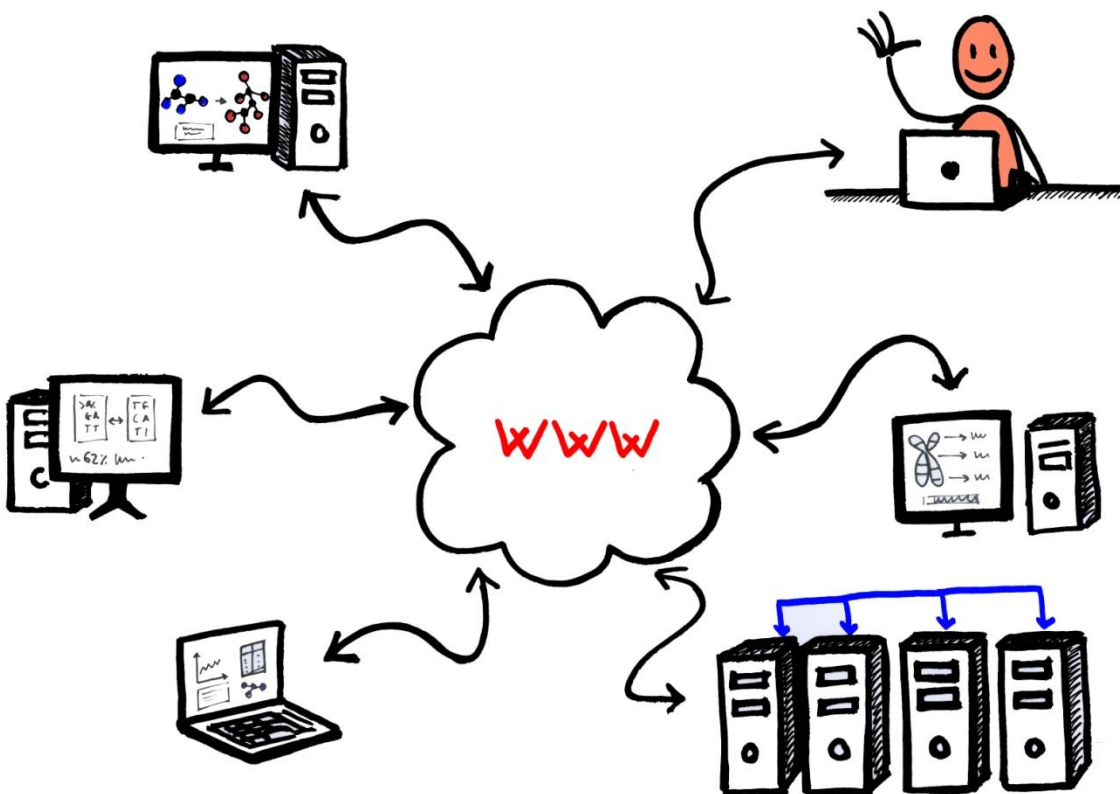
Before you share, check your ethical and legal rights and requirements.

For example, protect the privacy and confidentiality of research subjects, protect national security, and comply with regulations under HIPAA and Institutional Review Boards.

NSF recognizes the importance of intellectual property rights and data with potential commercial value.

You can propose a data management approach that may include reasonable binding agreements, and this proposal will undergo peer review and program management ([NSF, 2011](#)).

If you have the right to and the interest, you can share your data with unlimited use. Make this explicit with a [CC0 Declaration](#).



5.4. Find tools for sharing



Question to ask

1. Are there tools, software, or services needed for data creation, sharing, processing, analysis, and visualization?



Technologies to consider

Email	Attach your data when you reply to requestors
Website	Upload your data to a lab website or a web storage service via Google, Amazon, or Microsoft
Journal publication	Find a journal that lets you publish your data as supplementary materials
Repository or archive	Find one at the University of California, or search for a subject-specific public repository – browse this directory too

↓
*Improving
preservation*

Tools and services for sharing at the University of California

[UC3](#), the University of California Curation Center, provides tools and services for managing, curating, and preserving the information resources from your scholarly activities.

[Merritt](#) is a repository service for the University of California community. It lets you manage, archive, and share digital content. It also features an easy-to-use interface for deposits and updates, offers persistent URLs for access, and provides tools for long-term management and permanent storage.

Check out other [UC3 data management tools and services](#).

UC Berkeley's IST [Data Services](#) can manage data repositories and provide data presentation, visualization, analysis, and collaboration tools. Here are their services for [data centers and servers](#).



5.5. Store data



Questions to ask

1. Do you have a strategy for storing and backing up data?
2. Are long-lived [file formats](#) used?
3. How secure are the data?
4. How long should the data be [retained](#)?
(3 to 5 years, 10 to 20 years, permanently)



Issues to consider

Choose a long-lived file format

When saving data, use a file format recommended for long-term preservation.

- File formats and naming ([University of Oregon Libraries](#), 2011)
- Data formats recommended – see page 9 ([UK Data Archive](#), 2009)

Store the data you generate ([University of Minnesota Libraries](#), 2010)

- There are **local storage options** like your computer hard drive, external drives, or your departmental server or network.
- **Campus-based storage options** may provide greater capacity or more services like backup.
- **Cloud-based storage** stores data on remote servers and may reduce the effort for management, access, and collaboration. For example, [Amazon S3](#), [Jungle Disk](#), [Dropbox](#), and [Google Docs](#). However, be cautious with private data on any third party server.

Backup and secure your data

Have a backup strategy or use backup tools. And if you have sensitive information, here is a guide for securing your data ([California Digital Library](#), 2011).

Locally, UC Berkeley IST offers [storage and backup](#) services.



Create a long-term identifier for your data files

Think of this as a social security number for your files. A long-term identifier remains the same even if your data file moves to a different online space – and the identifier will direct researchers to the new location. At the University of California, there is the [EZID](#) service.

6. Find inspiration from sample plans

Sample NSF data management plans are available from [UC San Diego](#).

Data plans written for different research programs and funding agencies may provide inspiration.

- [ICPSR](#) examples
- [NIH](#) examples
- [CalTech](#) example



7. Write your plan with a template

NSF-specific templates

- Templates with sample language from the [Odum Institute](#) and [UNC University Libraries](#)
- A checklist for compliance prepared by the [Odum Institute](#)
- A template with guiding questions from the [UVa Library](#) and [Data Conservancy](#)

General templates for data management planning

- A template with sample language from [ICPSR](#)
- Checklists from [MIT Libraries](#), the [California Digital Library](#), and [DCC](#)

After writing your plan, check that your proposal has addressed the data plan requirements of your funding opportunity, your NSF unit, and the general NSF policy (see page 3).



NEED ASSISTANCE?



Find answers in the
NSF Data Management Plan [FAQ](#)



To consult a UC Berkeley librarian, please contact
data-consult@lists.berkeley.edu

Images used and adapted with permission
listed by page number

2: [Chris Khamken](#)

5: [ibuch](#)

6: [ibuch](#)

7: (from top to bottom and left to right)

[Shuxi Dai et al.](#), [NASA](#), [NASA/ESA](#), [R.H. Mao et al.](#), [Stacey Lapp et al.](#), [Peter Maye et al.](#),
and [Guoping Sui et al.](#)

10: [ibuch](#)

11: [Lawrence Berkeley National Laboratory](#)

All other images from
Microsoft Office 2010 clip art

Credit to [MIT Libraries](#) and [UC3](#)
for permission to use and adapt their pages

February 2011

Please [send us any comments](#) about this guide

