

Research Data Management

1. Research Data Management Plan

1.1 Start

Research Data Management Basics

To begin, click the START button.

1.2 Introduction

Welcome to the *Research Data Management Basics* Tutorial!

This tutorial is not specific to any particular funder, discipline, or type of data so we are going to keep it simple.

You can listen to or view the contents of this tutorial on the left menu.

To navigate forward click the NEXT button located on the bottom right hand corner of this screen.

1.3 Learning Outcomes

Upon completion of this tutorial you will be able to:

- Describe the importance of a data management plan
- Recall best practices for managing research data

- **Recognize the importance of project data documentation**
- **Establish good file name standards and version control**
- **Prevent data loss and plan for digital preservation**
- **Know when to use data cleaning and maintenance tools to refine your data**
- **Locate appropriate digital repositories and understand their role in publishing and archiving**
- **Identify where to receive further assistance with your research data management needs**

1.4 Creating a Plan

Creating a data management plan is not just about fulfilling funding agency requirements. It's about making your work reusable and verifiable.

Collecting research data usually represents an immense investment of time and energy. With some commonsense strategies in place the data can be preserved, reused and accurately represent your collected work.

A Research Data Management Plan, or RDMP, is not just about storing and preserving data. It is also about making decisions that enable future researchers to have consistent access to your shared work.

“Good data is good research... and getting better at data is a part of getting better at research itself.”

1.5 Data Sharing Plan

Research Data Management planning is simply about improving access, discovery and reuse of your research.

The foundation of the plan describes the types of data that will be captured or created, and how that data will be shared. It indicates what parts of your data need to be accessible and who will have access. There are a variety of well established practices, tools, and disciplinary repositories to help you organize and manage your data from creation to publication.

Identifying which solution is right for you may seem daunting, but the ASU Library can help you write a data management plan and identify resources that will best serve your project.

1.6 Establish Guidelines and Practices

In the same way that a car requires regular maintenance to reliably function over a long period of time, investing a little bit of time in best practices and routine data maintenance is necessary throughout the research data life cycle. This will ensure your data is backed up, organized, and ready for sharing at the end of the project.

1.7 Keep Things Organized

Consider how your data are collected and determine a method for keeping things organized.

Although your research datasets may make sense to you now, months or years in the future, you or your research team may forget the details of how you organized your datasets.

Your data documentation should note your organizational methods. This could be based on the method provided by a funding agency or publisher. Prospective Principal Investigators (PIs) should review the specific proposal request documents and funder requirements.

1.8 Documentation

Documentation is one of the most important management steps towards ensuring the ability to reuse your data.

Provide readme files, [data dictionary](#), or other contextual information such as an appendix of support material to clarify your research project for future users.

In addition to organization, explain your file naming conventions, metadata schemas and coding sheets.

Contextual documentation should include a brief description of each type of data to be generated. For example, experimental, qualitative, raw, processed are various types of data that you might consider in your documentation. You should also know which data you will share and at what stage (raw, processed, reduced, or analyzed).

1.9 File Naming

File names should be informative.

People tend to take file naming for granted. However, if researchers have not followed a structured naming convention they may run into problems identifying a file's contents without first opening it.

Use a consistent file naming structure at the beginning of your project, document it, and ensure all of the project members understand it. This will help all of your team quickly identify files and their contents. Some suggestions include inserting the date, type, specific sample or other

unique identifying aspect of the file's contents to help identify what the file contains without opening.

Be consistent and establish a plan at the beginning!

1.10 Common Filename Mistakes

Avoid some common file naming mistakes by keeping the filenames descriptive but not too long, preferably under 25 characters.

Use unique file names. Files with the same exact names can confuse data collecting software and cause accidental overwriting, misinterpretations or deletion.

As mentioned before be consistent and establish a plan before data collecting begins. For the best interoperability, use underscores and not spaces, clarify the date format and avoid using special characters in your documentation.

If you are outsourcing your research collecting, work with those partners to ensure the files you receive conform to your project's naming conventions so that they are easily organized.

1.11 Version Control

Inevitably you may have different versions of a dataset, or file that require some sort of differentiation between them.

Using a simple numbering system or date-naming the files are the most effective methods of version control.

Avoid ambiguous terms such as “rough draft”, “final draft”, “final final” “submitted” or other commonly used terms. These generally only end up confusing end-users or other members of your project team.

1.12 3-2-1 Storage Rule

Now let's look at the basics for storage frameworks.

With digital data of any kind one should at the minimum follow the 3-2-1 rule:

- **3 copies of your data**
- **2 different locations**
- **more than 1 storage medium**

This helps prevent losses due to natural disasters, storage media failures or human errors.

1.13 Security and Storage

Don't put all your data in one basket and rely on just your computer's hard drive or an external drive.

Devices like SD cards, optical media and other portable media are also highly vulnerable to loss, damage, and obsolescence.

Plan to include off site, cloud-based options in your management plan. ASU provides cloud service options such as Dropbox and Google Drive for

free to ASU affiliated research projects.

Consult ASU Research Support Services at research.asu.edu or the ASU Library for more information to get your 3-2-1 plan started.

You may also consider investigating other additional commercial options for added security and storage and include these costs in your project proposals.

1.14 Correct, Clean and Check

Correct, clean and check for errors in your dataset.

In most cases you may have to clean up your data, to ensure it is accurately represented in the software you use to visualize or organize the data.

For example: if working with spreadsheets, you could use software such as OpenRefine to clean up messy datasets or transform the data from one format to another.

1.15 Redact Sensitive Information

You may have an ethical or legal obligation to maintain the confidentiality of information within your datasets. Data classification is of high importance and most data repositories will not accept data with sensitive information such as social security numbers, birth dates or other information that would conflict with privacy and confidentiality of research subjects.

In addition to redacting certain information you may also have other circumstances requiring secure data storage or restricted access to data, such as licensing restrictions that prohibit data sharing or if the data is part of a research project with commercial potential.

1.16 Choose What You Share

Avoid a data deluge. When you are ready to share your data, you probably don't need to share everything you have collected.

Share the data you actually used to produce your final results such as the data behind the figures and charts in your publications. Avoid unnecessary derivatives and focus on what your conclusions are actually based on. Choosing what you share may also greatly reduce the storage needs of your project and reduce backend costs.

1.17 Depositing and Sharing

A good place to store and share your data is within a research data repository. These include domain or topic-specific repositories, as well as institutionally supported options. All of these solutions allow you to meet funder obligations, connect your publications to your data, and facilitate reuse and increased impact. They also allow anyone to confidently cite your data sets.

Re3data.org is a directory of more than 1,500 research data repositories. It is a good place to start your search for the repository that best fits your needs.

Repositories may have fees associated with storing your files. ASU has in-house options such as the **ASU Digital Repository for interdisciplinary research and the **Digital Archaeological Record (tDAR)**, an international**

archaeological digital archive and repository by the **Center for Digital Antiquity**.

There's a lot of resources here so know that you are not on your own and help is available from the ASU Library.

1.18 Writing Your Plan

Putting it all together

Writing your data management plan is not difficult once you think through these steps and align them with your funding or publisher requirements, which are typically spelled out within the funding solicitation. This should really be the easiest component of your proposal writing. Typically a data management plan is less than two pages of documentation and there are many examples of previously written plans you can use as the foundation for your own.

The ASU Library provides access to resources and services designed to help you write your data management plan (or DMP). For starters, you can use the DMPTool to do it yourself or work with us on a data management checklist to get started developing your plan.

To do it yourself, go to dmptool.org, and when prompted, select ASU as your institution.

1.19 Where Can I Get Help?

If you want to know how to best manage your research data, or have other questions related to your project, please use the ASU Library Research and Publication Services page at lib.asu.edu/research.

The Research and Publication Services Team can assist you in ensuring

your research is accessible and reusable.

1.20 Learning Outcomes

Now that you have completed this tutorial, you can:

- **Describe the importance of a data management plan**
- **Recall best practices for managing research data**
- **Recognize the importance of project data documentation**
- **Establish good file name standards and version control**
- **Prevent data loss and plan for digital preservation**
- **Know when to use data cleaning and maintenance tools to refine your data**
- **Locate appropriate digital repositories and understand their role in publishing and archiving**
- **Identify where to receive further assistance with your research data management needs**

1.21 Congratulations!

Congratulations! you've completed the *Research Data Management Basics* tutorial!