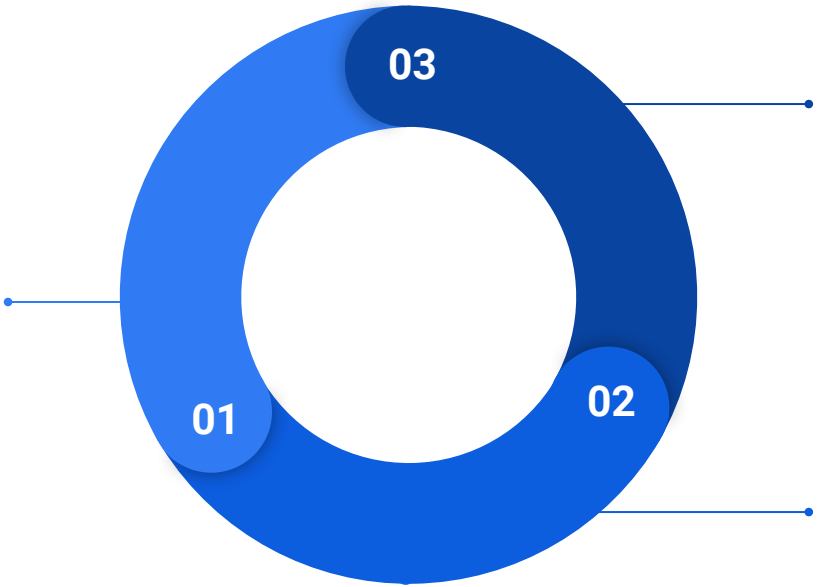
# Keys to a Successful Data Management Plan

## Creating a DMP & Using DMPTool for NSF-Funded Projects

Clare Michaud and Morgan Witte

**35 MINUTES**

- ➔ What is a data management plan (DMP)?
- ➔ Why are DMPs important?
- ➔ What components make a DMP effective for NSF grant applications?




**10 MINUTES**

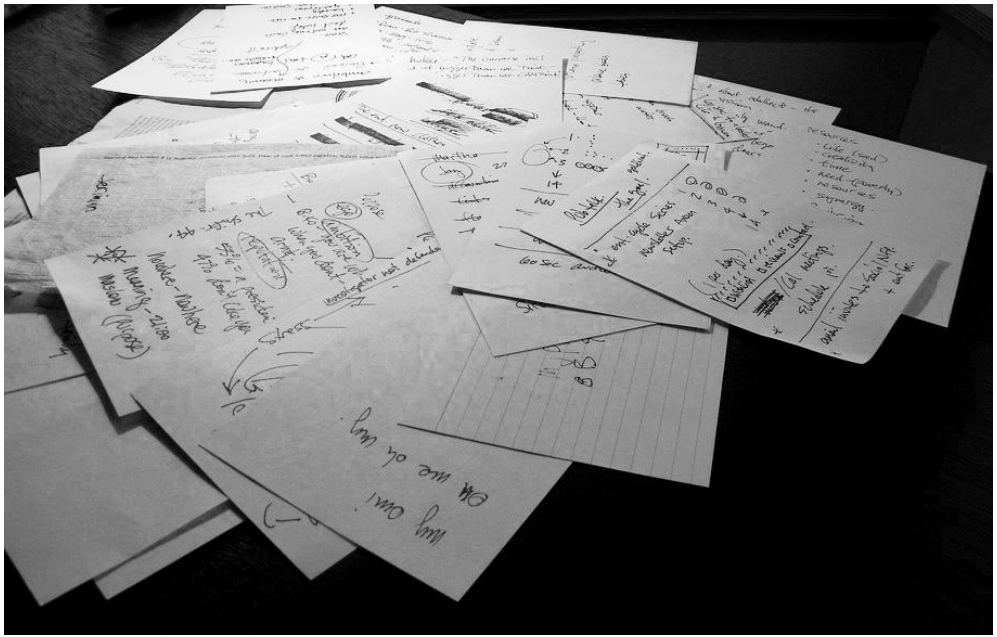
- ➔ Q&A
- ➔ Feedback questionnaire

**15 MINUTES**

- ➔ What is DMPTool?
- ➔ How can I use the features of DMPTool to write a DMP?

# What is a Data Management Plan (DMP)?

- A brief document describing what you will do with your data during your research project and after the project is over



# Why Create a DMP?

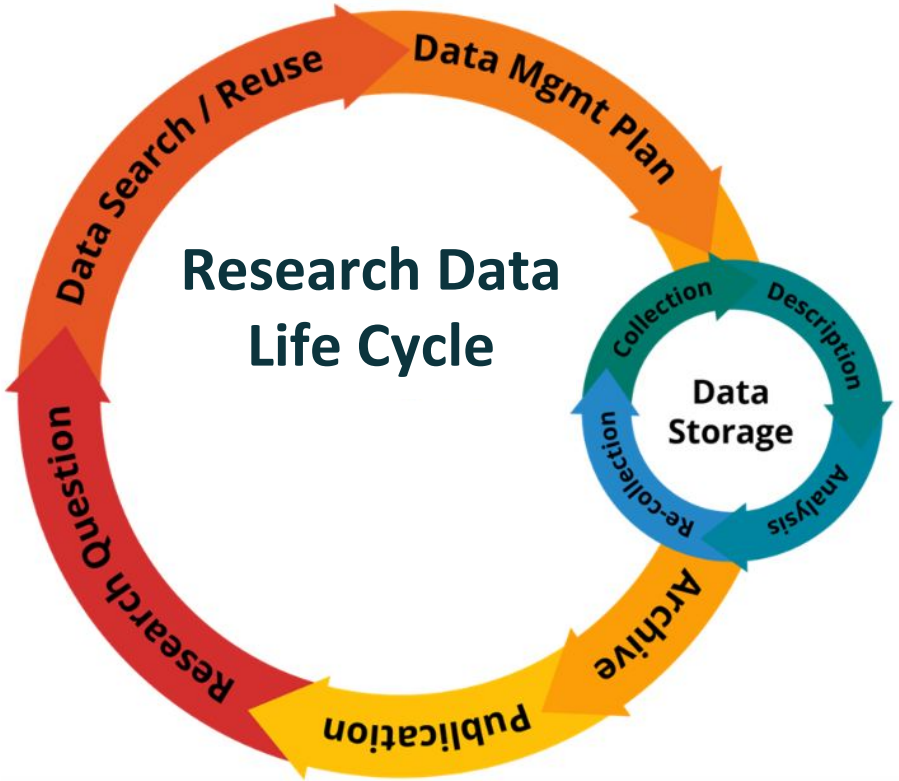
- It is required to submit one when applying for any NSF grant funding
- It is useful to have whenever creating and managing data
- It can make your research workflows easier
- It can help others interpret and use your data if you share it

# Components of a DMP for NSF Grants

1. Data types
2. Data documentation: formats and standards
3. Data sharing and access
4. Data archiving and preservation
5. Data management roles and responsibilities

**NSF Research Directorates**

- Biological Sciences
- Computer and Information Science and Engineering
- Education and Human Resources
- Engineering
- Geosciences
- Mathematical and Physical Sciences
- Social, Behavioral, and Economic Sciences



# Data Types: Identifying Your Research Data

The data from your research project will vary based on **form**, **type**, and **stage** in the research process

### Forms

- Measurements generated by sensors or lab instruments
- Computer modeling
- Simulations
- Observations/field studies
- Specimens
- Survey responses
- Demographics

### Types

- Observational
- Experimental
- Simulation (test models)
- Compiled/derived
- Reference/canonical

### Stages of Research Data

- Raw Data
- Processed Data
- Finalized Data
- Reuse or Combine with Existing Data

# Storing and Backing Up Your Data

- **Storage**: keeping working files in a secure location that can be accessed readily
- **Backup**: keeping additional copies of files in physical or cloud locations, separate from the working files that are in storage
  - Copies to access in case of data loss
  - Create a schedule for either manual or automatic backup
- This is an important section for specifying a role: who will be responsible for backing up data?

# Storing and Backing Up Your Data: Best Practices

Think of back-up in the rule of three:

1. **One copy in active storage.** This is a copy you are regularly accessing and working on during your research. It will likely be on your computer or a lab's shared network drive.
2. **A second copy on a different device on- or off-site**, such as an external hard drive in your office or a backup server provided by your IT department.
3. **A third copy, preferably off-site.** This might be on a cloud application like Box, Google Drive, or another appropriate cloud solution.

# Storing and Backing Up Your Data: Data Security

- How will you store your data?
  - **Examples**: in a secure, locked room, or managed technologically on secure, university-affiliated machines
  - This is another opportunity to designate a role
- Policies that may affect how you store your data:
  - HIPAA (Health Information Portability and Accountability Act)
  - FERPA (Federal Educational Rights and Privacy Act)

# Organizing Your Research Data

- Good methods for file-naming and organizing are some of the easiest ways to manage research data
- Funders typically look for brief information about how you plan to organize your research data
- Naming files consistently and accurately simplifies searching for files, distinguishing between versions, working with files, and sharing files
- Is there a role for this in your research group?

# Organizing Your Research Data: File Naming

- **Be brief**: choose 3-4 key pieces of information about the file to use in a file name

**Mendota_Buoy6_20180722_v3**

1. The lake and the buoy that the data was collected from
2. The date that the data was gathered (written in a standard format)
3. The version number of the document

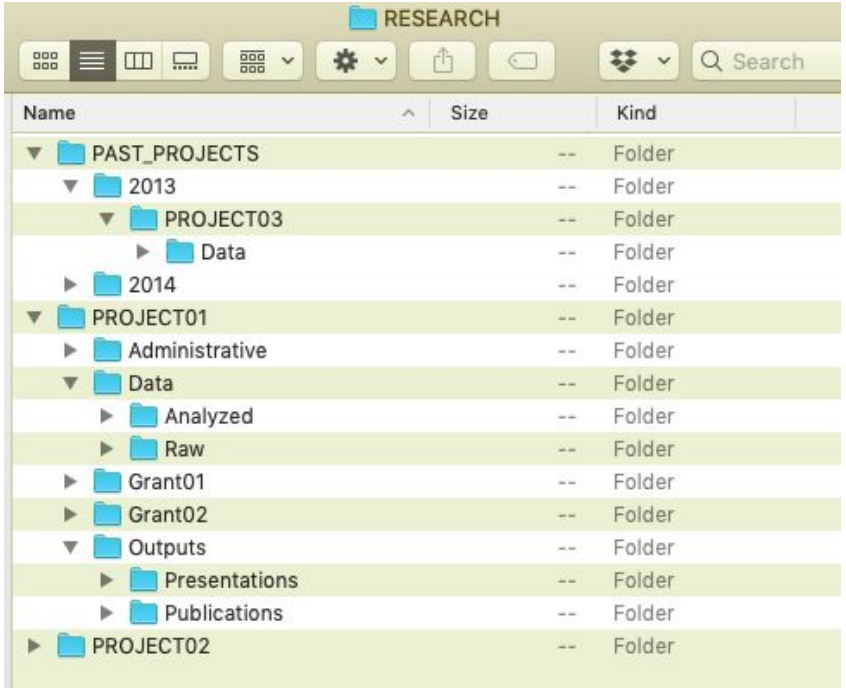
# Organizing Your Research Data: File Naming

**Other tips**

- File names that are too long may not work well with certain software
- Avoid special characters: ! @ # $ % ^ & * ( ) ` | { } [ ] < > / ? “ ‘
- Don’t use spaces
  - Instead, use:
    - Underscores: file_name.xxx
    - Dashes: file-name.xxx
    - No separation: filename.xxx
    - Camel case: FileName.xxx

# Organizing Your Research Data: File Organization

- A well-organized hierarchical folder structure should align with your file-naming conventions
- Balancing *breadth* and *depth* in creating a hierarchy
  - Limit the number of top-level folders and the number of nested folders
  - Too many nested folders → data becomes difficult to access
  - Too many files in a folder → data becomes cluttered and difficult to find



# Documenting Your Research Data

- Documentation for your data should contain the minimum information required to be able to reuse the data that it describes

  - Examples of methods for documenting data:
    - Data dictionary
    - README file
    - Embedded metadata (descriptive information about data)
    - Data paper
    - Codebook

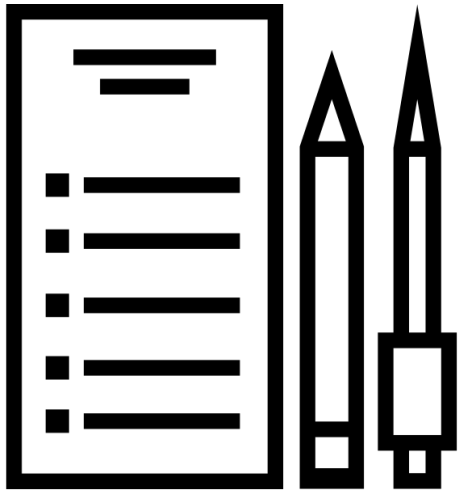
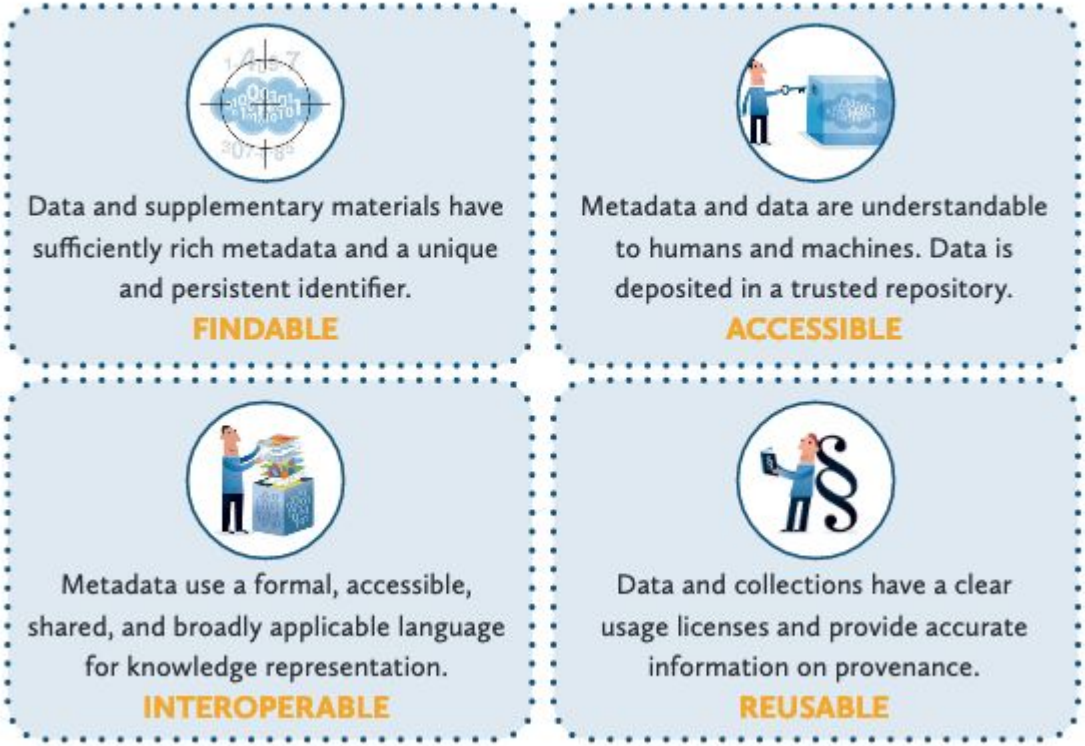


*Image attribution: 'documents' by Susannanova from the Noun Project*

# Why Should You Document Your Research Data?

- FAIR data practices and NSF's Growing Convergence Research



# Documenting Your Research Data:
## Data Dictionaries & Codebooks

- Provide key information about the data that you will be collecting
- Used to explain variable names and values in a dataset
    - Data types, what each variable describes, etc.
- Most commonly used when working with tabular data or creating a database
- Recommended resource: How to Make a Data Dictionary (OSF)

# Documenting Your Research Data: README Files

- Used to describe software packages, programming scripts, and datasets, and can also be used for research projects
- Key information to include:
  - Information about the creators of the data (researchers)
  - Basic funding details
  - A list of the files included in the dataset
  - Citation for the data
  - Citations for the research outputs (articles, presentations, etc.)

# Documenting Your Research Data: Data Papers

- Data papers ≠ Research papers
- Used to present large or expansive datasets
- Includes metadata about the content, context, quality, and structure of the data
- *Scientific Data* is an example of a publisher of data papers

# Documenting Your Research Data:
## Embedded Metadata

- An additional file with metadata can contain information that describes your project, including its data and other outputs and technical specifications
- Follow a disciplinary schema or a schema that fits the type of data you're collecting
- Examples for geospatial and data from surveys or observational methods
  - ISO 19115
  - Data Documentation Initiative


# Sharing Your Research Data

**Why?**

- Fulfill funder requirements
- Raise interest in publications
- To accelerate research and discovery rates

**When?**

- Within 12 months of article publication, or the end of a research project

**How?**

- Publish it through a trusted repository

# Sharing Your Research Data: NSF Policies

- Expected to share primary data, samples, physical collections, other supporting materials during the course of the project
  - At no more than incremental cost
  - Within a reasonable time
- Requirement for NSF-funded PIs to use NSF-PAR (Public Access Repository)
  - Published, peer-reviewed journal articles
  - Juried conference papers

# When Research Data Cannot Be Shared

- It contains potentially identifying information of human subjects
- It contains the locations of endangered/threatened species and will only be shared with trusted parties who agree to reuse criteria
- It cannot be released until the patents related to this research are issued

# Archiving and Preserving Your Data

- Each research directorate has different expectations about data preservation and sharing
- Identify a repository in your field
- Information to include in a DMP:
  - Access (or restrictions) to preserved data
  - Timeline for how long it should be preserved
  - Who in your research group is responsible for ensuring the data is preserved

# Roles and Responsibilities

- Required to include in DMPs for most NSF research directorates
- Areas that benefit most from role-and-responsibility assignments:
  - Data storage and backup
  - Data organization
  - Data archiving and preservation
- Recommended to include a contingency plan if project personnel leave

# What is DMPTool?


A free online tool that supports drafting of funder-compliant, functional DMPs by:

- Guiding you through the process of creating a DMP
- Helping you comply with funder and institutional requirements
- Providing links to funder information, guidance, and sample DMPs
- Making collaboration seamless through FREE access for anyone

UW-Madison is a participating institution.

**DMPTool DOES NOT submit your DMP to funders for you!

# Today We Will:

DMPTool
Build your Data Management Plan

- Log in with your NetID
- Explore DMPTool's resources
- Create a test (or real!) data management plan
  - View guidance provided by funders, institutions, and DMPTool
  - Learn how to share plans with collaborators
  - Request feedback on a completed plan
    - Only if you don't already have an RDS consultant as a collaborator!
  - Export our test plan into a useable format
    - Remember: DMPTool DOES NOT submit your DMP to funders for you!

DMPTool

Build your Data Management Plan

https://dmptool.org/

# Feedback on DMPTool


If you already have an RDS consultant added as a collaborator on your DMP:

- DO NOT request feedback!
- Your RDS consultant will lose their collaborator status

If you still need to request an RDS consultant to review your DMP

- DO request feedback!
- This will ensure you get assigned to an RDS consultant to start reviewing your DMP as quickly as possible

RDS Consultants can't be feedback providers AND collaborators. If you do both, your consultant will have restricted access!

# Questions?

GitHub with slides, handout, and resources

https://github.com/uw-madison-data-management/dmp-nsf

DMPTool help and FAQs

https://dmptool.org/help

https://dmptool.org/faq

Research Data Services (RDS) at UW-Madison

http://researchdata.wisc.edu/how-to-create-a-dmp/

http://researchdata.wisc.edu/contact-us/

# Feedback for Us!

http://bit.ly/2Fk7euz