

Data Management Plans

Spring 2024
Tom Boyle



Agenda

- Data Management Plans
 - For NSF
 - For NIH
- DMPTool
- KSU Resources

What is a DMP or DMS?

- A data management plan is a supplementary document that outlines what you plan to do with your data during your research project and beyond.
- Most federally funded grants and many private foundations require some sort of plan in a proposal.
- In the US, data management plans are usually limited to 2-page documents.
- In 2023, the NIH policy ([Data Management and Sharing](#)) reflects their emphasis on the sharing of scientific data. The push is to clearly demonstrate additional value in their awards from the resulting sharing.

NSF: Data Management Plan

- Proposal & Award Policies & Procedures Guide (PAPPG)
 - [PAPPG \(NSF 23-1\) dated January 30, 2023](#)
 - [PAPPG \(NSF 24-1\)](#) (mandatory for proposals submitted on or after 5/20/2024)
- Chapter 2: Preparation
 - Section D: Proposal Content
 - Special Information and Supplementary Documentation
 - [Plans for Data Management and Sharing of the Products of Research](#)

NSF Recommended Sections

Types of data produced

Data and metadata standards

Policies for re-use, re-distribution and derivatives

Policies for access and sharing

Plans for archiving and preservation

The next few slides will give the NSF explanation for each section and an example from a sample proposal

NSF: Types of Data Produced

NSF Generic

The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project.

KSU Example

This project includes storage infrastructure to host research datasets. The researchers will be asked to develop and share materials/slides, source code, and their analysis along with datasets. Data will be produced based on research studies, experiments, and computational simulations...

NSF: Data and Metadata Standards

NSF Generic

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).

NSF Specific

No data will be stored in connection with the subject name or any other Personally Identifiable Information (PII).

The plots, images, and videos will be saved as the jpeg, mpeg, pdf, docx, ppt, or xls formats, as appropriate. Large data can be compressed in the formats such as *.tar, gzip, zip.

Current required metadata includes Title, Authors, Publication Date (Year is required; season, month, and day are optional), and File.

In addition, the following metadata is available: Department, Keywords, Disciplines (links the dataset to bepress networking 'commons' and makes work more findable to those interested in the field), Abstract, Journal Information (Title, Volume, Issue, ISSN), DOI, Comments, and Supplemental Files. However, additional metadata can be added to increase findability and reusability.

NSF: Policies for Access and Sharing

NSF Generic

Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

KSU Example

The data to be shared will be hosted on the KSU storage server and read or write access will be provided to relevant parties based on the dataset shared and collaborators of researchers. The permissions for accessing data will be determined by research project groups and the owners of datasets. Once the relevant publications are out, we will encourage dataset owners to make their datasets public.

NSF: Policies for re-use, re-distribution, derivatives

NSF Generic

Policies and provisions for re-use, re-distribution, and the production of derivatives.

KSU Example

All software and materials developed for the proposed research will be freely distributed using the training platform and training materials.

The PIs will maintain a WiKi at Kennesaw State University, on which all project related activities will be posted, including relevant literature, software, data, and discussions.

Any data products produced by these analyses will be recorded and are considered metadata for this project.

NSF: Plans for archiving and preservation

NSF Generic

Plans for archiving data, samples, and other research products, and for preservation of access to them.

KSU Example

KSU's UITS is also providing support for archiving research datasets on Azure Files cloud for archival purposes. All archived data files will be indexed by both the name of the experiment and the date of acquisition in order to aid in retrieval.

NIH

NIH template

- DSP Policy
 - [Data Management and Sharing Policy | Data Sharing \(nih.gov\)](#)
- NIH offers a fillable form
 - <https://grants.nih.gov/sites/default/files/DMS-Plan-blank-format-page.docx>
- The following slides share the sections of the template with excerpts from a sample DSP.

NIH: Data Type

Element 1:

- A. Types and amount of data
- B. Scientific data preserved
- C. Metadata

KSU Example

- Tier 1 data refers to the identified versions of the data. Tier 2 data refers to the deidentified versions of the data. Tier 1 can include individual names, addresses, criminal activity details, and health details. Tier 2 removes the names and addresses but preserves other details.
- Response reports and associated analyses will be preserved during the duration of the signed MOU with the research client organizations plus six months following expiration of the MOU.
- Information about the workflow that connects outputs to the inputs will be stored, indicating the stages of the workflow with tags on the inputs and outputs of each stage.

NIH: Related Tools, Software and/or Code

Element 2:

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data. Name them and describe their access.

KSU Example

We are building custom analysis software designed to identify potential behavioral health cases within the data using the Python programming language and related programming libraries. We expect to package the software for easier usage at some point using the de-identified source data. Additionally, we have been using Prodigy coding software for truth sample development, which produces one of the associated analyses of the cases.

NIH: Standards

Element 3

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project.

KSU Example

No consensus standards exist for this type of data at present. We are aware of NIBRS, IAED, and related data standards used in this type of crisis reporting [reference]. However, for detection of such health crisis dynamics without a clinical diagnosis determination, there is not a clear standard to follow.

NIH: Data Preservation, Access and Associated Timelines

Element 4

- A. Repository where scientific data and metadata will be archived.
- B. How scientific data will be findable and identifiable.
- C. When and how long the scientific data will be made available.

KSU Example

- No plan to archive data beyond the term of the MOU plus 6 months.
- No Tier 1 data will be findable due to the sensitive nature of the data. Tier 2 examples may be shared and findable via presentations and publications.
- The Tier 1 and Tier 2 source data will not be available. The Tier 1 source data will not be available

NIH: Access, Distribution, or Reuse Considerations

Element 5

A. Factors affecting subsequent access, distribution, or reuse of scientific data

B. Whether access to scientific data will be controlled

C. Protections for privacy, rights, and confidentiality of human research participants

KSU Example

- Tier 1 data will not be shared. Tier 2 data will be referenced as de-identified examples.
- Tier 1 data will only be available to the research team on high-encryption hard drives accessed without internet connectivity.
- Tier 2 data will be available on KSU devices and resources following KSU security policies.

NIH: Oversight of Data Management and Sharing

Element 6

Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles)

KSU Example

In addition to Primary Investigator monitoring and management, the Tier 1 data in this project will be monitored and remotely managed via the IronKey tool dashboard managed daily by the HIPAA compliance research data security officers within the KSU University Information Technology Services office.

While less than the Tier 1 data, risks associated with Tier 2 data access, handling and storage will be the responsibility of the Primary Investigators of the project.

What is DMP Tool?

- A website: <https://dmptool.org/>
- An online tool to create your own DMP
- An archive of example DMPs.
- View Funder requirements.
 - Example: NSF
 - [DMP Templates \(dmptool.org\)](https://dmptool.org/)
 - See templates for different disciplines
 - Typically, there are links to the PAPPG and FAQs

DMPTool: Login

- There is no KSU SSO option, create your own password.

DMP Tool Public Plans Funder Requirements About [Login](#) [Sign Up](#)

Login

Email address *

For SSO, use your institutional address.

[Continue](#)

Problems signing in? [Contact us.](#)

DMP Tool Quick links External links



DMP Tool: Create New Plan

Once you have selected a primary funding organization, you will be presented with the available templates.



[Dashboard](#) [Create Plan](#) [Public Plans](#) [Funder Requirements](#) [About](#)   [Logout](#)

Kennesaw State University (kennesaw.edu)

Create a new plan

Before you get started, we need some information about your research project to set you up with the best DMP template for your needs.

* What research project are you planning?

If applying for funding, state the project title exactly as in the proposal.

mock project for testing, practice, or educational purposes

* Select the primary research organization

Research organization

- or -

No research organization associated with this plan or my research organization is not listed

* Select the primary funding organization

Funder

- or -

No funder associated with this plan or my funder is not listed

[Create plan](#)

[Cancel](#)



DMP Tool: NSF Templates

* Select the primary funding organization

Funder

National Science Foundation (nsf.gov)

- or -

No funder associated with this plan or my funder is not listed

Which DMP template would you like to use?

✓ Arctic Data Center: NSF Polar Programs

BCO-DMO NSF OCE: Biological and Chemical Oceanography

NSF-AGS: Atmospheric and Geospace Sciences

NSF-AST: Astronomical Sciences

NSF-BIO: Biological Sciences

NSF-CHE: Chemistry Division

NSF-CISE: Computer and Information Science and Engineering

NSF-DMR: Materials Research

NSF-DMS: Mathematical Sciences

NSF-EAR: Earth Sciences

NSF-EHR: Education and Human Resources

NSF-ENG: Engineering

NSF-GEN: Generic

NSF-PHY: Physics

NSF-SBE: Social, Behavioral, Economic Sciences

We found multiple DMP templates corresponding to your funder.

External links

[GitHub](#)

[Blog](#)

[Accessibility](#)

DMPTool: New Project

Project Title

Project Abstract

Research Domain

Project Start / End

Funder

Funding Status

Funding Opportunity Number


Grant number / URL

Project Details Collaborators Write Plan Research outputs Finalize Download

Project title *

mock project for testing, practice, or educational purposes

Project abstract

B *I* 12pt **A** 

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean blandit tortor a purus cursus, sit amet accumsan lorem aliquet. Etiam sodales dolor et enim euismod tincidunt sagittis vitae libero. Nunc massa dolor, sollicitudin non eleifend id, pulvinar at turpis. Etiam vitae erat ipsum. Duis vestibulum convallis enim a luctus. Etiam pellentesque magna nisl, ut rhoncus justo tempor vitae. Sed pulvinar, mi nec blandit facilisis, mauris quam feugiat magna, nec rhoncus justo diam sed orci. Vestibulum nec rhoncus mi. Mauris ut bibendum erat. Morbi at eros congue, scelerisque orci ut, pulvinar risus.

Press Alt 0 or Option 0 for help using the rich text editor with keyboard only.

Research domain

Project Start Project End



DMPTool: New Project Collaborators

DMP Collaborators

Invite Collaborators

Permissions

- Co-owner
- Editor
- Read only

Project Details Collaborators Write Plan Research outputs Finalize Download

Project Contributors

Please list the project's Principal Investigator(s) and those responsible for data management.

No contributors have been defined.

[Add a contributor](#)

DMP Collaborators

Invite specific people to read, edit, or administer your plan. Invitees will receive an email notification that they have access to this p

Email address	Permissions
tboyle@kennesaw.edu	Owner

Invite collaborators

Email *

Permissions *

Co-owner

Editor

Read only

[Submit](#)



DMPTool: New Project Research outputs

List your anticipated research output(s)

DMPTool: New Project Finalize

Set plan visibility for finished plans

- Private
- Organization
- Public

Register your plan for a DMP ID (DOI)

- Only tied to the creator ORCID
- Funder has to be identified

DMPTool: New Project Download

Formats available

- Csv
- Html
- pdf
- text
- Docx
- json

Project Details Collaborators Write Plan Research outputs Finalize Download

Format
pdf

Download settings

- include a project details coversheet
- include the section headings
- include the question text
- include any unanswered questions

PDF formatting

Font

Face "Times New Roman", Times, Serif Size (pt) 11

Margin (mm)

Top	Bottom
25	25

Download Plan (new window)



Resources from KSU

- Our page with supporting materials for Documents you may need in your proposals
 - [Proposal Documents - Office of Research \(kennesaw.edu\)](https://www.kennesaw.edu/research/proposal-documents)
 - This includes notes for creating a generic plan
 - Specific examples from KSU researchers for NIH and NSF proposals.
- UITS has a matrix for their available storage resources.
 - <https://uits.kennesaw.edu/document-management/>

What could be in your DMP?

- Description of the Data
- MetaData
- Access, Sharing and Privacy
- Strategy for Archiving and Preserving Data
- Plans for any Re-use and Re-distributing or Any Derivatives of the Data
- For each section, I have tried to provide a set of questions you can consider when creating your DMP.
- https://research.kennesaw.edu/computing/resources/dmp_notes.php

Questions?



Special Thanks

- Dr. Dominic Thomas for use of his DMP
- Dr. Ramazan Aygun for use of his DMP