

Data Management Plans

Wasila Dahdul, PhD

Data Curation Librarian, UCI Libraries

4/22/22

Overview

- Why plan for managing your data?
- Data Management Plan (DMP) basics
- DMP Deep Dive: NSF requirements and examples
- DMP resources

Why plan for managing your data?

Open access and data sharing requirements

The image shows two overlapping web pages. The top page is the National Science Foundation (NSF) website, featuring a search bar and navigation tabs for Research Areas, Funding, Awards, Document Library, News, and About NSF. The bottom page is the NIH Grants & Funding website, displaying the NIH Data Sharing Policy. A yellow highlight on the NIH page reads: "NEWS: New NIH Policy on Data Management and Sharing. On October 29, 2020, NIH issued the NIH Policy for Data Management and Sharing which will require NIH funded researchers to prospectively submit a plan outlining how scientific data from their research will be managed and shared. This policy will be effective January 25, 2023 and at that time will replace the 2003 NIH Data Sharing Plan. Learn more about the new policy." Below this, text states: "Data sharing is essential for expedited translation of research results into knowledge, products and procedures to improve human health. The Final NIH Statement on Sharing Research Data was published in the NIH Guide on February 26, 2003. This is an extension of NIH policy on sharing research resources, and reaffirms NIH support for the concept of data sharing. The new policy becomes effective with the October 1, 2003 receipt date for applications or proposals to NIH." A bullet point lists a "Data Sharing Regulations/Policy/Guidance Chart for NIH Awards (08/30/2006) - (MS Word - 58 KB)" as a quick guide.

See [2013 OSTP Memorandum](#) on increasing public access to the results of research funded by the federal government

NSF example: Dissemination and Sharing of Research Results

- Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data...created or gathered in the course of work under NSF grants
- Investigators and grantees are encouraged to share software and inventions created under the grant or otherwise make them or their products widely available and usable

NSF Proposal and Award Policies and Procedures Guide [Chapter XI.D.4](#)

Benefits of data management and sharing

- Maximize Data Collection
- Transparency and accountability for taxpayer funds
- Accelerate the Research Enterprise
 - Spark new research collaborations
 - Enable unique data combinations
 - Facilitate study validation
 - Stimulate new research inquiries



From NASEM and NIH [Workshop on Changing the Culture of Data Management and Sharing, April 28-29, 2021](#)

Data Management Plan Basics

What is a DMP?

- A short, typically 2 page document that describes how data will be collected and managed **during** your research, and how it will be preserved and shared **after** you complete your project
 - Required by some funding agencies
 - A guide for daily activities

Components of a typical DMP

1. Types of data
2. Standards for data and metadata
3. Policies for access and sharing
4. Policies for re-use, re-distribution, and the production of derivatives
5. Plans for archiving data
6. Roles and responsibilities

NSF PAPPG [Chapter II.C.2.j](#)

Find data sharing requirements by funder

- SPARC Open Data (Federal)
 - <http://datasharing.sparcopen.org/data>
- DMPTool (Federal or Private)
 - https://dmptool.org/public_templates



NSF data sharing requirements

Find specific requirements by unit, program, or directorate:

<https://www.nsf.gov/bfa/dias/policy/dmp.jsp>

- Biological Sciences Directorate (BIO)
 - [Directorate-wide Guidance](#)
- Computer & Information Sciences & Engineering (CISE)
 - [Directorate-wide Guidance](#)
- Education & Human Resources Directorate (EHR)
 - [Directorate-wide Guidance](#)
- Engineering Directorate (ENG)
 - [Directorate-wide Guidance](#)
- Geosciences Directorate (GEO)
 - [Directorate-wide Guidance](#)
- Mathematical and Physical Sciences Directorate (MPS)
 - [Division of Astronomical Sciences](#)
 - [Division of Chemistry](#)
 - [Division of Materials Research](#)
 - [Division of Mathematical Sciences](#)
 - [Division of Physics](#)
- Social, Behavioral and Economic Sciences Directorate (SBE)
 - [Directorate-wide Guidance](#)
- Cross-Directorate Programs
 - [Guidance for Designing Materials to Revolutionize and Engineer our Future \(DMREF\)](#)

DMP Deep Dive: NSF requirements and examples

Components of the DMP

1. Types of data
2. Standards for data and metadata
3. Policies for access and sharing
4. Policies for re-use, re-distribution, and the production of derivatives
5. Plans for archiving data
6. Roles and responsibilities

PAPPG [Chapter II.C.2.j](#)

1. Types of Data

- Describe the types of data to be produced in the course of the project
- How will data be processed?
- If applicable, what are your plans to obtain IRB approval?

1. Types of Data

- Data definition:
 - The recorded factual material commonly accepted in the scientific community as necessary to validate research findings
 - This includes original data, but also typically includes “metadata” (e.g., experimental protocols, code written for statistical analyses, etc.)

2. Standards for data and metadata

- What formats will the data or products be stored in (e.g., csv, TIFF)?
 - How may the data be converted to more accessible formats?
- What metadata (i.e., documentation) are needed to make the actual data products useful and reproducible for the general researcher?
- Where existing standards are absent or inadequate, this should be documented along with any proposed solutions or remedies

3. Policies for access and sharing

- Describe provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements
- Are there legal and ethical restrictions on access to non-aggregated data?
- What steps are you taking to ensure data privacy (e.g., de-identification?)

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

- Consider any conditions that may affect access or re-use of your data
- Will controlled or restricted access of data be required?
- What license or data use agreement will be applied to data and/or software? (e.g., [CC0](#), [CC-BY](#), [MIT Open Source](#))

5. Plans for archiving data

- What data repository will you use to make data accessible for the long term?
- When will you make the data available?
- What data transformations need to occur before preservation?
- What metadata will be submitted with the datasets?

5. Plans for archiving data

- Discipline-specific repositories
 - [PLOS](#) recommended repositories
 - [Scientific Data](#) guide to data repositories
 - [NIH](#) List of domain-specific repositories
 - [re3data](#) Registry of Research Data Repositories
- Restricted access repositories for social sciences data
 - [openICPSR](#)
 - [Qualitative Data Repository](#) (\$)
 - [DSDR](#) (Data Sharing for Demographic Research)

5. Plans for archiving data

- Generalist repositories
 - [Dryad](#)
 - [Figshare](#)
 - [Harvard Dataverse](#)

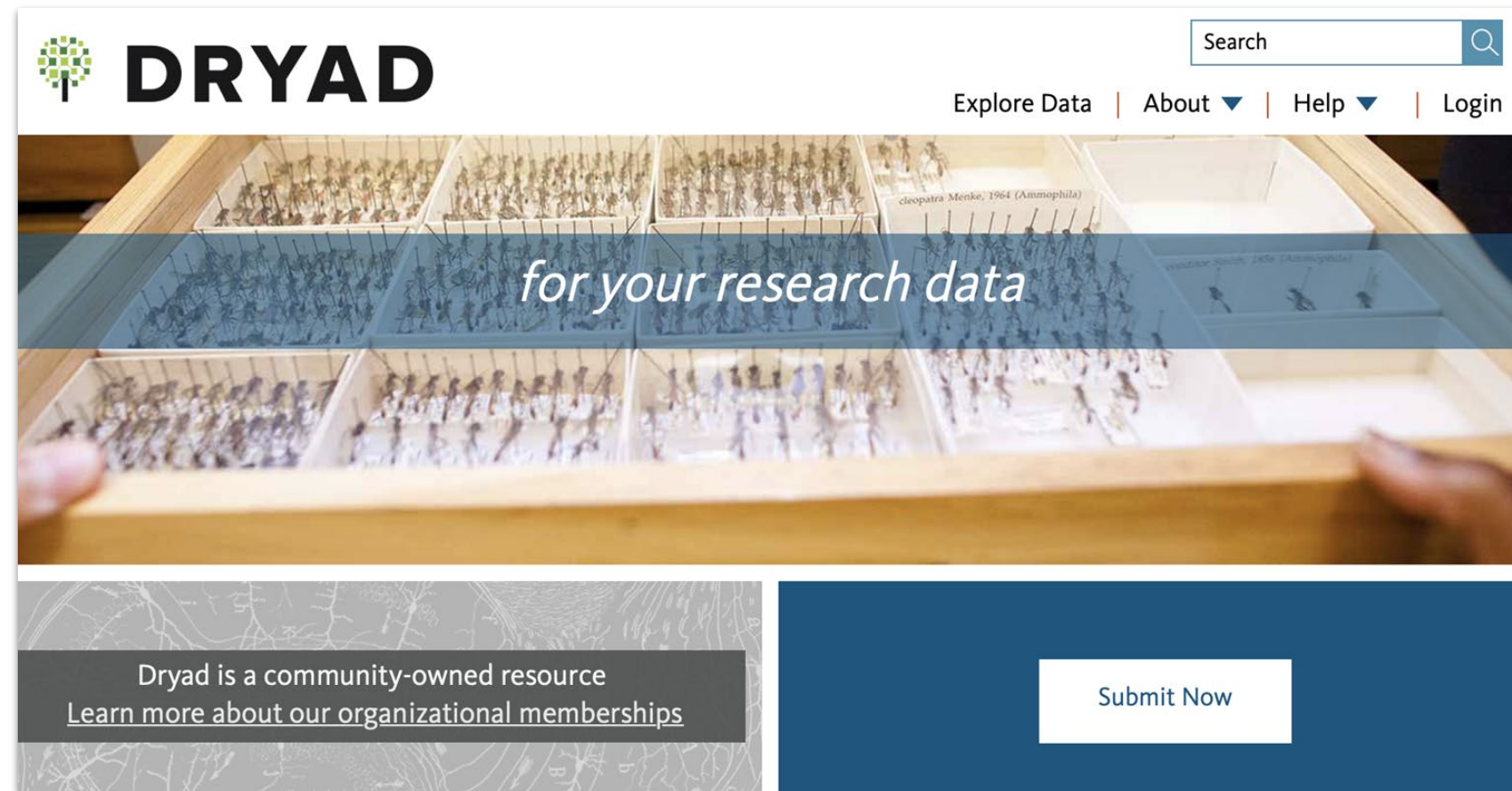
CSU Institutional Repositories



- All CSU campuses have open access institutional repositories
- Contact your campus library for data sharing policies
- Find your campus repository: <https://libraries.calstate.edu/ir/>

Dryad Digital Repository

- Generalist repository for the University of California and CSUEB
- Free to use for UC and CSUEB researchers
- Integrated with journals
- Curation process to check for FAIR compliance
- Open (CC0) license



<https://datadryad.org/stash>

6. Roles and responsibilities

- Who will be responsible for managing the data during each phase of research?
- Examples of possible roles:
 - project director
 - data collector
 - data analyzer/statistician
 - data model or database designer
 - computing staff responsible for backup and storage
 - staff responsible for running instruments

Budget

- Costs of data preparation and documentation
- Hardware, software
- Personnel involved in managing and curating data
- Repository fees

Example DMP

- Nabatchi, T. & R. McLain. (2021). Data Management Plan for NSF Proposal. Zenodo. <https://doi.org/10.5281/zenodo.4596772>

Data Management Plan

Roles and Responsibilities

The core project team consists of the principal investigator (PI), Dr. Nabatchi (Syracuse University), a co-principal investigator (co-PI), Dr. McLain (Portland State University), and a postdoctoral fellow (postdoc) (Syracuse University). The core team will design and oversee data collection instruments and activities, data analysis, and data management, as well as the generation of project products, including the database, data collection instruments and protocols, database user manual, publications, and presentations, among others. A doctoral student at Syracuse University and a master's student at Portland State University will assist with various tasks, such as producing instruments and protocols, coordinating and participating in data collection and analysis, and producing and disseminating publications, among other activities. In addition, the PI and co-PI, with input from the postdoc, will identify potential members and convene an advisory group to provide input on the data collection instruments and activities. Finally, we will work with the Syracuse University Qualitative Data Repository (QDR) to assist with organizing, curating, and making available the data and data collection instruments and protocols associated with the Atlas of Collaboration database. (The role of QDR is discussed in more detail below.)

Expected Data

The proposed project will produce several types of data in various data formats, as well as data collection instruments and protocols and a database user manual, most of which will be made available through QDR.

Data. The project will produce several types of quantitative and qualitative data.

- i. *Web-collected data* will be collected on the structural characteristics, physical geography, and social geography for approximately 300 collaborative governance regimes (CGRs) and dozens of collaborative platforms. The web-collected data will be gathered by visiting organizational websites, with data coded and entered into .xlsx or .csv files and deposited with QDR. Methods for inter-coder reliability will be used to ensure data quality.
- ii. *Survey data* will be collected from approximately 5,000 CGR participants. Survey data will be collected via Qualtrics, saved in .csv files, and deposited with QDR.
- iii. *Interview data* will be collected from approximately 75 CGR participants and leaders, collaborative platform managers, collaborative governance practitioners, and policymakers. The interviews will be recorded and transcribed, and qualitative software (e.g., NVivo or ATLAS.ti) will be used to code and analyze the data. Once transcribed, the recordings will be destroyed. We will remove direct identifiers from transcripts before analyzing them and store a re-identification key on a protected server at Portland State University. De-identified transcripts will be deposited with QDR along with coded data in REFI-QDA.

Example DMP

Roles and Responsibilities

The core project team consists of the principal investigator (PI), Dr. Nabatchi (Syracuse University), a co-principal investigator (co-PI), Dr. McLain (Portland State University), and a postdoctoral fellow (postdoc) (Syracuse University). The core team will design and oversee data collection instruments and activities, data analysis, and data management, as well as the generation of project products, including the database, data collection instruments and protocols, database user manual, publications, and presentations, among others. A doctoral student at Syracuse University and a master's student at Portland State University will assist with various tasks, such as producing instruments and protocols, coordinating and participating in data collection and analysis, and producing and disseminating publications, among other activities. In addition, the PI and co-PI, with input from the postdoc, will identify potential members and convene an advisory group to provide input on the data collection instruments and activities. Finally, we will work with the Syracuse University Qualitative Data Repository (QDR) to assist with organizing, curating, and making available the data and data collection instruments and protocols associated with the Atlas of Collaboration database. (The role of QDR is discussed in more detail below.)

Example DMP

Data. The project will produce several types of quantitative and qualitative data.

- i. *Web-collected data* will be collected on the structural characteristics, physical geography, and social geography for approximately 300 collaborative governance regimes (CGRs) and dozens of collaborative platforms. The web-collected data will be gathered by visiting organizational websites, with data coded and entered into .xlsx or .csv files and deposited with QDR. Methods for inter-coder reliability will be used to ensure data quality.
- ii. *Survey data* will be collected from approximately 5,000 CGR participants. Survey data will be collected via Qualtrics, saved in .csv files, and deposited with QDR.
- iii. *Interview data* will be collected from approximately 75 CGR participants and leaders, collaborative platform managers, collaborative governance practitioners, and policymakers. The interviews will be recorded and transcribed, and qualitative software (e.g., NVivo or ATLAS.ti) will be used to code and analyze the data. Once transcribed, the recordings will be destroyed. We will remove direct identifiers from transcripts before analyzing them and store a re-identification key on a protected server at Portland State University. De-identified transcripts will be deposited with QDR along with coded data in REFI-QDA.
- iv. *Coded legal texts* of approximately 13 pieces of state legislation and agency policy will be collected. The legal texts will be coded with qualitative software (e.g., NVivo or ATLAS.ti). The coded data will be deposited with QDR in REFI-QDA format.

Example DMP

Data Formats. We will enter and store data in standard formats (e.g., .docx, .xlsx, .csv), which are used widely and easy to work with in a wide range of applications. For long-term storage, we will follow QDR's preservation policy and convert data files and outputs to the format most appropriate for the dataset and the repository. The research team will decide on those formats in collaboration with QDR and in response to their expert advice.

Data collection instruments and protocols. The research team will prepare at least three sets of instruments and protocols, including: (i) a protocol for gathering web-collected data on CGRs and collaborative platforms; (ii) a protocol for administering the CGR participant survey instrument; and (iii) a protocol for using the interview instrument for actors engaged with collaborative governance. These instruments and associated protocols will be made available via QDR for use by others to contribute to the Atlas of Collaboration.

Database user manual. The research team will work with QDR to develop a manual to guide prospective users of the Atlas of Collaboration in conducting data collection and entry. The manual also will provide information about the types of analyses one can do with the data, with the goal of making the database a publicly available tool that scholars and practitioners will find useful. The database user manual will be deposited with QDR.

Example DMP

Data Storage, Preservation, and Sharing

Back-ups and Data Security. All data and materials will be kept on the password-protected work computers belonging to the PI (at Syracuse University) and the co-PI (at Portland State University), which are backed up on the universities' secure networks. In addition, the data will be synced through the Dropbox app using the research team's professional account, which provides multi-factor authentication and version history.

Data Preservation and Sharing. The PI and co-PI have communicated with the Qualitative Data Repository (QDR) regarding their plan to deposit in QDR the data, data collection instruments and protocols, database user manual, and other documentation generated through the research project. QDR staff have confirmed these project-generated materials are suitable for archiving with QDR. As the designated archive, QDR will take responsibility for managing the data and documentation after they are deposited and will make them available to the broader social science community as openly as possible. In particular, the web-scraped data, survey data, and coded legal texts will be freely accessible without restrictions. De-identified interview transcripts will be made available with appropriate access controls based on their sensitivity. The controls will be developed by the research team in consultation with the advisory group and QDR. Curation and storage fees as quoted by QDR are included in the grant's budget.

Example DMP

Period of Data Retention. All data will be deposited with QDR within 12 months of project completion or with the publication of relevant articles and papers, whichever comes earlier. QDR guarantees a minimum retention of 25 years for all deposited data, while aiming to ensure their accessibility in perpetuity.

More DMP examples shared by researchers

- Qualitative Data Management Plan Competition ([DMPTool blog](#))
- Public plans shared by researchers using DMPTool (not vetted):
https://dmptool.org/public_plans

DMP Resources

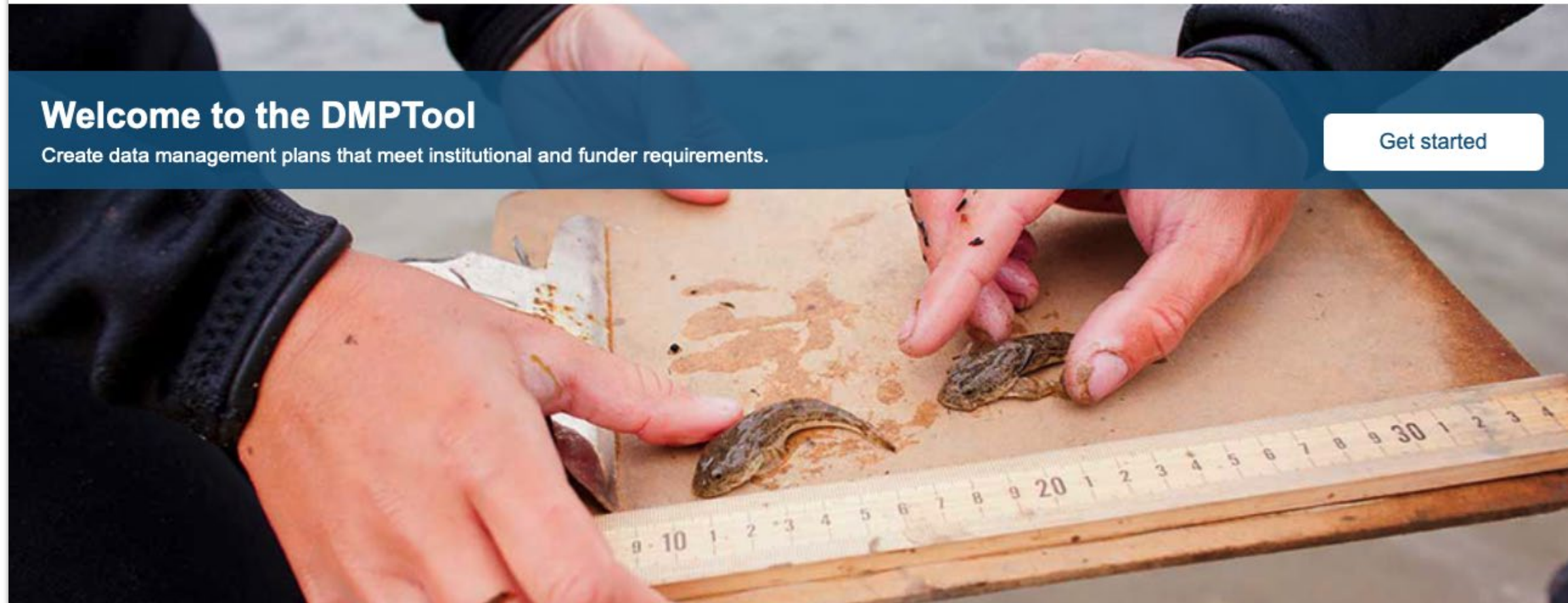


- Free online application for DMP creation and guidance
- Customized templates for US funders (NSF, NIH, DOE, etc...)
- All UC and some CSU campuses are members (not required for use)
- Use institutional login for customized guidance from your campuses
 - <https://dmptool.org>

Welcome to the DMPTool

Create data management plans that meet institutional and funder requirements.

Get started



DMPTool by the Numbers



59,708
Users



57,344
Plans
[More](#)



296
Participating Institutions
[More](#)

Top Templates

Digital Curation Centre
Template USP - Baseado no DCC
Template USP - Mínimo
UNICAMP-GENÉRICO: Aplicável a todas as áreas
USDA-NIFA: National Institute of Food and Agriculture

[More](#)

Available at: <https://dmptool.org>



National Science Foundation (NSF)

- Project Details
- Collaborators
- Write Plan
- Research Outputs
- Request feedback
- Download
- Finalize / Publish

This plan is based on the "NSF-BIO: Biological Sciences" template provided by National Science Foundation (NSF) - (ver: 5, pub: 2020-08-13).

expand all | collapse all

0/6

- + Data and Materials Produced (0 / 1) +
- + Standards, Formats and Metadata (0 / 1) +
- + Roles and Responsibilities (0 / 1) +
- + Dissemination Methods (0 / 1) +
- + Policies for Data Sharing and Public Access (0 / 1) +
- + Archiving, Storage and Preservation (0 / 1) +

National Science Foundation (NSF)

- Project Details
- Collaborators
- Write Plan
- Research Outputs
- Request feedback
- Download
- Finalize / Publish

This plan is based on the "NSF-BIO: Biological Sciences" template provided by National Science Foundation (NSF) - (ver: 5, pub: 2020-08-13).

expand all | collapse all

0/6

+ Data and Materials Produced (0 / 1) -

Describe the types of data, physical samples or collections, software, curriculum materials, and other materials to be produced in the course of the project. (For collaborative proposals, the DMP must cover all the various data types being collected by each collaborator.)

B *I* [List icons] [Link icon] [Table icon]

Save

- Guidance
- Comments

NSF DMPTool

- NSF-BIO Guidance on DMPs
- NSF Proposal & Award Policies & Procedures Guide (PAPPG)
- NSF plans for data management and sharing of the products of research (PAPPG)
- NSF Frequently Asked Questions (FAQs) for Public Access

+ Standards, Formats and Metadata (0 / 1) +



Funder Requirements

Templates for data management plans are based on the specific requirements listed in funder policy documents. The DMPTool maintains these templates, however, researchers should always consult the program officers and policy documents directly for authoritative guidance. Sample plans are provided by a funder or another trusted party.

Template Name	Download	Organization name	Last Updated	Funder Links	Create a new plan	Sample Plans (if available)
Alfred P. Sloan Foundation		Alfred P. Sloan Foundation	09-10-2020	Sloan Grant Proposal Guidelines	Requires login	
Arctic Data Center: NSF Polar Programs		National Science Foundation (NSF)	08-13-2020	NSF Arctic Data Center DMP Resources NSF Proposal & Award Policies & Procedures Guide (PAPPG)	Requires login	
BCO-DMO NSF OCE: Biological and Chemical Oceanography		National Science Foundation (NSF)	09-04-2020	NSF OCE Sample and Data Policy NSF GEO Directorate Guidance	Requires login	
Department of Defense (DOD)		United States Department of Defense (DOD)	09-13-2018	DOD Public Access Plan Data Archiving Plans for NIJ Funding Applicants	Requires login	
Department of Energy (DOE): Generic		United States Department of Energy (DOE)	07-13-2020	DOE Policy for Digital Research Data Management DOE Suggested Elements for a Data Management Plan DOE Office of Science Statement on Digital Management DOE Office of Energy Efficiency and Renewable Energy (EERE) Policy on Digital Data Management	Requires login	
Department of Energy (DOE): Office of Energy Efficiency and Renewable Energy (EERE)		United States Department of Energy (DOE)	08-02-2019	EERE Digital Data Management Suggested Elements for an EERE Data Management Plan	Requires login	
Department of Energy (DOE): Office of Science		United States Department of Energy (DOE)	07-13-2020	DOE Office of Science Statement on Digital Data Management DOE Suggested Elements for a Data Management Plan Additional Requirements and Guidance from Office of Science Program Offices	Requires login	
Digital Curation Centre		Digital Curation Centre (DCC)	05-29-2018		Requires login	

Links to funder requirements and specific plan templates (no sign-in required to access)

https://dmptool.org/public_templates

Your campus library!

- Connect with your campus data person
 - Data Services or Scholarly Communications librarians
 - UC list of contacts: <https://uc-love-data-week.github.io/#contact>
- Refer to data-related library research guides
 - For example: <https://libraryguides.fullerton.edu/DMP>

Questions?

- Get in touch!

Email: wdahdul@uci.edu