

GUIDANCE FOR DATA MANAGEMENT PLANS

Rachel Adams, University of Colorado Boulder

Maria Esteva, The University of Texas at Austin

Lori Peek, University of Colorado Boulder

This check sheet offers guidance on writing a data management plan based on current National Science Foundation (NSF) requirements. This page is also intended for researchers who plan to publish their data, data collection protocols, or instruments via the NSF-funded [DesignSafe Cyberinfrastructure](#).

What is a data management plan?

The NSF requires researchers to submit a data management plan (DMP) as part of the proposal process. This is a supplementary document of no more than two pages that describes how the proposed research will conform to NSF policy on the dissemination and sharing of research results. The NSF will not accept any proposal that is lacking a DMP. Even if the research is purely theoretical and no data are to be produced, a DMP must still be uploaded as part of the proposal submission process (in this instance, it would simply state that no data are to be produced, collected, or analyzed as part of the research). However, it should be noted that papers, presentations, websites, and learning materials produced during a project are considered publishable research outputs.

What information should be included in a data management plan?

Your data management plan (DMP) should address the following areas:

1. *Data Description*

“Data” are defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings. This includes original data (e.g., samples, physical collections, survey responses, interview transcripts, observations, curriculum materials, and other materials produced in the course of the study) and documentation about your project (e.g., experimental protocols, survey instruments, interview guides, code written for statistical analyses, etc.). To indicate the scope of your dataset we recommend that you record the data formats and prospective sizes.

2. *Data Documentation and Metadata Standards*

To ensure long-term availability and interoperability between computer systems, software, and your data, it is recommended that you use [open data formats](#). Your DMP should clearly explain your plans, if any, to convert proprietary data to open formats.

For your data to be retrieved and reused by others, you must describe it using metadata standards. Metadata standards are developed to establish a common understanding of the meaning of the data across different research projects. In general, social scientists use the [Data Documentation Initiative](#) (DDI) standard. DesignSafe-CI has incorporated DDI elements in its Field Research data model. In addition, an extensive vocabulary to describe the different types of research datasets was compiled to aid in the description of natural hazards datasets. DesignSafe-CI complies with the [Dublin Core](#) and [DataCite](#) metadata standards.

Non-existing or inadequate standards as related to specific projects should be documented in the DMP along with any proposed solutions or remedies. For example, if the existing metadata elements in DesignSafe-CI do not satisfy exhaustive data description, you may include documentation (i.e., data dictionary, codebook, help files) about why, how, and when the data was gathered, and about the meaning and semantics of the data. This will help assure proper interpretation of the data by other users.

DesignSafe-CI is designed to facilitate your data description through one of their data models that map the structure of your research project with all the information required for others to reuse your data. Social scientists, engineers, and other users who conduct field research will be able to select the “Field Research Data Model.” If you want to publish datasets based on other primary or secondary data sources, presentations, white papers, or other resources, you may use the “Other” data model.

3. **Data Security and Sharing**

The DMP should clearly articulate when and how you will share your research data with your team members during the active research project by outlining the provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements. Data in DesignSafe-CI is geographically replicated and monitored following security best practices. Storing your data in DesignSafe will allow you to access and share your data with your team members for the duration of the active research project.

Human Subjects Research

- * If conducting human subjects research, you must indicate in the DMP what extra precautions for storing and sharing protected data will be followed to comply with the procedures approved by your Institutional Review Board (IRB) and your own ethical commitment to participants. This includes data that will be stored in DesignSafe-CI or elsewhere while you are conducting research. Included are privacy protection and back-up best practices. There are some considerations before storing raw protected data in DesignSafe. You can store data with [Personally Identifiable Information](#) (PII) as long as the data is not protected under [HIPAA](#) and [FERPA](#) regulations, nor if it contains [sensitive information](#). If your data fall under those categories, you may still use DesignSafe, but you will have to request [protected storage services](#). If you are unsure about what category your data fits into, you can submit a [help](#) ticket or attend [virtual curation office hours](#).

4. **Data Distribution/Publication and Reuse**

The DMP should describe the data dissemination approaches to make data and metadata available to others outside your research study. Policies for public access and sharing should be described, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

DesignSafe-CI allows users to make data publicly available via community data (i.e., this is non-curated data that researchers make available) and published data (i.e., this is curated data). For published data, and as you go through the data curation and publication process, you will be able to select the license through which you wish to distribute your data. Licenses are fundamental to ensure that your data is used appropriately according to the nature of your research project. Importantly, when you publish your data you will receive a permanent Digital Object Identifier (DOI) and a citation. The former enables data authors to cite their datasets in papers and receive credit when their data is reused by others.

Specifically, for publishing protected data, you will have to consider some of the following procedures which you may describe in your DMP.

- a. Do not publish HIPAA, FERPA, PII data, or other sensitive information in DesignSafe.
- b. To publish protected data and any related documentation (i.e., reports, planning documents, fieldnotes, etc.) it must be properly anonymized. No direct identifiers are allowed. These include items such as participant names, participant initials, facial photographs unless expressly authorized by participants, home addresses, social security numbers, and dates of birth. Up to three indirect identifiers are allowed in a published dataset. These are identifiers that, taken together, could be used to deduce someone's identity. Examples include gender, household and family composition, religious beliefs occupation, places of birth, or year of birth/age.
- c. If a researcher needs to restrict public access to data because it includes highly sensitive personal information, or because removing the indirect identifiers will impair the data understandability, you may only publish metadata and other documentation about the data in DesignSafe. Users interested in the data will contact the PI (your email address will be published) to request access to the data and to discuss the conditions for its reuse.

While types of data and culture around publishing data often vary by scientific discipline, NSF is strongly committed to the underlying principle of timely and rapid data distribution. Time to release and processes for sharing data should be addressed in the data management plan. Research centers and major partnerships with industry or other user communities must also address how data are to be shared and managed with partners, center members, and other major stakeholders. See recommended [Responsibilities and Timeline](#) for publishing data in DesignSafe-CI.

5. Plans for Archiving and Presentation

The DMP should also include plans for long-term archiving of your published data and other research materials. This involves how the materials will be preserved and accessed over time. When you deposit your data and associated materials in the DesignSafe-CI repository, the Data Depot, you are meeting requirements for preservation and access. Your data management plan should note that DesignSafe-CI will maintain all uploaded data on storage resources at the Texas Advanced Computing Center (TACC), which ensures the authenticity, integrity, security, and persistence of datasets for open access.

6. Data Management Governance and Support

Researchers are encouraged to specify the roles of team members in managing and publishing data and how they will be supported throughout the research. If using DesignSafe-CI, the curation team provides support to researchers during office hours with advice on how to organize, describe, and publish your datasets.

For more information on the DesignSafe-CI - CONVERGE - RAPID joint field research data model for social scientists, engineers, and interdisciplinary teams, please visit this page: <https://converge.colorado.edu/data/data-publication>.

Frequently Asked Questions

- *Can I still use this information if I am required to write a data management plan for another funding agency?*
This check sheet draws on NSF requirements and guidance for data management. However, we think that some of this information can be useful regardless of funding source. You should always read and adhere to the guidance for your specific funder.
- *Can I still use DesignSafe resources if I do not have NSF funding?*
Yes. DesignSafe is the cyberinfrastructure for the natural hazards and disaster research community. You can learn more and sign up for a free account at the [DesignSafe website](#).

- *How can I learn more?*

The information presented here is derived from NSF's [Proposal and Award Policies and Procedures](#), DesignSafe's [Data Management Plan](#) and [Data Publication Guidelines](#), and the NSF's [Social, Behavioral, and Economic Sciences Directorate](#) and [Engineering Directorate](#) guidance. Investigators should be aware that due to disciplinary nuances, different directorates at NSF may offer slightly different guidance regarding data management plans. For this reason, investigators are encouraged to review [directorate-level guidance](#) on data sharing and data management as well.

Additional considerations for data management issues for the social, behavioral, and economic sciences are available in a workshop report, [Public Access to NSF-Funded Research Data for the Social, Behavioral, and Economic Sciences](#).

REFERENCES:

DesignSafe Cyberinfrastructure. (n.d.) *DesignSafe Data Management Plan Guide*. DesignSafe-CI. <https://www.designsafe-ci.org/rw/user-guides/>.

National Science Foundation. (2019). *Proposal and Award Policies and Procedure Guide* (OMB Control Number 3145-0058). National Science Foundation. https://www.nsf.gov/pubs/policydocs/pappg19_1/nsf19_1.pdf.

National Science Foundation. (2018). *Data Management for NSF SBE Directorate Proposals and Awards*. National Science Foundation. https://www.nsf.gov/sbe/DMP/SBE_DataMgmtPlanPolicy_RevisedApril2018.pdf.

National Science Foundation. (2018). *Directorate for Engineering Data Management Plans Guidance for Principal Investigators*. National Science Foundation. https://nsf.gov/eng/general/ENG_DMP_Policy.pdf.

Suggested Citation: Adams, R., Esteva, M., & Peek, L. (2020). Guidance for Data Management Plans. CONVERGE Extreme Events Research Check Sheets Series. DesignSafe-CI. <https://doi.org/10.17603/ds2-ycz2-xc47>.

