

NSF Research Experiences for Undergraduates (REU) Data Management Plan Template

# Data Management Plan Requirements

Proposals submitted must include a supplementary document of **no more than two pages** labeled “**Data Management Plan**”. This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See the NSF [Data Management Plan Requirements](https://www.nsf.gov/bfa/dias/policy/dmp.jsp) and [Proposal and Award Policies and Procedures Guide (PAPPG)](https://www.nsf.gov/publications/pub_summ.jsp?ods_key=pappg) for more information about this policy.

Each directorate may also have specific guidelines that address unique data management issues within the respective community. Be sure to look at the proposal details and main directorate and/or division website for additional guidelines. If you have any questions regarding directorate or program specific Data Management Plan guidelines, please contact your designated Program Officer listed in the program solicitation.

***Note:*** *Data Management Plans for REU proposals may differ from those in other NSF programs. Data that is generated/collected from lab activities and data collected from student applications and assessments are typically the focus of REU DMPs. However, it is important to discuss DMP expectations with your division/directorate Program Officer in addition to referring to the REU solicitation and NSF PAPPG. This template is intended to serve as a general guide and reference when writing your DMP.*

# Help with Your Data Management Plan

To create your Data Management Plan, go through the sections on the next page. Start by answering the questions within each section (numbered under each section description). After answering all of the questions, remove the questions, leaving just your answers. Modify the answers into prose that makes sense as a paragraph below each Roman numeral header (include the bold text as the header to each of your sections in your Data Management Plan). The library also has a guide on DMPs that might help provide additional information and guidance:

<https://guides.library.tamucc.edu/c.php?g=1046360&p=7595962>.

*If you need additional help with your Data Management Plan, please contact Jayne-Marie Linguist (*[*jayne-marie.linguist@tamucc.edu*](mailto:jayne-marie.linguist@tamucc.edu)*) or (*[***researchdevelopment@tamucc.edu***](mailto:researchdevelopment@tamucc.edu)*).*

# Data Management Plan

## I. Types of data

Samples, physical collections, software, curriculum materials, and other materials to be produced during the project.

1. What data will be generated in the research? (Give a short description, including amount – if known and the content of the data).
2. What data types will you be creating or capturing? (e.g., experimental measures, observational or qualitative, model simulation, processed etc.)
3. How will you capture or create the data?
4. If you will be using existing data, state that fact and include where you got it. What is the relationship between the data you are collecting and the existing data?

## II. Data and Metadata Standards

Standards to be used for formatting data and metadata (information about other data) and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies).

1. Which file formats will you use for your data, and why?
2. What contextual details (metadata) are needed to make the data you capture or collect meaningful?
3. How will you create or capture these details?
4. What form will the metadata take?
5. Which metadata standards will you use?
6. Why have you chosen particular standards and approaches for metadata and contextual documentation? (e.g., staff expertise, Open Source, accepted domain-local standards, widespread usage)

## III. Policies for access and sharing and provisions for appropriate protection/privacy

1. How will you make the data available? (Resources needed: equipment, systems, expertise, etc.)
2. When will you make the data available? (Give details of any embargo periods for political/commercial/patent reasons.)
3. What is the process for gaining access to the data?
4. Will access be chargeable?
5. Will there be a cost to the user to access the data or is it free?
6. Does the original data collector/ creator/ principal investigator retain the right to use the data before opening it up to wider use?

Provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

1. Are there ethical and privacy issues?
2. If so, how will these be resolved? (e.g., anonymization of data, institutional ethical committees, formal consent agreements.)
3. What have you done to comply with your obligations in your IRB Protocol?
4. Is the dataset covered by copyright? If so, who owns the copyright and other intellectual property?
5. How will the dataset be licensed if rights exist? (e.g., any restrictions or delays on data sharing needed to protect intellectual property, copyright, or patentable data.)

## IV. Policies and provisions for re-use, re-distribution

1. Will any permission restrictions need to be placed on the data?
2. Which bodies/groups are likely to be interested in the data?
3. What and who are the intended or foreseeable uses / users of the data?
4. Are there any reasons not to share or re-use data? (Suggestions: ethical, non- disclosure, etc.)

## V. Plans for archiving and preservation of access

1. Plans for archiving data, samples, and other research products, and Preservation of access to them.
2. What is the long-term strategy for maintaining, curating, and archiving the data?
3. Which archive/repository/central database/ data center have you identified as a place to deposit data? (See **TAMU-CC Repositories Available** section below)
4. What transformations will be necessary to prepare data for preservation / data sharing? (e.g., data cleaning/anonymization where appropriate.)
5. What metadata/ documentation will be submitted alongside the data or created on deposit/ transformation in order to make the data reusable?
6. What related information will be deposited (e.g., references, reports, research papers, fonts, the original bid proposal, etc.)?
7. How long will/should data be kept beyond the life of the project?
8. What procedures does your intended long-term data storage facility have in place for preservation and backup?

# TAMU-CC Repositories Available

You may want to consider using one of these repositories for your data:

## 1. The Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC)

The Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC) is a team of researchers, data specialists and computer system developers who are supporting the development of a data management system to store scientific data generated by Gulf of Mexico researchers. The [Master Research Agreement](http://gulfresearchinitiative.org/about-gomri/master-research-agreement/) between BP and the Gulf of Mexico Alliance that established [the Gulf of Mexico Research Initiative (GoMRI)](http://gulfresearchinitiative.org/) included provisions that all data collected or generated through the agreement must be made available to the public. The Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC) is the vehicle through which GoMRI is fulfilling this requirement.

GRIIDC is housed at the [Harte Research Institute for Gulf of Mexico Studies (HRI)](http://www.harteresearchinstitute.org/) at [Texas A&M University – Corpus Christi (TAMUCC)](http://www.tamucc.edu/). Administrative staff including the GRIIDC Director and Program Manager are supported by a technical staff team that includes software engineers, data analysts, web developers, subject matter experts and partners at Texas A&M University and the Florida Fish and Wildlife Research Institute. Guiding the work of staff at GRIIDC is an Advisory Committee that includes one representative from each the GoMRI Research Consortia, the GoMRI Administrative Unit, GoMRI Research Board Data Committee, the National Oceanic and Atmospheric Administration’s (NOAA) National Centers for Environmental Information (NCEI) as well as GoMRI’s Chief Scientific Officer and the Director of GRIIDC. The Advisory Committee meets on a bi-monthly basis to provide GRIIDC direction and feedback through evaluation of past activities. Working groups are formed by the Advisory Committee to investigate issues and develop recommendations on GRIIDC policies and program direction.

The GRIIDC team is building a data management system for scientific data generated by Gulf of Mexico researchers. The GRIIDC data management system provides researchers with a variety of tools to help manage data throughout the lifecycle of a project. For example, the GRIIDC Dataset Information Form (DIF) is a resource designed to assist researchers with data management planning. While the system assists researchers with multiple phases of data management, the main functions of the system are storing and sharing data. Researchers from diverse fields of study, including biology, chemistry, physical oceanography, sociology, political science, and public health, are able to store their data in the GRIIDC system. Through the GRIIDC Search page, researchers, policy makers, and the general public are able to search for and download this data. This shared data can be used to address innovative scientific research questions, assess policies and programs, and in educational initiatives. By providing a forum for both storing and sharing data the GRIIDC system increases the impact of scientific research in the Gulf of Mexico and beyond for the benefit of society.

## 2. TAMU-CC Repository

The TAMU-CC Repositoryis the university’s digital institution repository focused on collecting, managing, and disseminating important digital content produced by the members of an institution like a university. The TAMU-CC repository (<https://tamucc-ir.tdl.org/>) is open to all members of the university community to contribute content for persistent long-term access and management. Management activities include preservation activities, application of metadata standards to promote discovery and access, and persistent links to content for use in citing and linking to works.

The TAMU-CC Repository collects and provides long term public access to the scholarly output of the students, staff, faculty, and researchers of Texas A&M University - Corpus Christi in order to increase the exposure and impact of scholarship at TAMU-CC and to raise the profile of TAMU-CC scholarship on the World Wide Web. The repository is also the main digital dissemination platform for library special collections.

Users can post open access versions of their scholarship in the repository for online access by the public. The objective of including an open access version of faculty works in an institutional repository is to enable open versions of academic literature to be found, re-used, and have impact beyond what is possible if works are only available behind a paywall. Essentially, hosting published works in a repository makes them more accessible, which means works will be seen, read, and used by more users, expanding the reach of published works beyond who has access to the particular journal in which the work was published. The library can help users verify permissions and rights for posting scholarly works online. A persistent link for each item in the repository is available for use in linking to the work from personal and departmental web pages, CVs, and other documents.