

# How to Develop a Data Management Plan for the Gulf of Mexico Research Initiative Information & Data Cooperative

## What is a data management plan?

The data lifecycle model illustrates the processes involved in managing data during a research project (Figure 1<sup>1</sup>). A data management plan is a document that describes how data will be handled and organized during each phase of the data lifecycle.

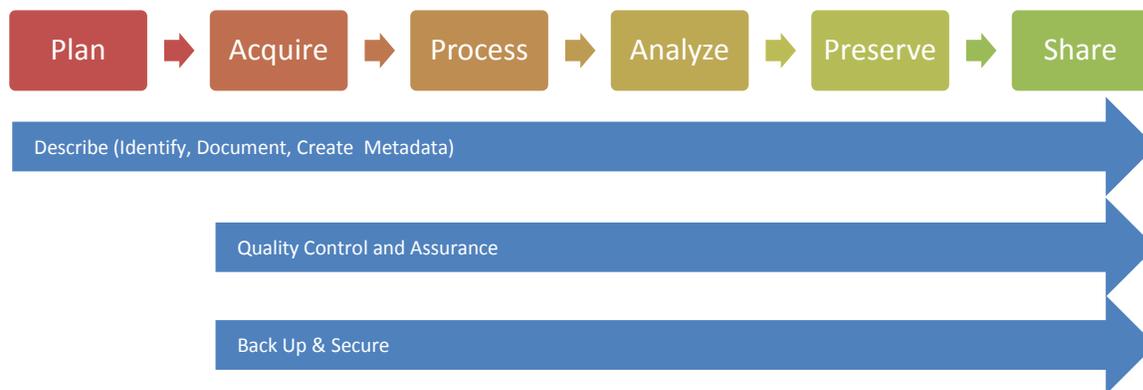


Figure 1: The Data Lifecycle Model. Adapted from Faundeen *et al*, 2013.

A data management plan should describe what data will be produced, and how data will be organized, documented, and stored. Additionally, a data management plan should clearly define roles and responsibilities for data management. A data management plan does not necessarily include details of methods that will be used to process or complete statistical analysis on data, but should include information about, for example, how processed data will be backed up.

## Why develop a Data Management Plan?

Research projects often have a finite lifespan, from the time the grant funding is received to the time the grant is closed-out. However, data are often used and re-used after the research project that created them is completed. By preparing and implementing a data management plan, data can be organized and documented in a way that facilitates re-use. Implementing a data management plan during the course of a project is easier, more efficient, and more cost effective than attempting to organize and document data at the end of a project.

<sup>1</sup> Faundeen, J.L., Burley, T.E., Carline, J.A., Govoni, D.L., Henkel, H.S., Holl, S.L., Hutchison, V.B., Martin, Elizabeth, Montgomery, E.T., Ladino, C.C., Tessler, Steven, and Zolly, L.S., 2013. The United States Geological Survey Science Data Lifecycle Model: U.S. Geological Survey Open-File Report 2013-1265, 4p., <http://dx.doi.org/10.3133/ofr20131265>.

The Gulf of Mexico Research Initiative (GoMRI) Data Management Policy Framework requires that each Research Consortium and non-affiliated Principal Investigator develop a data management plan to be approved by the Gulf of Mexico Research Initiative Information & Data Cooperative (GRIIDC).

## Writing a Data Management Plan

GRIIDC has developed a data management plan template for each Research Consortium to use to create data management plans. The data management plan template is a checklist of information and questions that should be addressed in the data management plan. Data management plans required by GoMRI have three sections: Research Consortium Information, Detailed Task Information, and Dataset Information Forms. These sections should be completed through consultation with researchers reflecting the best knowledge available at the time.

## Advice for Writing a Data Management Plan

This section is intended to guide Research Consortium Data Managers, Principal Investigators, and others who have responsibilities for developing and writing a data management plan.

### 1. Create a Data Management Plan Team

The first step in writing a data management plan is designating a team that will be responsible for coordinating and writing the plan. This team will likely include the data manager and principal investigator(s), and may include project coordinators, and other research team members.

The team should determine who will be responsible for overall plan preparation and writing, and what roles other team members will play in plan development.

### 2. Prepare and Organize

The team should review and understand the information required through the GRIIDC data management plan template. The team should determine what information is needed to complete the data management plan and who to consult to gather this information. The team may need to consult with Research Consortium scientists, institution administration, or IT departments to complete different sections of the data management plan. The team should agree on consultation methods that may be used to gather information.

If the Research Consortium is using the online version of the GRIIDC data management plan template (<https://dmptool.org/>) the team should log into the system and become familiar with the template and system functionality before beginning the consultation process.

### 3. Consult and Collaborate

Once the team has determined who must be consulted and the information that is required to complete the plan, the team can initiate the consultation process. Early consultation and communication can help build and strengthen working relationships within the Research Consortium, in particular between researchers and the data manager. Consultation will likely be an iterative process, involving back and forth communication to clarify details.

### 4. Use Existing Support

The GRIIDC data management plan template is the primary tool to use while developing the Research Consortium data management plan. Institutional resources may be available through academic library services to assist in data management planning. Additionally, IT resources at institutions within the Research Consortium may be able to provide technical advice about matters such as back-up policies and procedures.

GRIIDC staff members are available to provide assistance and support to Research Consortium, as needed, as they develop and implement data management plans. GRIIDC staff can be consulted during the writing of the data management plan to make sure that the plan will meet GRIIDC requirements.

## **5. Provide training**

Section 3 of the GRIIDC DMP Template requires that Dataset Information Forms (DIFS) be completed to identify datasets that will be generated. The DIF is a data management planning tool that helps researchers consider elements of data management early in their project. Additionally, identifying datasets early helps GRIIDC plan the design of the data management system and its infrastructure, including estimating the amount and type of storage the system requires. GRIIDC encourages researchers to complete DIFs for datasets they plan to collect or generate. When researchers complete DIFs they become more familiar and comfortable using the GRIIDC system. To encourage the use of the GRIIDC system, GRIIDC offers periodic training sessions to data managers and researchers. However, researchers may require extra support to complete tasks in the GRIIDC system, such as completing DIFs, and data managers should be prepared to provide additional training and support as required.

Additionally, data managers should provide data management practices training to members of the Research Consortium, including researchers, scientific assistants, laboratory technicians, post-doctoral fellows and graduate students. By encouraging the use of data management best practices during the acquisition, creation, processing, and analysis of data, data managers can reduce time and effort required to organize and document data for submission to the GRIIDC system.

## **Once the Data Management Plan is Approved**

The data management plan must be approved by GRIIDC. When the Research Consortium Data Management Plan is approved it should be distributed to all members of the Research Consortium. It may also be posted on the Research Consortium website to make it easily accessible to all Research Consortium members. All members of the Research Consortium should have knowledge of the data management plan and of their responsibilities to manage data to facilitate submission to GRIIDC or national data archives, such as NIH NCBI GenBank.

The final step in data management is implementing and maintaining the approved data management plan. As the plan is implemented, changes may be required to reflect the most up-to-date and accurate information. For example, roles and responsibilities for data management may change when the project is implemented. The Research Consortium may schedule periodic reviews and updates to the data management plan, however all changes to the data management plan must be approved by GRIIDC.



## Notes