

Dataset Information Form (DIF) Guide

What is a Dataset Information Form?

A Dataset Information Form (DIF) is a data management planning tool required for datasets that will be submitted to GRIIDC. The DIF should be created before data are collected or generated, and helps researchers plan and prepare for data collection and organization of data submission. These details also assist GRIIDC in designing the data management system, determining storage capacity, and preparing to distribute your dataset. GRIIDC understands that the information provided in the DIF is preliminary and subject to change. After researchers have collected or generated data, they will use dataset submission to submit data and can update and add to the information that was provided in the DIF.

Section 2: User Account

To submit a DIF, you must first have an account with the GRIIDC data management system.

Section 3: Dataset Information Form - Datasets List

The GRIIDC DIF contains an interactive form on the left and a list titled “Datasets” on the right (Figure 1). In the list of datasets on the right, there will be one or more folders. Each folder represents a project that you are associated with in the GRIIDC system. Projects in the GRIIDC system are obtained from grant funding agreements from organizations that have contracted GRIIDC to provide data management services. If a DIF has been saved or submitted for a project, the dataset GRIIDC Unique Dataset Identifier (UDI) and title will be listed under the project folder. If a project has no DIFs saved or submitted, the folder will not have any datasets listed under it and the folder icon will be grayed out (e.g., ).

One of three symbols, described below, will appear beside each dataset title within a folder.

-  Open: A DIF has been started and saved for this dataset but has not yet been submitted to GRIIDC.
-  Submitted: A DIF has been submitted for this dataset and is in review by GRIIDC staff. The DIF is locked to prevent edits at this stage.
-  Approved: A DIF has been approved for this dataset. The DIF is locked to prevent edits at this stage. To unlock an approved DIF, please use the “Unlock” button found at the bottom of the form. Please note that the information provided in the DIF can be edited when you submit your dataset. There is usually no need to unlock the DIF to provide updated information at this stage.



Figure 1: Dataset Information Form, example view. In this hypothetical example, the user is associated with one project in the GRIIDC system: GRIIDC Example Research Group and 3 DIFs are listed for this project.

Section 4: Creating a New Dataset Information Form

Required fields have a red asterisk (*). To begin, select the project that the dataset is associated with from the drop-down list. For more details about what information you should provide for each field, please see Section 6 below or hover over the information icon (i) found on the form. To clear all the fields and start over, select “Reset Form”, found at the bottom of the form.

You have two options to end your session:

1. **Save and Continue Later:** If you are interrupted or need to collect additional information before you can complete your DIF, click “Save and Continue Later”. When you return to the DIF page, the dataset title will appear in the dataset list on the right under the project folder you previously selected. Click the dataset title to load the information into the form and continue to add and edit information as required.
2. **Save and Submit to GRIIDC:** When you have finished editing the DIF, select “Save and Submit to GRIIDC”. This will submit your DIF for review and approval by GRIIDC staff. The DIF will be locked to prevent additional edits.

When a DIF is saved, it is assigned a Unique Dataset Identifier (UDI), a unique string of numbers and letters that can be used to reference your dataset in the GRIIDC system. When communicating to the GRIIDC team about your dataset, please reference the UDI.

NOTE: The GRIIDC Unique Dataset Identifier (UDI) is not a persistent identifier like a Digital Object Identifier (DOI). It should be used only as a reference to a DIF record and should **not** be used to cite datasets in peer-reviewed journal articles.

You will receive an email confirmation after submitting a DIF. The DIF will then be reviewed by a GRIIDC staff member for completeness and errors. If additional information or changes are required, a GRIIDC staff member will contact you via email. Once the DIF has been reviewed and determined to be complete, you will receive an email notification that your DIF has been approved. The DIF will be locked to prevent any changes after approval.

Section 5: Editing a Submitted DIF

The DIF is primarily a planning document and the information you provide is subject to change. You are not required to update the DIF. Please plan to update information about your dataset, including the title and abstract, when the dataset is submitted to GRIIDC using Dataset Submission. Submitted DIFs are locked to prevent edits after submission. Unapproved DIFs can be updated by emailing GRIIDC to request that the DIF be unlocked. Please reference the dataset UDI in your request. Approved DIFs can be unlocked by selecting “Request to Unlock” from the bottom of the form. When the DIF is unlocked, the status will revert to open and the icon will be reverted to a red **x**.

Section 6: Fields and Examples

Project Title



*Project Title:
[PLEASE SELECT A PROJECT] 

This is a required field. Project titles in the GRIIDC data management system have been obtained from funding proposals and/or agreements of organizations that have contracted GRIIDC to provide data management services. The drop-down menu lists all projects that you are associated with in the GRIIDC system.

Funder



*Add Funder(s) 
Choose Funder...
Additional Funders 
[]

This is a required field. Select one or multiple Funders from the list. This field represents the organization that provided funding for the dataset. If the Funder is not on the list, type in the name under the Additional Funders section.

Dataset Title



*Dataset Title: 
Dataset Title (200 Character Maximum)

This is a required field. Please provide a descriptive title that briefly explains the contents of your dataset. This may include the location where, and the time period when the research will be completed. It should be understandable by a user unfamiliar with your methods, collection sites, or research platforms. Note this field has a 200 character maximum.

Example Titles

(1) Ecological

Aerial survey data for the assessment of the distribution of cownose rays (*Rhinoptera bonasus*) in the Eastern Gulf of Mexico, from May to October 2023

(2) Chemical/Molecular Engineering

Image sequences of rising bubble plumes from laboratory study to calculate bubble velocity vector under various gas flow rates and vertical density gradients

(3) Oceanographic

Conductivity, temperature, and depth data for 12 northwestern Gulf of Mexico locations, May to July 2023

(4) Model/Numerical Model

Galveston Bay Circulation Study: Stanford Unstructured Nonhydrostatic Terrain following Adaptive Navier-Stokes (SUNTANS) models simulations for 2022-2023

(5) Sociology

Cross-sectional household survey response data to assess health and wellbeing of residents of coastal Louisiana, April 2023

(6) Genetics

Mitochondrial DNA control region sequences (297 base-pairs) from 140 northern red snapper (*Lutjanus campechanus*) collected from the Gulf of Mexico, 2022-2023

(7) Biochemistry

Polycyclic aromatic hydrocarbon concentrations in liver and muscle tissue from barracuda, escolar, and common dolphin fish, northeastern Gulf of Mexico, 2022-2023

Primary Data Point of Contact

*** Primary Point of Contact:** 

This is a required field. The Principal Investigator (PI) is usually the Primary Point of Contact. This is the person responsible for answering questions associated with the dataset and responding to requests about the dataset. Please select the appropriate person's name from the drop-down list.

Secondary Data Point of Contact

Secondary Point of Contact:

[PLEASE SELECT TASK FIRST] i

The Secondary Point of Contact will be contacted if the Primary Point of Contact cannot be reached or is unable to respond to questions about the dataset. Please select the appropriate person’s name from the drop-down list.

Dataset Abstract

***Dataset Abstract:**

Please provide a brief narrative describing: What, where, why, how, and when the data will be or have been collected/generated? (4000 Character Maximum) i

This is a required field. The dataset abstract should summarize the rationale for collecting the dataset, methods that will be used to collect or generate the dataset, time period of data collection, and what the dataset will contain. It does not need to include details about conclusions drawn or statistical analyses completed using the dataset. The abstract should answer what, why, how, when, and where the data will be or have been collected and/or generated. Do not submit a publication abstract. Note that this field has a limit of 4,000 characters.

Field of Study

Field of Study: i

<input type="checkbox"/> Ecological/Biological	<input type="checkbox"/> Physical Oceanography	<input type="checkbox"/> Atmospheric
<input type="checkbox"/> Chemical	<input type="checkbox"/> Human Health	<input type="checkbox"/> Social/Cultural/Political
<input type="checkbox"/> Economics	Others: <input style="width: 150px;" type="text"/>	

Datasets may be categorized based on the discipline(s) for which they are collected. Please check all the boxes that categorize the discipline(s) of your dataset. Note that datasets can be multidisciplinary so select all disciplines that apply. If your field of study is not listed, please specify additional details in “others”.

Approximate Dataset Size

Approximate Dataset Size: i

< 1GB
 1GB-10GB
 10GB-200GB
 200GB-1TB
 1TB-5TB
 >5TB

Please provide an estimate of the total size of the dataset files you expect will be created. The response provided in this field will help GRIIDC formulate strategies to facilitate access, storage, and distribution of your dataset. This field is important to provide if you will be submitting a very large dataset (>200 GB) to GRIIDC, so that GRIIDC can make plans to store your data.

Data Parameters and Units

Data Parameters and Units: i

Examples: wind speed (km/hr), salinity (ppt), temperature (°C), PCB concentrations in eggs from a specified species (ng/g wet weight), Ionic Strength (mM)

This is a required field. Please expand on the information you have provided in the abstract by giving more exact details about all parameters and units that you plan to measure, collect, or generate. If your data are submitted in a spreadsheet, please make sure all header information are included in this section. Please separate list items using a comma.

Examples of Data Parameters and Units: wind speed (km/hr), salinity (ppt), temperature (degrees Celsius), PCB concentrations in eggs from a specified species (ng/g wet weight), Ionic Strength (mM).

Method(s) for Collecting or Generating the Data

Method(s) for Collecting or Generating the Data: i

Field Sampling Simulated/Generated Laboratory
 Literature Based Remote Sensing Others:

Please select all the procedures or methods that you will use to collect or generate data in the dataset. If not available on the list, please provide additional information in the “Others” field.

Estimated Data Sampling/Generation Period

Estimated Data Sampling/Generation Period: i

Start Date to End Date

This is a required field. Please provide the estimated time period when the data is expected to be collected and/or generated. These dates do not need to be exact; an approximate time period is acceptable as unforeseen events often require project timelines to change.

Geographic/Study Area

Geographic/Study Area: i



[Define Spatial Extent](#)

The Spatial Extent Wizard is a guided tools to create the Spatial Extent of a dataset. Users can 1) Create a description to indicate that the data has no spatial or temporal extent, or 2) Create a spatial extent geometry by either providing a list of coordinates or drawing a geometry using a map.

Please use the GRIIDC Spatial Extent Wizard to provide the approximate geographic area where the data will be collected or generated, for example the area to be included in a model. If your dataset will be generated entirely in the laboratory, and therefore has no relevant geographic information, you will be able to describe this using the Spatial Extent Wizard. Please reference the Spatial Extent Wizard user guide for additional assistance using the Wizard.

Identify a National Data Archive for your Dataset, if applicable:

Identify a National Data Archive for your Dataset, if applicable:

- National Oceanographic Data Center
- US EPA Storet
- Global Biodiversity Information Facility
- National Center for Biotechnology Information
- Data.gov Dataset Management System

Others:

Datasets should be submitted directly to GRIIDC. However, your dataset may also be submitted to a recognized National Data Archive. If your dataset is submitted to a National Data Archive you still need to complete Dataset Submission so that the dataset can be catalogued with other datasets produced by your research project. If you will be submitting your dataset to a National Data Archive not on the list provided, please specify the data center and associated URL in the “Others” field.

Will this dataset include any data governed under the IRB or HIPAA, or data with other ethical issues?

Will this dataset include any data governed under the IRB or HIPAA, or data with other ethical issues?

- No
- Yes
- Uncertain

If yes or uncertain, please explain:

This question pertains to datasets that contain sensitive information and may have ethical and/or privacy issues that prevent the data about human subjects from being shared or distributed. If such conditions exist, please select yes. If you do not know whether your dataset has any ethical or privacy issues, please check uncertain and provide a short description of potential issues. GRIIDC staff will then follow up to help determine if privacy or ethical issues apply to your dataset.

Remarks

Remarks:

Please provide any additional information that would help GRIIDC understand the contents and format of your dataset that would assist in GRIIDC obtaining, storing, or distributing your dataset.