Section 1: What is a Dataset Information Form?
A Dataset Information Form or DIF is a planning tool required for datasets that will be submitted to the Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC) data management system. Currently, GRIIDC can only accept data from researchers who receive funding from organizations that have contracted GRIIDC to provide data management services. A DIF record (hereafter referred to simply as a DIF) helps researchers plan ahead and prepare for data collection and organization of data submission. Additionally, the details provided in the DIF assist GRIIDC in designing the data management system, determining storage capacity, and preparing to distribute your dataset. The GRIIDC team understands that the information provided in the DIF is preliminary and subject to change. Researchers should provide the most accurate and up-to-date information when dataset files are submitted to the GRIIDC data management system.

Section 2: User Account
In order to complete the dataset identification and submission process, you must first have an account with the GRIIDC data management system. To request an account please visit https://data.gulfresearchinitiative.org/account.

Figure 1: GRIIDC Account Request Page https://data.gulfresearchinitiative.org/account.

The account request page (Figure 1) will prompt you to enter the email address that is associated with your funding agreement. If you do not remember the email address or need to update your email address associated please contact GIRIIDC at griidc@gomri.org.
Once you have entered your email address, the screen will display a message verifying that an email has been sent to your email address (Figure 2).

Figure 2: Screen Verification showing that account request email has been sent to the email address provided.

The email you receive from GRIIDC will contain a link. Please follow the link to finish the account registration process.

Section 3: Datasets List
When you visit the GRIIDC DIF (https://data.gulfresearchinitiative.org/dif) you will see an interactive form on the left side of your screen and a list titled “Datasets” on the right side of your screen (Figure 3).

In the datasets list on the right side of the screen, you will see a list of 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., 🗂).

To facilitate work flow and communications one of three symbols, described below, will appear beside each dataset title within a folder.

- 🚫 Open: A Dataset Information Form has been started for this dataset, but has not yet been submitted to GRIIDC.

- ⚠️ Submitted: A Dataset Information Form has been submitted for this dataset and is in review by GRIIDC staff. The DIF is locked to prevent edits at this stage. If edits are required please contact griidc@gomri.org and the form will be unlocked to allow edits.

- ✔️ Approved: A Dataset Information Form has been approved for the dataset. The DIF is locked to prevent edits at this stage. To unlock an approved DIF please use the “Request to Unlock” button found at the bottom of the form. Once your DIF is approved, information should be updated when you submit your dataset. There is usually no need to unlock the DIF to provide updated information at this stage.
Figure 3: Dataset Information Form, example view. On the left side of the screen is the interactive Dataset Information Form and on the right side of the screen is a list of projects and datasets associated with the signed in user. 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. The dataset titled “Example GRIIDC DIF 1, September 2016” has a DIF that has been started but not yet submitted to GRIIDC for review and approval (red x). The dataset titled “Example GRIIDC DIF 2, September 2016” has a DIF that has been saved and submitted to GRIIDC for review and approval, but this DIF has not yet been approved (yellow triangle). For the dataset titled “Example GRIIDC DIF 3, September 2016” the DIF has been submitted and approved by GRIIDC staff (green check mark). To start a new DIF select the project the dataset is associated with from the drop-down list under the “Project Title” on the left side of the screen. To load an existing DIF, double click on the dataset title in the list on the right side of the screen under “Datasets”.

Section 4: Creating a New Dataset Information Form

To create a new DIF, please visit https://data.gulfresearchinitiative.org/dif. The left side of the screen is an interactive form. Field titles that are prefixed with a red asterisk (*) are required. To start filling out the form, select the project that the dataset is associated with from the drop-down list (Figure 3). 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. At any time during the creation of your DIF you can select “Reset Form”, found at the bottom of the form, to clear all the fields and start over.

You have two options to end your session:

1. **Save and Continue Later**: If you are interrupted while completing your DIF or need to collect additional information before you can complete your DIF, please select “Save and Continue Later”. When you sign in to the GRIIDC system again and visit the DIF page, the dataset title will appear in the list on the right side of your screen under the project you have selected for the dataset. You can then click on 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 are finished completing the DIF with the most complete and current information, select “Save and Submit to GRIIDC”. This will lock your DIF to prevent additional edits and your record will be submitted for review and approval by GRIIDC staff.
When a DIF record 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.

When you have submitted your DIF, you will receive an email confirmation that the DIF has been submitted. 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 all the information has been reviewed and determined to be acceptable, you will receive an email notification that your DIF has been approved. The DIF will be locked to prevent any changes after approval; however, if you wish to make changes to your DIF after it has been approved you can load the form and select “Request to Unlock” from the bottom of the form. Since the DIF is a planning document only, you are not required to update the information. Information should be updated when the dataset is submitted to GRIIDC. When the DIF is unlocked, the status will revert back to open and the icon will be reverted to a red x.

**Section 5: Editing a Submitted DIF**

The DIF is primarily a planning document, and information you provide in the DIF 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. The DIF is locked to prevent edits after submission; however, if you would like to update the information provided in the DIF before the DIF is approved you can email griidc@gomri.org to request that the DIF be unlocked; please reference the dataset UDI in your request.

If your DIF has been approved you can load the form and select “Request to Unlock” from the bottom of the form. You will receive an email notifying you when GRIIDC has unlocked the form. When the DIF is unlocked, the status will revert back to open and the icon will be reverted to a red x.

**Section 6: Fields and Examples**

**Project Title**

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 lists all projects that you are associated with in the GRIIDC.
system. If you have questions about what projects you are associated with, please contact griidc@gomri.org.

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 location where, and 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 2008

(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 2012

(4) Model/Numerical Model
Galveston Bay Circulation Study: Stanford Unstructured Nonhydrostatic Terrain following Adaptive Navier-Stokes (SUNTANS) models simulations for 2007-2011

(5) Sociology
Cross-sectional household survey response data to assess health and wellbeing of residents of coastal Louisiana, April 2012

(6) Genetics
Mitochondrial DNA control region sequences (297 base-pairs) from 140 northern Red Snapper (Lutjanus campechanus) collected from the Gulf of Mexico, 1998–2001

(7) Biochemistry
Polycyclic Aromatic Hydrocarbon concentrations in liver and muscle tissue from barracuda, escolar, and common dolphin fish, northeastern Gulf of Mexico, 2011-2012

Primary Point of Contact

*Primary Point of Contact:

[PLEASE SELECT TASK FIRST]

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, if the person’s name does not appear on the dropdown list, please contact GRIIDC (griidc@gomri.org).
Secondary Point of Contact

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. If the person’s name does not appear on the list, please contact GRIIDC (griidc@gomri.org).

Dataset Abstract

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. This may be similar to a publication abstract; however, it does not need to include details about conclusions drawn or statistical analyses completed using the dataset. It should address the following questions:

- What data will be or have been collected and/or generated?
- Why the data will be or have been collected and/or generated?
- How the data will be or have been collected and/or generated?
- When the data will be or have been collected and/or generated?
- Where the data will be or have been collected and/or generated?

Note that this field has a limit of 4,000 characters.

Field of Study

Datasets may be categorized based on the discipline(s) for which they are collected. Please check all the boxes that categorize the discipline(s) for which your dataset will be created for. 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

Please provide an estimate of the total size of the dataset files you expect will be created. This field is an estimate based on data files you have created in the past. 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.

**Phenomenon/Variables Observed or Generated**

Please expand on the information you have provided in the abstract by giving more exact details about all variables or phenomena that you plan to measure, collect, or generate. Please separate list items using a comma.

**Examples of Phenomena or Variables Observed/Generated:** 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**

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**

The estimated time period when the data is expected to be collected and generated based on the information provided in your GoMRI proposal. 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**

Please use the GRIIDC Spatial Extent Wizard to provide the approximate geographic area where the data will be collected or geographic area the data will be generated for or about, 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:**

All research projects funded by the Gulf of Mexico Research Initiative (GoMRI) are required to make data publicly accessible. 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 GoMRI research projects. 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?**

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 or not 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**

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.