

Dataset Information Form (DIF) Quick Reference Guide

The *Gulf of Mexico Research Initiative Information and Data Cooperative* (GRIIDC) tasked at developing the *Research Database* that facilitates the discovery and access of the scientific data expected to be collected and generated by GoMRI Scientists. To facilitate the design and deployment of services, GRIIDC is requiring all GoMRI scientists to complete the DIF.

User Identification

The forms require user identification and GRIIDC maintains a single-sign on system (*Central Authentication Service*). A user needs to enter their authentication parameters to access the forms (see Figure below). A link is provided (red marking) to formally request access.

The Form Panels and Color Codes

The form has two panels to facilitate the review of tasks and encoding of inputs. The left panel contains the fields to complete (see figure below). Note that a field prefixed with * (asterisk), indicates that the field is required. The right panel presents the list of tasks and datasets being edited, created, in review or has been submitted successfully. For more information in using this for, a PDF version of this document is also available for download (see A).

To facilitate the review of the forms, the system uses three colored balls to indicate the status. Red ball (●) is the default to mean that the dataset form has not been submitted (see B). A yellow ball (●) indicates that the form has been submitted for review and the green ball (●) indicates that the

form has been accepted and is now on file. The information on this form can be modified at any time. If the document has been locked for review or has been submitted, GRIIDC can unlock the record for editing.

The screenshot shows a web form interface. On the left, a light blue box contains a 'NOTE' about the 'Dataset Information Form' and a 'Download a User Guide' link with a PDF icon. A red arrow labeled 'A' points to this link. Below the note is the 'Dataset Information' section, which includes a 'NOTICE' and a 'Task Title' dropdown menu. A red arrow labeled 'B' points to a red dot in the top right corner of the 'Dataset Information' section. On the right, a sidebar titled 'Tasks and datasets for Xxxxxxxx Xxxxxxx' is visible.

Field Definition

Task Title

A close-up of the 'Task Title' field, which is a dropdown menu with a red asterisk indicating it is required. The text '[SELECT A TASK]' is visible in the dropdown. A red dot is located in the top right corner of the field.

This is a required field. The task titles will be provided as a drop-down list from where you choose from. If your task is not listed, please email GRIIDC (gomri.help@gomri.org) for assistance.

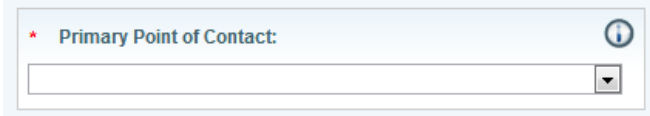
Dataset Title

A close-up of the 'Dataset Title' field, which is a text input box with a red asterisk indicating it is required. A red dot is located in the top right corner of the field.

This is a required field. It is in the discretion of the researcher to define the level of data aggregation to define a dataset. If this level of data aggregation has not been identified, it is recommended to start by answering the '*what, how, when, where*'. It is also not recommended to aggregate data too much that the data attributes can no longer be segregated and discoverable.

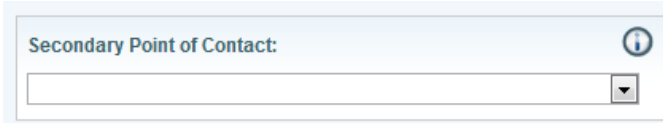
Example: Hydrodynamics: ADCP Data for June – July 2012 in Station 42001

Primary Point of Contact

A form field for the Primary Point of Contact. It has a light blue header with a red asterisk and the text "Primary Point of Contact:". Below the header is a white text input field with a small downward arrow on the right. An information icon (i in a circle) is in the top right corner.

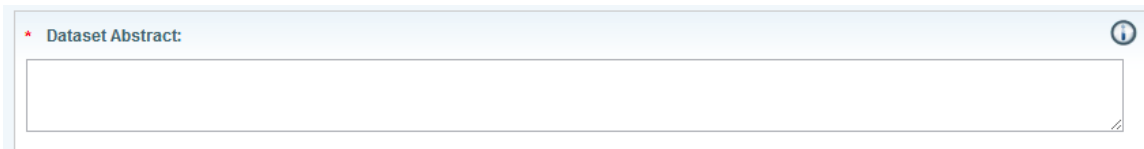
By default, the Principal Investigator (PI) is the *Primary Point of Contact* (pPOC). The pPOC is the person responsible for answering questions associated to this dataset. If the PI wants to designate a new person, a list is provided. However, if the name of the person is not on the list, contact GRIIDC (griidc.help@gomri.org) for assistance.

Secondary Point of Contact

A form field for the Secondary Point of Contact. It has a light blue header with the text "Secondary Point of Contact:". Below the header is a white text input field with a small downward arrow on the right. An information icon (i in a circle) is in the top right corner.

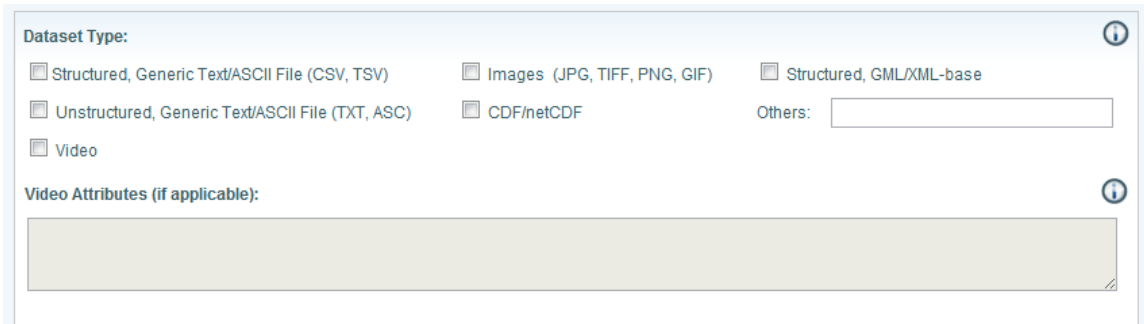
By default, the consortia director, if applicable, is the *Secondary Point of Contact* (sPOC). The sPOC will be contacted if the pPOC cannot be reached or unable to respond to queries regarding the dataset. As with the pPOC, a list is provided. However, if the name of the person is not on the list, contact GRIIDC (griidc.help@gomri.org) for assistance.

Dataset Abstract

A form field for the Dataset Abstract. It has a light blue header with a red asterisk and the text "Dataset Abstract:". Below the header is a large white text area with a small icon in the bottom right corner. An information icon (i in a circle) is in the top right corner.

This field should describe the rationale of collecting the dataset, procedure/process how this dataset will be created, period of data collection and what it will contain. Note that some of the fields that follow in this form are or may be components of this field.

Dataset Type

A form field for the Dataset Type. It has a light blue header with the text "Dataset Type:". Below the header are several checkboxes: "Structured, Generic Text/ASCII File (CSV, TSV)", "Images (JPG, TIFF, PNG, GIF)", "Structured, GML/XML-base", "Unstructured, Generic Text/ASCII File (TXT, ASC)", "CDF/netCDF", "Video", and "Others:". There is a text input field next to the "Others:" label. Below these is a section titled "Video Attributes (if applicable):" with a large white text area. Information icons (i in a circle) are in the top right and bottom right corners.

A dataset can be submitted in many different forms. Check the appropriate boxes that best describes the dataset that is expected to be submitted. If the type is not on the list, please provide a short description in the *Others* field.

If the file is a video, it is important to identify attribute of the dataset format (e.g. MPEG, AVI) and compression-decompression technology (e.g. H.264, DivX) to the community identify requirements from their end in using the dataset.

Dataset For:

Dataset For:

☐ Ecological/Biological

☐ Physical Oceanography

☐ Atmospheric

☐ Chemical

☐ Human health

☐ Social/Cultural/Political

☐ Economics

Others:

Datasets are often classified or created for a field of science. Choices are provided (checkboxes) and check all boxes that best classify the dataset to be created. If the field is not represented, please provide an input the *Others* field.

Approximate Dataset Size

*** Approximate Dataset Size:**

☐ < 1GB

☐ 1GB-10GB

☐ 10GB-200GB

☐ 200GB-1TB

☐ 1TB-5TB

☐ >5TB

Provide an approximation as to the size of the dataset expected to be created. This field will help GRIIDC formulate strategies in the access and/or collection of the dataset.

Phenomena/Variables Observed

Phenomenon/Variables Observed or Generated:

This is a free-text field and you may list as much as possible all the variables or phenomena that will be measured, collected or generated. Separate the items in the list using a comma.

Example: wind direction, wind speed, sea surface temperature

Procedure to Acquiring the Data

Procedure to Aquiring or Creating the Data:

☐ Field Sampling

☐ Simulated/Generated

☐ Laboratory

☐ Literature Based

☐ Remote Sensing

Others:

This filed should contain the procedure or method used to collect or generate the data in the dataset. Depending on the dataset being created, list all of the procedures expected to be used in acquiring the data.

Data Sampling/Generation Period

Data Sampling/Generation Period:

Start Date

Jan. ▾

2012 ▾

to

Jan. ▾

2012 ▾

End Date

This is the approximate period that the data is expected to be collected as per the proposal. If these dates need to be modified, email GRIIDC (griidc.help@gomri.org) to unlock the file for modification.

Geographic/Study Area

Geographic/Study Area:

This is the relative geographical zone that the data will be collected or will be generated for. Although the form will allow descriptive inputs, the use of a point location in latitude and longitude (in decimal degrees) pair or a series to create a closed polygon are preferred. You may use online interactive maps for these inputs (e.g. <http://www.birdtheme.org/useful/googletool.html>) to get these coordinates.

Example of a descriptive input: Coastal waters close to New Orleans

Example of a polygon: -91.252441,30.600094,-88.417969,30.675715,-88.308105,29.745302,-89.626465,28.574874,-91.538086,29.554345

References to Referenced Historical Data (if applicable)

References to Historical Data to Use (if applicable):

If applicable, all referenced data should be listed using standard bibliographic referencing format. Using a *Digital Object Identifier* (DOI) is preferred but not required. The hyperlinks (URL) to the online data source (if applicable) are acceptable inputs.

Metadata Editor to Use

Metadata Editor to Use:

All dataset that will be submitted will require a metadata. List the metadata editors that will be use or planned to be used for the creation of the metadata. An option to indicate that the project has not determined the editor is also available.

Example: ESRI ArcCatalog, NCDDC MERMAid

Metadata Standards to Use

Metadata Standards to Use: ⓘ

☐ ISO 19115

☐ FGDC-CSDGM

☐ Dublin/Darwin Core

☐ Ecological Metadata Language (EML)

Other:

This is an optional field but if the metadata standard to use has been identified, check all that apply. If an unlisted format/framework will be used, please identify in the *Others* field.

Data Access Points

Data Access Point: ⓘ

☐ File Transfer Protocol (FTP)

☐ THREDDS Data Server (TDS)

☐ Environmental Research Division's Data Access Program (ERDDAP)

Other:

It is very important that GRIIDC is aware on how the data will be made accessible to the world (when it is ready). This field answers the question, 'How will you make these data available?' Options are provided based on the proposal submitted. Check all that apply. If the data transfer protocol is not on the list, please identify in the *Others* field.

National Data Center(s) that the Dataset will be Submitted to

National Data Center(s) that the Dataset will be Submitted to: ⓘ

☐ National Oceanographic Data Center (<http://www.nodc.noaa.gov>)

☐ US EPA Storet (<http://www.epa.gov/storet/wqx>)

☐ Global Biodiversity Information Facility (<http://www.gbif.org>)

☐ National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>)

☐ Data.gov Dataset Management System (<http://www.dms.data.gov>)

☐ Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC) (<https://griidc.gomri.org>)

Others:

As part of the contractual obligation of all projects of GoMRI, this dataset needs to be submitted to a recognized national data archival center. GRIIDC is recognized as an acceptable substitute to national centers if none can be identified. If another center can be identified not on the list, identify the center and its URL in the *Others* field.

Ethical or Privacy Issues

Ethical or Privacy issues:

☐ No
☐ Yes
☐ Uncertain

If yes or uncertain, please explain:

Some of the data that will be collected may carry ethical and/or privacy issues preventing it to be shared or distributed. If such condition exist, check the 'yes' check box if certain that such issue exist but if uncertain, check the 'uncertain' box, and please provide a short description of the issue.

Remarks

Remarks:

Provide general remarks that will help GRIIDC prepare itself for the downloading (if needed), use or distribution of the dataset being identified.

Submission and Completion

Two buttons are available at the bottom of the form: *Save and Continue Later* and *Submit and Done*. The first button allows users to save their work and to continue later. Once a form has been submitted, the data can no longer be edited. The data will be forwarded to GRIIDC for review of the inputs. Users will be notified if the form has been accepted and no further inputs are expected. Users will also be notified if error or problem with the inputs is detected. In which case, the status of the dataset will be demoted (showing red button again) to allow the user to modify the dataset and correct the problem.

For more information, please contact GRIIDC at griidc.info@gomri.org