

GRIIDC Dataset Submission User Guide

Note: Before a dataset can be submitted to the GRIIDC data management system, it must first be identified through a Dataset Information Form (DIF). To complete and submit a DIF for your dataset please visit <https://data.gulfresearchinitiative.org/dif>. You will be able to submit your dataset file after you receive notification that your DIF has been reviewed and approved by GRIIDC staff.

Contents

<i>GRIIDC Dataset Submission User Guide.....</i>	<i>1</i>
<i>Section 1: What is a Dataset Submission?</i>	<i>1</i>
<i>Section 2: What Do I Need Before I Can Submit a Data File?</i>	<i>2</i>
<i>1. User Account</i>	<i>2</i>
<i>2. Dataset Information Form (DIF).....</i>	<i>2</i>
<i>Section 3: Dataset Submission – Introduction.....</i>	<i>2</i>
<i>Section 4: Dataset Submission – Field Descriptions and Examples</i>	<i>2</i>
<i>Dataset Contact.....</i>	<i>2</i>
<i>Dataset Information</i>	<i>3</i>
<i>Keywords</i>	<i>5</i>
<i>Data Extent</i>	<i>5</i>
<i>Section 5: Dataset Submission – Dataset File Transfer Details</i>	<i>6</i>
<i>1) Upload.....</i>	<i>6</i>
<i>2) Enter URL.....</i>	<i>7</i>
<i>Restrictions</i>	<i>8</i>
<i>Finish your Submission.....</i>	<i>8</i>
<i>What’s Next?.....</i>	<i>8</i>
<i>Appendix A: Recommendations: Dataset File Names.....</i>	<i>8</i>

Section 1: What is a Dataset Submission?

GRIIDC has developed a data management system to store datasets and related information collected and generated by researchers. Datasets can be submitted to the GRIIDC data management system directly or if housed at a National Data Archive (e.g., National Centers for Environmental Information) a link to the dataset may be provided. Dataset submission is the process that researchers use to submit data directly to the GRIIDC data management system or to notify GRIIDC that a dataset is available through a National Data Archive.

Section 2: What Do I Need Before I Can Submit a Data File?

1. 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/pelagos-symfony/account>. If you have a GRIIDC account, please visit the GRIIDC website (<https://data.gulfresearchinitiative.org/>) and click the login link in the upper right corner of the screen to open the log in form. Enter the username and password for your account.

2. Dataset Information Form (DIF)

Data providers must first have an approved Dataset Information Form (DIF) for a dataset before it can be submitted to GRIIDC. In order to complete and submit a DIF for your dataset please visit <https://data.gulfresearchinitiative.org/dif>.

Section 3: Dataset Submission – Introduction

After a data provider has created an account and received email notification that a Dataset Information Form (DIF) has been approved, they can continue to Dataset Submission:

<https://data.gulfresearchinitiative.org/dataset-submission>.

At the top of the Dataset Submission start page in the text box below “Unique Dataset Identifier (UDI)”, enter the UDI for the approved Dataset Information Form (DIF) that you will be submitting data for, and select “Load Dataset”. Alternatively, you can search for and select your dataset from the list of approved Dataset Information Forms (DIFs) displayed below the UDI field. This will pre-fill a number of fields including Dataset Contact, Dataset Title, Dataset Abstract, Supplemental Information-Data Parameters and Units, and Data Extent with information provided when you completed your DIF. Please update these fields as appropriate. At any time you can select the Save and Continue later button if you are not ready to submit your data. For more details about what information should be provided in the fields, hover over the information  icon on the webpage.

Section 4: Dataset Submission – Field Descriptions and Examples

The Dataset Submission page consists of four tabs of information. Each tab focuses on a different component of descriptive information about the dataset. Required fields are indicated with a red asterisk. A data file cannot be submitted until all the required fields are entered. A red X will be displayed on a tab if required fields have not been filled out. When all required information has been entered for a tab, a green check will appear on the tab. Once all required information has been entered for every tab, indicated by green checks on each tab, a data file can be submitted.

Dataset Contact

The Dataset Contact tab collects information about the individual who is responsible for answering questions about the dataset. Usually this is the primary investigator of the project.

Primary Point of Contact*: (Required) The person you designated as the primary point of contact in the DIF will be entered into this field. You may select another person associated with your project using the drop-down list. If a member of your research group or project is not listed in the drop-down list or if the

contact information displayed is incorrect, please contact griidc@gomri.org. You may also select additional points of contact by selecting the “Add Contact” button.

Role*: (Required) This field is the person’s relationship to the dataset. There are three options: Point of Contact, Principal Investigator, and Author. Point of contact is the person who can be contacted for acquiring information about the data or to acquire the data. Principal investigator is the key person responsible for gathering information and conducting research. Author is the person who authored the dataset.

Dataset Author(s)*: (Required) This field is a list of authors who should be acknowledged if these data are cited in published materials. Please use the Modern Language Association (MLA) style, demonstrated below, when listing dataset authors.

Example Dataset Author(s)

One originator: Last Name, First Name (e.g. Smith, Joe)

Two originators: Cross, Susan and Christine Hoffman

Three originators: Lowi, Theodore, Benjamin Ginsberg, and Steve Jackson

More than three originators: Gilman, Susan *et al.*

Dataset Information

The Dataset Information tab collects information that allows others to understand your dataset.

Dataset Title* (Required): Please provide a descriptive title that briefly explains the contents of your dataset and, if applicable, dates and geographic area. It should be understandable by a user unfamiliar with your methods, collection sites, or research platforms.

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 2020

(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 5 - July 24 2020

(4) Model/Numerical Model

Galveston Bay Circulation Study: Stanford Unstructured Nonhydrostatic Terrain following Adaptive Navier-Stokes (SUNTANS) models simulations for 2020-2021

(5) Sociology

Cross-sectional household survey response data to assess health and wellbeing of residents of coastal Louisiana, April 2020

(6) Genetics

Mitochondrial DNA control region sequences (297 base-pairs) from 140 northern Red Snapper (*Lutjanus campechanus*) collected from the Gulf of Mexico, 2018–2020

(7) Biochemistry

Polycyclic Aromatic Hydrocarbon concentrations in liver and muscle tissue from barracuda, escolar, and common dolphin fish, northeastern Gulf of Mexico, 2020-2021

Short Title: This is an alternative short name or other language name by which the dataset might be known. This could be the title that the lab references the dataset by and could include abbreviations or cruise and platform names.

Dataset Abstract*: (Required) A narrative summary of the dataset's contents. The dataset abstract should summarize what data the dataset contains, methods used to collect or generate the dataset, and time period and location of data collection. This may be similar to a publication abstract; however, it does not need to include details about results, conclusions, or statistical analysis completed using the dataset. It should address the questions:

- What data have been collected and/or generated?
- How have the data been collected and/or generated?
- When were the data collected and/or generated?
- Where were the data collected and/or generated? If model data, the location for which the dataset was generated.

If your dataset relates to another dataset you have provided to GRIIDC, please provide the DOI of the related dataset in the abstract. Additionally, if you would like to reference a publication that uses the dataset, please include the full reference of the publication in the abstract.

Purpose* (Required): Summary of the reasons or intentions for which the dataset was created or generated.

Supplemental Information – Data Parameters and Units* (Required): Please provide descriptions of all reported data parameters/variables, list and define abbreviations for each parameter/variable, and define units of measurement for each parameter. If you provide multiple files that report different parameters/variables, indicate which files report which parameters/variables.

Supplemental Information – Methods* (Required): Please provide a description of the methods used to create and/or generate the data in the dataset.

Supplemental Information – Instruments: Please provide a description of the instruments and equipment used to create and/or generate the data in the dataset.

Supplemental Information – Sampling Scales and Rates: Please provide a description of the spatial and temporal scales and rates that were used to collect and/or generate the data, if applicable.

Supplemental Information – Error Analysis: Please provide a description of any error or uncertainty analysis completed on the final data and the results of the error analysis.

Supplemental Information – Provenance and Historical References: If existing historical data were used as part of the dataset, please provide a description of the historical data used, including who created the original dataset (person and/or organization) and from where the historical data can be obtained.

Keywords

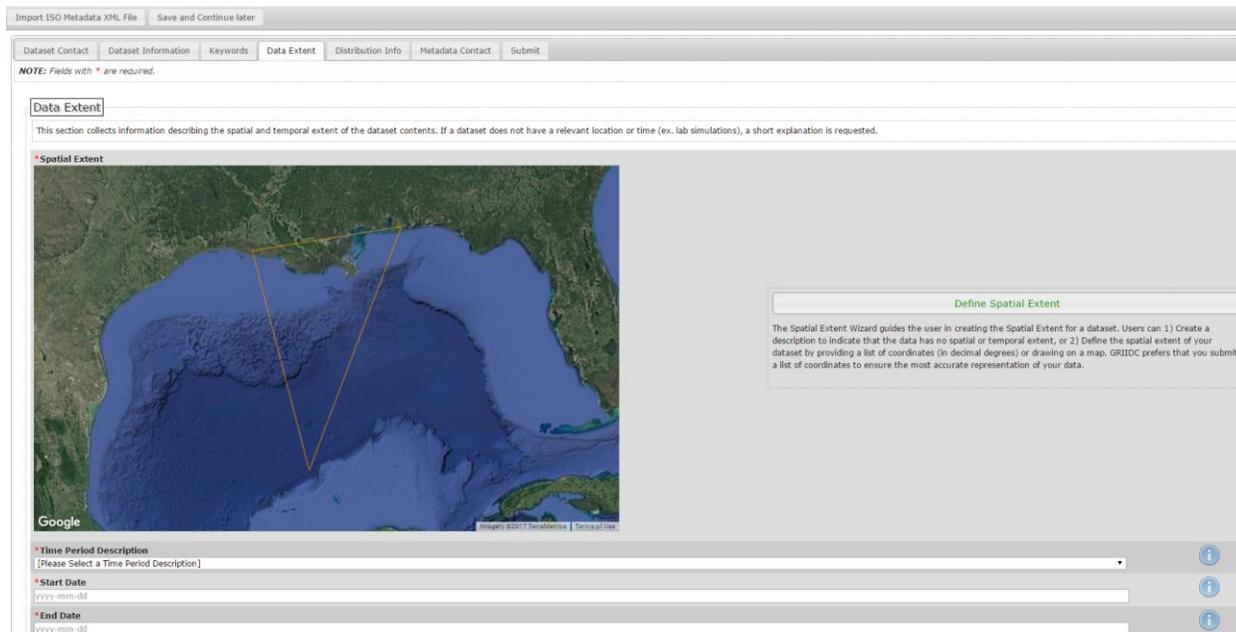
Theme Keywords* (Required): Please include commonly used words or short phrases that describe the themes or subjects related to the dataset. These may be the keywords used to describe the publication associated with the dataset. Do not include keywords that relate to the location or geography of the data.

Place Keywords* (Required): Please include commonly used words or short phrases that describe the geographic areas, locations, or places that are associated with the data. Leave blank if the dataset does not have a relevant place keyword (e.g. lab mesocosms).

Topic Category Keyword* (Required): These keywords are pre-defined by the ISO 19115-2 metadata standard that GRIIDC uses to describe datasets. Please select all the relevant topic categories from the keyword list. Hover over the keyword to see a definition.

Data Extent

All fields in this tab are required. This tab collects information about the spatial and temporal extent of your data (Figure 1).



The screenshot shows a web-based metadata form with several tabs: 'Dataset Contact', 'Dataset Information', 'Keywords', 'Data Extent', 'Distribution Info', 'Metadata Contact', and 'Submit'. The 'Data Extent' tab is active. A note states: 'NOTE: Fields with * are required.' Below this, the 'Data Extent' section is titled and includes a sub-note: 'This section collects information describing the spatial and temporal extent of the dataset contents. If a dataset does not have a relevant location or time (ex. lab simulations), a short explanation is requested.' The main content area features a map of the North Atlantic Ocean with a yellow polygon defining a spatial extent. To the right of the map is a 'Define Spatial Extent' button and a text box explaining the Spatial Extent Wizard. Below the map are three required fields: 'Time Period Description' (with a dropdown menu), 'Start Date' (with a date input field), and 'End Date' (with a date input field). Information icons are present on the right side of the form.

Figure 1: Data Extent tab displaying a spatial extent

Define Spatial Extent* (Required): Please use the GRIIDC Spatial Extent Wizard to provide the most accurate 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, please describe this using the

Spatial Extent Wizard. Please see the Spatial Extent Wizard User Guide for more information on use of the Wizard.

Time Period Description *: (Required) This is a description of what the start and end date you will provide below represents. This field has three standard values:

1. Ground condition: the data represents the actual condition of things on the ground during the time period specified. Samples or data were collected during the time period.
2. Modeled period: the data represents simulated conditions during the time period (could be future modeled period or past modeled period).
3. Ground condition and modeled period: the dataset includes data that represents the actual condition of things on the ground during the time period specified and simulated conditions during the time period.

Start Date* (Required): The beginning date when the data were collected. Alternatively, for a model the earliest date for which the data were generated.

End Date* (Required): The end date when the data were collected. Alternatively, for a model the end date for which the data were generated.

Section 5: Dataset Submission – Submitting Data Files

Once all required fields on the Dataset Submission page are completed with the most accurate and up-to-date information, you can select the Submit tab to submit your dataset file to GRIIDC. You will be unable to select the Submit tab if the Dataset Submission is missing required fields; however, your work will be saved and you can continue your Dataset Submission at a later time if you do not have all of the required information or the dataset file. Currently, there are two methods of providing data using the dataset submission interface: uploading data from your local computer or entering a URL for the data located at a National Data Archive. If your dataset can not be provided using the upload method due to the size of the dataset, please contact griidc@gomri.org for more information about alternative submission methods.

1) Upload

This is the simplest method for submission of dataset files to GRIIDC. You may upload a single or multiple files of any size using direct upload. Under the “Upload” tab, select the “Upload” drop-down to either upload a file or folder from your local computer (Figure 2). You may also drag and drop files and/or folders in the dialog box. Files and folders can be deleted, moved, or renamed using the toolbar button or by right clicking the file or folder.

HELPFUL TIP: Please use consistent and descriptive names when naming files. For more information about dataset file name recommendations please see [Appendix A](#).

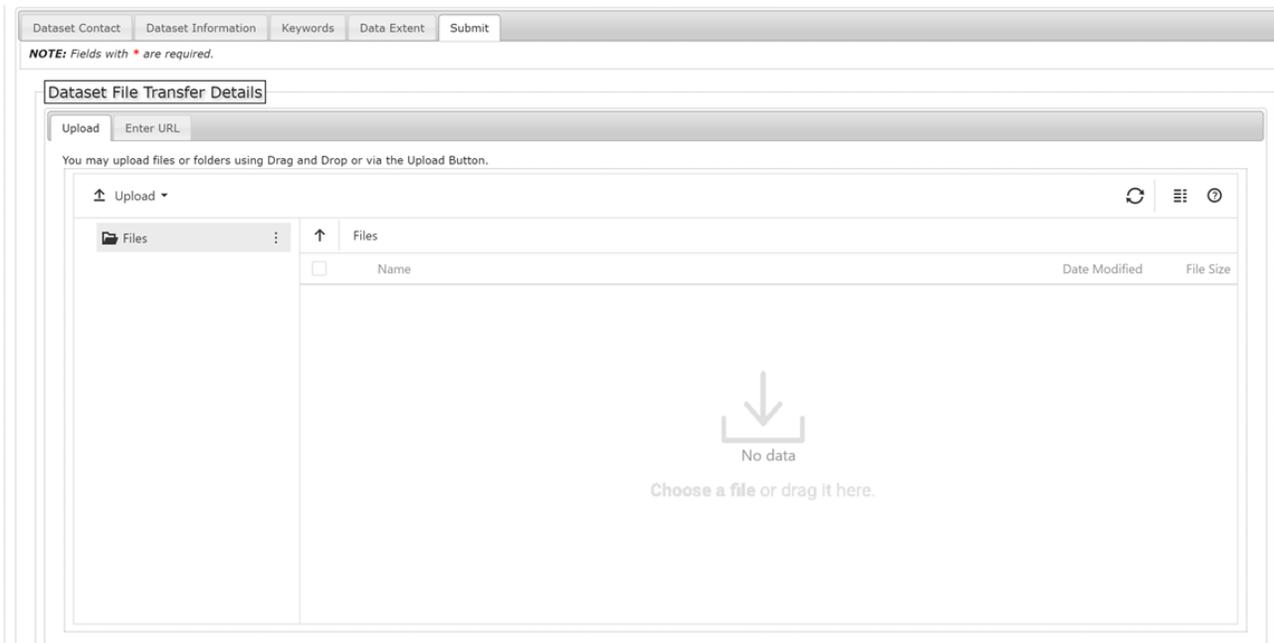


Figure 2: Submit tab: Dataset File Transfer Details, Upload

2) Enter URL

If your data have been submitted to a National Data Archive, you may provide the DOI URL for the dataset at the archive instead of providing the data file. If you do not know if the archive is considered a National Data Archive, please contact griidc@gomri.org.

Under the Submit tab, select the “Enter URL” tab under “Dataset File Transfer Details” (Figure 3). Please provide the web address (URL) of the DOI of the data located at the archive. If the data do not have a DOI, please provide the URL to the home page or data landing page to your dataset. If you are providing a link to a National Data Archive, your data should be publicly accessible through the archive. By providing a link to a dataset available at a National Data Archive your data will be catalogued in the GRIIDC system.

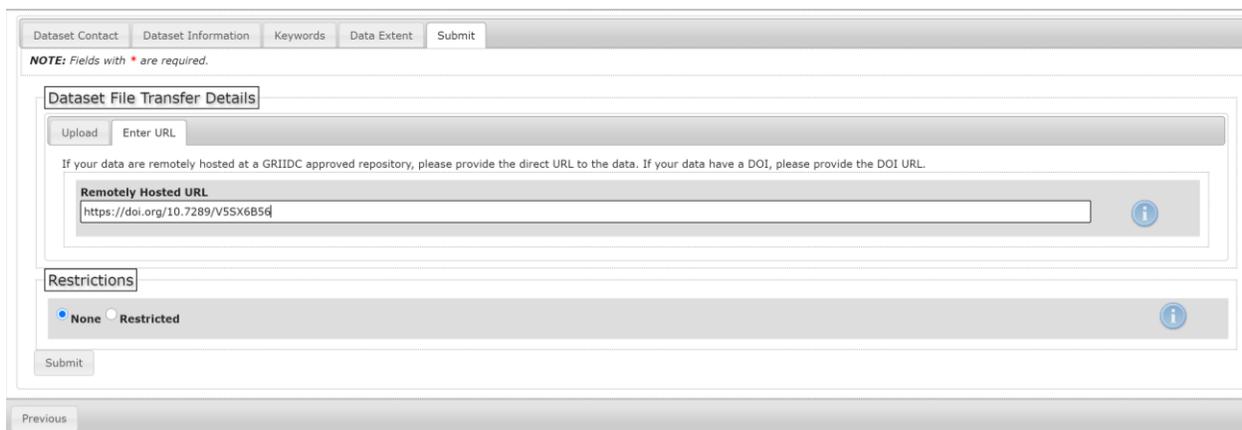


Figure 3: Submit tab: Dataset File Transfer Details, Enter URL

Restrictions

The default selection for restrictions is “none”, indicating the dataset file is available for the general public to download after GRIIDC has performed a dataset package review. GRIIDC does allow short term restrictions to be placed on datasets. If you select “Restricted” your dataset file will not be available for download after dataset package review. It is your responsibility to make sure your data are not restricted and publicly available in compliance with the terms of your grant award.

Finish your Submission

After you have uploaded your data either by upload or providing a URL to the data at a National Data Archive, you can select the Submit button at the bottom of the page to finish the dataset submission process. You will be taken to a confirmation page with your dataset’s information and you will also receive a confirmation email that you have submitted a dataset to GRIIDC.

What’s Next?

Your submission will be reviewed by the GRIIDC team. The Dataset Package Review process includes:

- Making sure all files can be opened
- Making sure no empty files or folders are included in the dataset
- Making sure the dataset does not contain published materials or other materials that may be subject to copyright
- Making sure the descriptive information fully and accurately describes the dataset files provided

You may be contacted during the Dataset Package Review process to answer questions related to the dataset file or dataset description. Please respond to requests promptly. Datasets are not made publicly available until GRIIDC has reviewed and accepted the submission.

Appendix A: Recommendations: Dataset File Names

When a user downloads your dataset file from the GRIIDC website, the dataset will be provided with the file name with which it was submitted. GRIIDC does not change dataset file names or names of individual files within datasets. Therefore, GRIIDC recommends using the following guidelines to name your dataset file; if your dataset includes multiple files these guidelines should be followed for all files within your dataset.

- Use descriptive names that indicate what the file contains
- Use short names (less than 50 characters)
- Use simple names that are easy to understand
- Use alphanumeric characters
- Use underscores (_) or dashes (-) rather than spaces
- Avoid special characters such as: \ ' = / , < > ^ ; : () # * ? % , " @ ! + { } ~ ` []
- Avoid using internal project codes or acronyms that individuals outside of your laboratory or research group would not understand
- Incorporating the temporal or spatial information when applicable is recommended

You may wish to include the UDI assigned to your dataset in the file name; however, please change the colon (:) to a dash (-). Table 1, below, provides examples of both good and bad dataset file names.

Table 1: Example Dataset File Names

	File Name	Issues
Bad example	#1.xls	- Undescriptive - Special Characters
Bad example	gomri data 12.zip	- Undescriptive - Spaces
Bad example	TLF5682.tar	- Undescriptive - Uses internal project code
Bad example	Sam37.pdf	Undescriptive Uses internal project name
Bad example	Dataset publication for the Journal of Examples volume 1 page 32 2012/05/06	Too long Special characters Spaces
Good example	PAHConcentrations_ClamTissue2020.csv	None
Good example	Smith_CommunitySurveyDataset_2020.txt	None
Good example	BacterialProduction_MobileBay_R6.x807.000.0002.zip	None
Good example	Johnson_BubblePlumeVelocity_Lab_Images.tar	None