

# GRIIDC Metadata Editor Guidance

## Contents

Section 1: What is the GRIIDC ISO 19115-2 Metadata Editor? .....	1
Section 2: Methods to Use the Metadata Editor .....	2
Section 3: Metadata Editor Fields .....	3
Tab 1: Dataset Contact.....	3
Tab 2: Dataset Information .....	4
Tab 3: Keywords.....	5
Tab 4: Data Extent.....	6
Tab 5: Distribution Information .....	7
Tab 6: Distribution Contact .....	7
Tab 7: Metadata Contact .....	8
Section 4: Saving your Metadata File.....	9
Appendix A: Example Metadata .....	10
Tab 1: Dataset Contact Examples .....	10
Tab 2: Dataset Information Examples.....	10
Tab 3: Keyword Examples .....	18
Tab 4: Data Extent Examples .....	19
Tab 5: Distribution Info Examples.....	21
Tab 6: Distribution Contact Examples* .....	21
Tab 7: Metadata Contact Examples* .....	22

## Section 1: What is the GRIIDC ISO 19115-2 Metadata Editor?

The GRIIDC Metadata Editor is a web-based tool that creates ISO 19115-2 compliant metadata files in XML file format. This tool is presented as a series of seven interactive forms that the user fills out. Once a user has filled in all the required information, the metadata can be saved as an XML file on their computer. No knowledge of XML is needed to create metadata using the interactive forms. It is important to note that the GRIIDC ISO 19115-2 Metadata Editor does not store any information outside of the XML file saved on the user’s local computer, so please download the file to your computer to save your work.

### What is metadata?

Metadata, or data about data, is information that describes the contents and context of dataset files. Its main purpose is to help people who use data find the data they need and determine how best to use the data. However, metadata is also used to support data management, archiving, and preservation.

### What is ISO 19115-2?

ISO 19115-2 is a metadata standard developed and adopted by the International Standards Organization (ISO) that defines how to document and describe information. Generally, the standard defines what information is to be included in the metadata and how it should be structured. GRIIDC has chosen to adopt the ISO 19115-2 metadata standard because it has wide acceptance and is used by many data repositories, including national data centers.

Note: For those individuals with expert knowledge of XML, GRIIDC uses the ISO 19115-2 schema published by NOAA’s National Geophysical Data Center (NGDC), please contact GRIIDC for details.

### What is XML?

Extensible Markup Language (XML) is a format for storing and transporting information in a structured way that is readable by both people and computers. XML uses ‘tags’ and is roughly similar to HTML; an example can be seen in Figure 1. ISO 19115-2 metadata is in XML format and the GRIIDC Metadata Editor only creates and accepts ISO 19115-2 XML metadata files. However, the GRIIDC ISO 19115-2 Metadata Editor (<https://data.gulfresearchinitiative.org/metadata-editor-start>) allows users to create, edit, and submit metadata files without ever needing to read and understand XML files.

```
<gmd:phone>
  <gmd:CI_Telephone>
    <gmd:voice>
      <gco:CharacterString>+1-361-825-3604</gco:CharacterString>
    </gmd:voice>
    <gmd:facsimile>
      <gco:CharacterString>+1-361-825-2050</gco:CharacterString>
    </gmd:facsimile>
  </gmd:CI_Telephone>
</gmd:phone>
```

Figure 1: Example of XML. In this example XML is being used to represent information for phone and fax numbers.

## Section 2: Methods to Use the Metadata Editor

### Ways to Use The Metadata Editor

(start with a blank form)

(load a file from your computer)

UDI:

URL:

Figure 2: GRIIDC ISO 19115-2 Metadata Editor. Methods listed to use the metadata editor found on the home page.

### 1. Start from Scratch

If you do not have a dataset identified or registered within the GRIIDC data management system, you can start from scratch. This option will load a blank form. This is ideal for using the Metadata Editor to create documentation for any project.

### 2. Load from File

If you have an existing ISO 19115-2 XML metadata file on your computer you can upload the file into the metadata editor using this method. This method is useful if you are continuing to work on or updating an existing metadata file.

When you select load from file, a popup will be displayed that allows you to select the file from your computer. Select the file name, and then select “open”. If your file is in the correct format, the ISO 19115-2 schema, the metadata editor will load with the information that is in the XML file.

### 3. Load from Registered Dataset

If your dataset has been identified in GRIIDC, you can load the metadata editor with the most recent information for the dataset by using the UDI. The UDI is the string of letters and numbers first assigned to your dataset when you submitted a Dataset Information Form. An example of an UDI is R4.x257.000:0001.

Type the UDI for the dataset that you are creating metadata for into the box beside “Load from Submitted Dataset” and select “Load from Submitted Dataset”. This will load the Metadata Editor with a number of fields pre-filled based on the most recent information GRIIDC has for the dataset.

### 4. Load from External URL

If your metadata file is available online, you can load the Metadata Editor by providing the URL for the XML metadata file. The URL must point directly to the file and the XML file must adhere to the ISO 19115-2 metadata schema.

Type or copy and paste the URL for the website into the field beside “Load from External URL and select “Load from External URL”. Your metadata form will be loaded with the information from the metadata file that is available at the URL.

## Section 3: Metadata Editor Fields

The GRIIDC Metadata Editor is presented in seven tabs. Each tab focuses on a different component of the metadata. Some fields are required by the GRIIDC ISO 19115-2 Metadata Editor. These fields are marked with a red asterisk beside the field title (\*), both in the form and in the descriptions below. Examples of the type of information that should be provided for each field can be found in Appendix A to this user guide.

### Tab 1: Dataset Contact

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

**Table 1: Dataset Contact Field Descriptions**

<b>Dataset Contact Field</b>	<b>Description</b>
Name*	First and last name of the individual who should be contacted if there are questions related to the dataset. Often this is the primary investigator.
Organization*	The primary organization with which the dataset contact is affiliated. Usually this is the University and Department where the dataset contact works on a daily basis.
Position*	The position that the dataset contact holds at their organization.
Role*	The individual's relationship to the dataset.
Telephone Number*	The dataset contact's telephone number at their organization, including the country code.
Fax Number	Fax number for the dataset contact or their organization.
Address*	Address line for the organization with which the dataset contact is affiliated. This field can accept multiple lines of text.
City*	The city that is associated with the address above.
State/Province/Region*	The state, province, or administrative area associated with the address.
Zip/Postal Code*	The administrative spatial code that is used in mail delivery.
Country*	The country of the organization's address.
Email Address*	Email address of the dataset contact.

**Tab 2: Dataset Information**

This tab collects information that allows others to understand your dataset.

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

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

**Date\***: (Required) The date the dataset was created, published, or revised. Usually the date the dataset was published in the GRIIDC data management system, using the current date. The format should be YYYY-MM-DD.

**Date Type\*** (Required) The date type is a description of the date provided and refers to the date the dataset was created, published, or revised.

**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 period and location of data collection. This may be similar to a publication abstract; however, it does not need to include details about conclusions drawn or statistical analysis completed using the dataset. It should address the questions:

- What data have been collected and/or generated?
- How were data collected and/or generated?
- When were data collected and/or generated?

- Where were 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 this information in the abstract. Additionally, if you would like to reference a publication that uses the dataset, please include this information 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/variable. If you provide multiple files that report different parameters/variables, indicate which files report which parameters/variables.

**Supplemental Information – Methods**: 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 was 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 where the historical data can be obtained from.

### **Tab 3: 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**: Please include commonly used words or short phrases that describe the geographic areas, locations or places that are associated with the data, if applicable. Datasets generated in the laboratory, for example through mesocosms, may not have any relevant place keywords.

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

**Tab 4: Data Extent\***

All fields in this tab are required. This tab collects information about the spatial and temporal extent of your data (Figure 3).

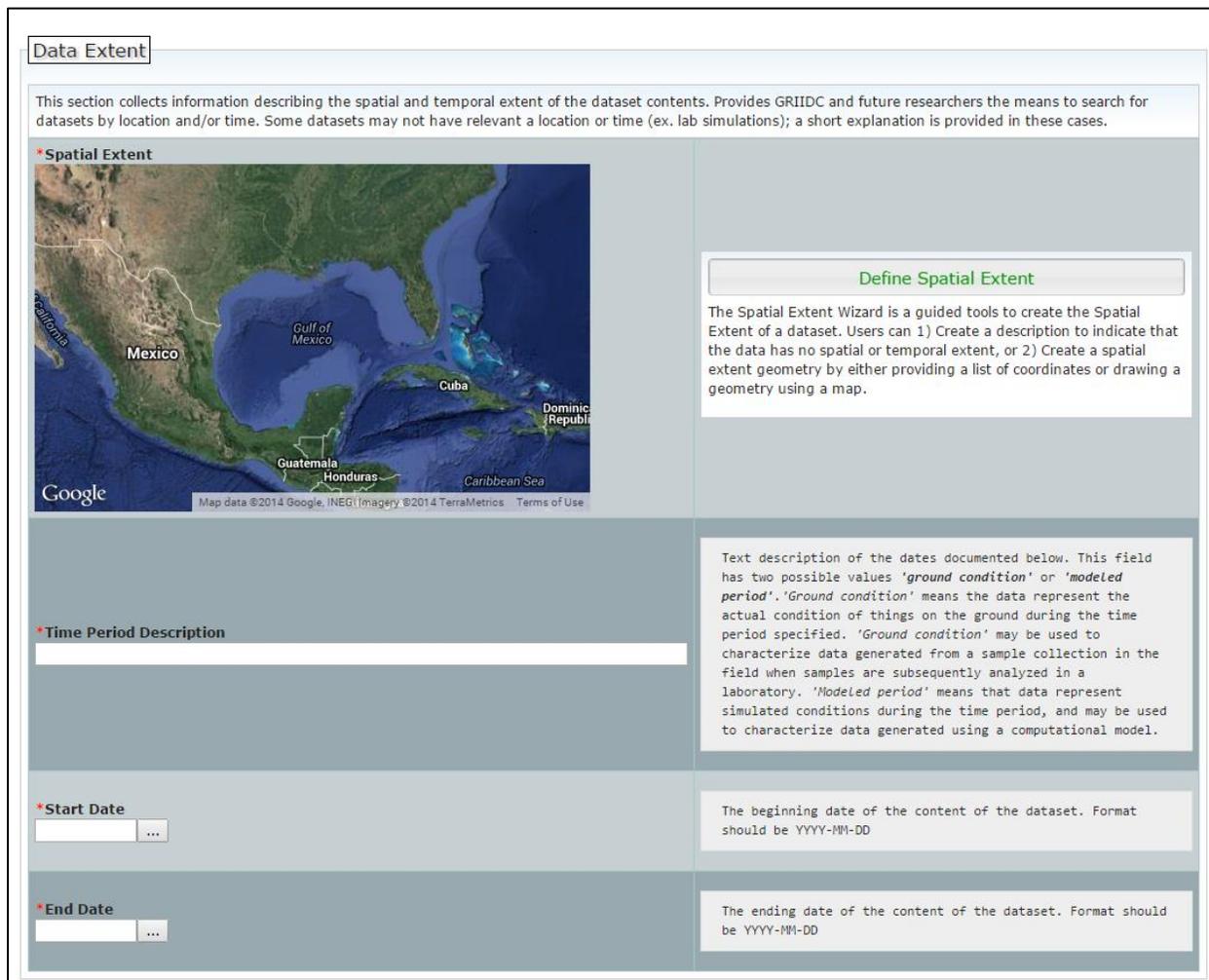


Figure 3: Data Extent Tab in GRIIDC ISO 19115-2 Metadata Editor.

**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 dates represent. 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 ending date when the data were collected. Alternatively, for a model the ending date for which the data were generated.

### **Tab 5: Distribution Information**

This tab collects information about the format and method that will be used to submit or register your dataset with GRIIDC.

#### **Format**

**Data Transfer Format:** The file types (extensions) of all the files included as part of your dataset. For common formats, you can use the extension and/or common name. If the format is less common or ambiguous, please identify the software program needed to read the data.

**File Decompression Technique:** If multiple files were packaged together to create a single dataset file, please provide the method used to package these files together. Common examples include zip, gz and tar. Please see GRIIDC Compression Guidance documents for Linux, Windows, and Mac.

#### **Transfer Options**

**Transfer Size\*:** (Required) The estimated size of the dataset file(s) expressed in megabytes. If multiple files are packaged together, this should be the overall package size.

**URL\*:** (Required) The URL where your dataset can be obtained from. If your dataset is registered with GRIIDC, the URL should be the data landing page URL - <http://data.gulfresearchinitiative.org/data/<UDI>>, where UDI is the GRIIDC UDI assigned to your dataset when you submitted a Dataset Information Form. (For example, <https://data.gulfresearchinitiative.org/data/R1.x139.144:0002/>).

If your dataset is available through a national data archive, please provide the URL where your data can be downloaded here.

### **Tab 6: Distribution Contact**

This tab collects information about the organization that is responsible for distributing your dataset.

If your dataset is, or will be available through the GRIIDC website, please left click to check the box beside “This data is hosted with GRIIDC”. This will fill in all the fields with the correct GRIIDC contact information.

If another organization is responsible for distributing your publicly available dataset, please provide the contact information for that organization. Field descriptions can be found in Table 2 below.

**Table 2: Distribution Contact Information Field Descriptions.**

<b>Distribution Contact Field</b>	<b>Description</b>
Organization*	The organization responsible for distributing the dataset.
Telephone Number*	The organization's telephone number, including the country code.
Fax Number	Fax number for the organization.
Address*	Address line for the organization, this field can accept multiple lines of input.
City*	The city that is associated with the address above.
State/Province/Region*	The state, province, or administrative area associated with the address.
Zip/Postal Code*	The administrative spatial code that is used in mail delivery.
Country*	The country of the organization's address.
Email Address*	Email address of the organization.
URL*	URL of the organization's website. Should be formatted starting with http:// or https://

**Tab 7: Metadata Contact**

This tab collects information about the person who is responsible for maintaining and answering questions about the dataset description (metadata) information.

**Table 3: Metadata Contact Information Field Descriptions.**

<b>Metadata Contact Field</b>	<b>Description</b>
Name*	First and last name of the individual who is responsible for answering questions about and maintaining the metadata. Often this is the scientist or researcher who created the dataset.
Organization*	The primary organization with which the metadata contact is affiliated. Usually this is the University and Department where this person works on a daily basis.
Position*	The position that the metadata contact holds at their organization.
Telephone Number*	The metadata contact's telephone number at their organization, including the country code.
Fax Number	Fax number for the metadata contact or their organization.
Address*	Address line for the organization with which the metadata data contact is affiliated. This field can accept multiple lines of text.
City*	The city that is associated with the address above.
State/Province/Region*	The state, province, or administrative area associated with the address.
Zip/Postal Code*	The administrative spatial code that is used in mail delivery.
Country*	The country of the organization's address.
Email Address*	Email address of the metadata contact.



## Section 4: Saving your Metadata File

There are two methods that you can use to save your metadata file. These options are both found in the metadata editor menu bar.

- 1) **Save to File:** This option allows you to save the contents of the form in an ISO 19115-2 compliant format and XML file format to your computer. The file will not be checked for completion of all required fields, and is typically used to save an in-progress metadata record so it can be completed later.
- 2) **Check and Save to File:** This option checks to make sure that all required fields are completed before saving the file to your computer. If required fields are incomplete, they will be highlighted in red and the tab(s) with missing fields will be displayed with a red triangle with an exclamation point inside (  ). While this method checks to make sure that all required fields are completed, it does not validate the field contents for accuracy. Once all required fields are completed, this option will save the form in an ISO 19115-2 compliant XML file.

## Appendix A: Example Metadata

The purpose of this appendix is to provide examples for required fields, and many of the optional fields, found in the GRIIDC ISO 19115-2 metadata editor. Many of these are actual examples from datasets submitted to GRIIDC from different scientific disciplines; the information for your dataset will be different. If you have any questions about what information you should provide for any metadata field or element, please contact [griidc@gomri.org](mailto:griidc@gomri.org).

**Tab 1: Dataset Contact Examples**

Dataset Contact Field	Description	Examples
Name*	First and last name of the individual who is responsible for answering questions about or requests related to the dataset. Often this is the primary investigator.	Charles Darwin Albert Einstein Marie Curie Thomas Edison
Organization*	The primary organization with which the dataset contact is affiliated. Usually this is the University and Department where this person works on a daily basis. This information can be found from <a href="http://research.gomri.org">research.gomri.org</a> .	Auburn University, Biology Department Smithsonian Institute Mexican Petroleum Institute Deltares
Position*	The position that the dataset contact holds at their organization.	Professor Research Scientist
Telephone Number*	The dataset contact's telephone number at their organization, including the country code.	1 555-555-1212 ext.0001 47 4555-5555
Fax Number	Fax number for the dataset contact or their organization.	456-566-2314
Address*	Address line for the organization with which the dataset contact is affiliated. This field can accept multiple lines of text.	XYZ Building 1234 Main Street
City*	The city that is associated with the address above.	Madison Corpus Christi
State/Province/Region*	The state, province, or administrative area associated with the address. Do not use abbreviations.	Wisconsin Texas
Zip/Postal Code*	The administrative spatial code that is used in mail delivery.	12345
Country*	The country of the organization's address.	USA
Email Address*	Email address of the dataset contact.	<a href="mailto:example@email.edu">example@email.edu</a>

**Tab 2: Dataset Information Examples**

**Dataset Title Examples\*:**

**(1) Ecology**

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) Physical Oceanography**

Conductivity, temperature and depth data for 12 northwestern Gulf of Mexico locations, May to July 2012

**(4) 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 175 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

**Abstract Examples\*:**

**a) Ecology**

The dataset contains distribution and abundance data for rays in the Gulf of Mexico collected through aerial surveys and satellite telemetry. Aerial survey data includes the date, time, latitude and longitude where rays were observed, and number and species of rays observed during 32 surveys which took place from 2008 to 2011. Satellite telemetry data contains data for 11 cownose rays that were fitted with Smart Positioning or Temperature satellite transmitters between March and November 2011. The satellite telemetry data includes information about the cownose rays that were tracked, including the location they were caught, disk length, disk width, weight, sex, the number of days that the satellite tag remained attached to the ray, the calculated total distance traveled by each ray while fitted with the tag, and the calculated mean and maximum rate of movement for each ray while fitted with the satellite tag. The data from ARGOS that was used to calculate total distance and rates of movement for the tagged cownose rays is included in this data set and includes location date, satellite that received the signal, latitude, longitude, number of messages received and data ARGOS used to calculate the error ellipse for each location.

**b) Chemical Engineering**

This dataset reports the volume, number, surface tension, interfacial tension, and adsorption coefficient for quartz sand aggregates in different hydrophobic liquids over time. Various hydrophobic liquids (decane, tetradecane, hexadecane, benzene, toluene, ethylbenzene, m-xylene, and 2-chlorotoluene) were added to either deionized or salt (34 ppt) water. Quartz sand was then added to the surface of these hydrophobic liquids in water. The quartz sand formed aggregates which were observed over the course of 10 days. This dataset supports the paper: Daria Boglajenko and Berrin Tansel (2016) Encapsulation of light hydrophobic liquids with fine quartz sand: Property based characterization and stability in aqueous media with different salinities, Chemical Engineering Science, 145, 90-96, doi: 10.1016/j.ces.2016.02.010.

**c) Physical Oceanography**

Forty three conductivity, temperature and depth (CTD) casts made from the RV Walton Smith in the northern Gulf of Mexico near DeSoto Canyon in July-August 2012 as part of the Grand Lagrangian Deployment (GLAD) experiment. These CTD casts were made to determine the hydrography of the upper water column as one way to characterize the meso- and submesoscale variability in the region where 297 CODE-type ocean drifters were launched in an attempt to measure multi-scale near surface dispersion. The drifters were deployed at 1 meter and most of them were launched in triplets (initially separated by roughly 100 meters). This dataset was created by the Consortium for Advanced Research on Transport of Hydrocarbon in the Environment (CARTHE). This research was made possible by a grant from BP/The Gulf of Mexico Research Initiative.

**d) Model**

A generalized additive modelling (GAM) approach is used to describe the abundance of 40 species groups (i.e. functional groups) across the Gulf of Mexico (GoM) using a large fisheries independent data set (SEAMAP) and climate scale oceanographic conditions. Predictor variables included in the model are chlorophyll a, sediment type, dissolved oxygen, temperature, and depth. The GAM approach was shown to be robust despite zero-inflated data. This data was used in the publication: Drexler, M., and Ainsworth, C.H. 2013. Generalized Additive Models Used to Predict Species Abundance in the Gulf of Mexico: An Ecosystem Modeling Tool. *PLoS ONE* (8):5:e64458. doi:10.1371/journal.pone.0064458.)

**e) Sociology**

The Community Oil Spill Survey (COSS) is a cross-sectional telephone survey of households randomly drawn from a listed sample of approximately 6000 households living in the coastal zip codes of Plaquemines, Lafourche, and Terrebonne Parishes and the City of Grand Isle in Louisiana. The data presented in this dataset were collected in April 2011, and represent the third wave of the COSS. In total 864 individuals participated in the third wave of the COSS.

**f) Genetics**

This dataset contains partial mitochondrial DNA isolate sequences from 140 *Lutjanus campechanus* (red snapper) collected from the Gulf of Mexico and 35 *Lutjanus campechanus* collected from the Atlantic coast of Florida. Red snapper were collected between 1998 and 2001. Fourteen partial sequences of mitochondrial control regions from *Lutjanus griseus* (gray snapper) are included in the dataset. The gray snapper were collected off Tampa Bay, Florida in 1999 and served as an out group. All sequences are the hypervariable region, ranging from 287 to 293 bp.

**g) Biogeochemistry**

This dataset contains values for nitrification potential, abundances of ammonia-oxidizing bacteria (AOB), ammonia-oxidizing archaea (AOA), and soil properties collected from 13 marshes in Terrebonne Bay (TB), western Barataria Bay (WB), and eastern Barataria Bay (EB), LA. In TB and EB we sampled two oiled marshes and two unoiled marshes, and in WB we sampled three oiled marshes and two unoiled marshes. At each marsh, we collected 5-cm deep soil cores along a 20-m transect from the marsh edge and collecting soil samples every 5 m. Each sample was analyzed for nitrification potential and a suite of physical (soil water content, bulk density) and chemical properties (pH, organic C, total N, molar C:N ratio, total P). Marshes in TB were sampled monthly from May to September 2012, and the WB and EB marshes were sampled in July and September 2012. Samples collected in July 2012 were analyzed for AOA and AOB.

#### **h) Geology**

These data are compiled from 21 different publications dated between 1963 and 2011. They contain stable carbon isotope ( $\delta^{13}C$ ) values measured in core-top sediments, surface grab sediments, and other sediment samples from the sediment-water interface. All samples in this compilation were taken before the 2010 Gulf of Mexico oil spill even if the publication date is after that event. The compilation also includes some stable isotope measurements from carbonate in the same samples and radiocarbon ( $^{14}C$ ) data. These data are useful for comparison to after-spill measurements of the same basic parameters. This dataset was created by the Consortium for Advanced Research on Transport of Hydrocarbon in the Environment (CARTHE). This research was made possible by a grant from BP/The Gulf of Mexico Research Initiative.

#### **Purpose Examples\*:**

##### **a) Ecology**

To assess potential impacts of the Deepwater Horizon incident on rays by comparing distribution and abundance of rays from aerial surveys in completed in 2008-2009 with those completed in 2010-2011 and to delineate fine scale habitat use and movement patterns of cownose rays in the Gulf of Mexico using satellite telemetry.

##### **b) Chemical Engineering**

The aim of this work is to elucidate the relative influences of initial interfacial tension and of dynamic changes in interfacial tension on the dispersion effectiveness of mixtures of sorbitan monooleate (Span 80), (PEO) 20 sorbitan monooleate (Tween 80), and dioctyl sodium sulfosuccinate (DOSS).

##### **c) Physical Oceanography**

These CTD casts were made as part of the GLAD experiment to determine the hydrography of the upper water column as one way to characterize the meso and submesoscale variability in the GLAD experiment region.

##### **d) Model**

The Coupled Regional Climate Model was used to generate accurate, real-time predictions of atmospheric and oceanic conditions in the Gulf of Mexico at a 3km resolution, as well as of passive diffusion pathways.

##### **e) Sociology**

The purpose of this data was to assess how community attachment influences mental and physical health impacts from the BP-Deepwater Horizon disaster over time. To address the following questions: (1) Has the level of negative mental and physical health impacts among residents of the spill affected region changed over time; (2) Does community attachment have an influence on the level of negative mental and physical health impacts people experience; (3) Have the level of negative mental and physical health impacts differed for those employed in the fishing and oil industries versus other occupations; and (4) Do relationships between community attachment, natural resource occupations, and negative health impacts change over time?

##### **f) Genetics**

These data were collected to assess the utility of the hypervariable mitochondrial DNA segment as a genetic marker by examining sequence variation among red snapper samples from geographically remote locations with specific focus on a potential enhancement release site off the Mississippi coast.

**g) Biogeochemistry**

This dataset was developed as part of a research project investigating the effects of the Deepwater Horizon oil spill on salt marsh biogeochemistry. This particular project was directed to determine: 1) if the marsh’s ability to cycle reactive nitrogen was inhibited (nitrification potential); 2) if there was a significant impact on AOB and/or AOA; and 3) if there were spatial (regional, within marsh) or temporal patterns in nitrification potential.

**h) Geology**

The purpose of this dataset is to provide users with a compilation of pre-Deepwater Horizon sedimentary data measured throughout the Gulf of Mexico for comparison to the post-spill bottom sediment. The data come from 20 peer-reviewed research articles and 1 government report.

**Supplemental Information – Data Parameters and Units Examples\*:**

**a) Ecology**

Excel File: MAjemian\_BP\_SGER\_aerial\_survey\_data\_rays.xlsx

[Survey Design Worksheet]

PROJECT: project for which the survey was funded	YEAR: year, YYYY
MONTH: Month, 1 = January and 12 =December	DATE: MM/DD/YYYY
ROUTE: designates the three routes used in each survey, see supplementary material for more detail	TRANSECTS: the number of North-South transects flown, distance in nautical miles
DISTANCE: Total distance of transect flown in nautical miles	WEATHER: General weather conditions, if reported
ANGLE?: Angle (in degrees) over which observer scanned from viewing window on aircraft	ALTITUDE1: Altitude (feet) of aircraft from which aerial survey was completed
SWATH_MI: Swath of observer sighting coverage in miles	SWATH_KM: Swath of observer sighting coverage in kilometers
SWATH_NM: Swath of observer sighting cover in nautical miles	

[Transect Tables Worksheet]

DATE: MM/DD/YYYY	ROUTE: designates the three separate routes used in each survey, see supplementary information
TRANSECT: The transect that the observation took place on	ID: Unique ID generated by combining Julian Date and transect number
Length (mi): Length of transect in miles	Length (km): Length of transect in kilometers
Altitude (ft): Altitude feet of the aircraft from which the aerial survey was complete	Swath (mi): swath of observer sighting coverage in miles
Swath (km): swath of observer sighting coverage in kilometers	Area (sqmi): Area covered by the transect in square miles
Area (sqkm): Area covered by the transect in square kilometers	

[Observations Worksheet]

PROJECT: project for which the survey was funded	YEAR: year
MONTH: Month where 1 = January and 12 = December	ROUTE: designates the three separate routes used in each survey, see supplementary information

TRANSECT: the transect that the observation took place on

SPECIES: species of ray observed

COUNT1: count of individual rays by given time and space, default value of 1 assigned when no value was given for an actual count

LAT: latitude in decimal degrees

DEPTH\_M: bottom depth of water (in meters) at sighting location of rays

TIME: 24-hour military time

COUNT: count of individual rays at the given space and time

COUNTCODE: binary text code given to data to indicate whether a count was actually given (COUNT) or if a count was not given and a default value of 1 was used (DEFAULT)

LONG: longitude in decimal degrees

### b) Chemical Engineering

Flow Rate Experiments and Dispersant to Oil Ratio DOR Experiments measuring Peak volume droplet diameter, outlet conditions, and exit velocity

Outlet conditions or nozzle diameter (millimeters)

Airl (L/min)

Exit Velocity U (meters/second)

Peak diameter (micro-meters)

oil (L/min  $\pm$  0.02)

DOR - Dispersant to Oil Ratio

Interfacial Tension IFT (milli-Newton/meters  $\pm$  0.2)

### c) Physical Oceanography

ASCII data files with 12 columns

Column 1 = depth (meters)

Column 2 = pressure, Digiquartz (db)

Column 3 = longitude (deg)

Column 4 = latitude (deg)

Column 5 = temperatures (ITS-90, deg C)

Column 6 = salinity (PSU)

Column 7 = potential temperature (ITS-90, deg C)

Column 8 = density (kg/m<sup>3</sup>)

Column 9 = oxygen, SBE 43 (mg/l)

Column 10 = nitrogen saturation (ml/l)

Column 11 = fluorescence, Seapoint

Column 12 = flag: 0.000e+00

### d) Model

AKs (meter<sup>2</sup> second<sup>-1</sup>) = salinity vertical diffusion coefficient

AKv (meter<sup>2</sup> second<sup>-1</sup>) = vertical viscosity coefficient

dye\_01 (kilogram meter<sup>-3</sup>) = dye concentration, type 01

evaporation (kilogram meter<sup>-2</sup> second<sup>-1</sup>)

Coriolis parameter at RHO-points = h (meter)

latent (watt meter<sup>-2</sup>) = net latent heat flux

mask\_psi = mask on psi-points

mask\_u = mask on U-points

pm (meter<sup>-1</sup>) = curvilinear coordinate metric in XI

rain (kilogram meter<sup>-2</sup> second<sup>-1</sup>) = rain fall rate

AKt (meter<sup>2</sup> second<sup>-1</sup>) = temperature vertical diffusion coefficient

EminusP (meter second<sup>-1</sup>) = bulk\_flux surface net freshwater flux (E-P)

dye\_02 (kilogram meter<sup>-3</sup>) = dye concentration, type 02

evaporation rate = f (second<sup>-1</sup>)

bathymetry at RHO-points

lward (watt meter<sup>-2</sup>) = net longwave radiation flux

mask\_rho = mask on RHO-points

mask\_v = mask on V-points

pn (meter<sup>-1</sup>) = curvilinear coordinate metric in ETA

rho (kilogram meter<sup>-3</sup>) = density anomaly

salt = salinity  
U = x-wind component

T = perturbation potential temperature ( $\theta - \theta_0$ )  
V = y-wind component

### e) Sociology

Responses to survey questions for 656 survey participants. Disposition (21 classes of response to phone call) (Don't Know/Not Sure, Refused are options for all questions),

1. home proximity to the coast (GoM coastline, GoM coastline waters, less than 1, 1-5, 5-10, 10-20, 20-50, greater than 50 miles),
2. community name,
3. satisfaction with place you live (1-5 dissatisfied to satisfied),
4. likely to recommend to a friend as a place to live (1-5 unlikely to likely),
5. general health (1-5 excellent to poor),
6. stress when oil spill occurred (1-10 none to extremely),
7. stress while the well was leaking (1-10 none to extremely),
8. stress since BP officially declared the well dead in September (1-10 none to extremely),
9. times experienced [anxiety, anger, happiness, worry, depression, sadness, excitement, nervousness, fear, hopefulness] because of the oil spill (1-4 constantly to never),
10. times experienced anger because of the oil spill (1-4 constantly to never),
11. times in the last week [sick to your stomach, diarrhea, headaches or migraines, aches and pains in your joints or muscles, loss of appetite, chest pains, shortness of breath] because of the oil spill (1-4 constantly to never),
12. experienced worry or concern because of the oil spill preventing your good night's sleep because of the oil spill, getting along well with friends, getting along well with family members, your usual daily chores, your being able to take care of your family as well as you would like] (Yes, No)
13. experienced worry or concern because of the oil spill, caused you to [drink alcohol, eat, curse, exercise, sleep, watch television, surf the internet more, get together with friends and family, pray, go to church] more than you usually do (Yes, No)

### f) Genetics

This dataset is made of mitochondrial DNA hypervariable sequences; 124 *Lutjanus campechanus* (red snapper) isolate mitochondrial control regions, partial sequences available on GenBank, accession numbers AF356881–7004, AF356750–776; 24 *Lutjanus campechanus* (red snapper) isolate mitochondrial control region, partial sequences, available on GenBank accession numbers AY1535001-3523; 14 *Lutjanus griseus* (gray snapper) clone mitochondrial control region, partial sequences. Accession numbers AY245003-5016.

### g) Biogeochemistry

Region- Region of collection (TB- Terrebonne Bay; WB- western Barataria Bay; EB- eastern Barataria Bay)  
Site- Site identifier  
Plot- Plots along transect  
Oiled- Indicates whether the marsh received Macondo Oil  
Distance- Distance from the marsh edge of the sampling plot (meters)

Nitrification- Nitrification Potential (nM NO<sub>3</sub>-N per dry gram per day)  
 Organic Matter- soil organic matter content (%)  
 Water Content- gravimetric soil water content (%)  
 Bulk density- soil bulk density (grams per cubic centimeter)  
 Organic C- soil organic carbon content (%)  
 Total N- soil total nitrogen content (%)  
 Total P- soil total phosphorus (micrograms P per gram soil)  
 C:N- soil molar carbon to nitrogen ratio  
 N:P- soil molar nitrogen to phosphorus ratio  
 PO<sub>4</sub>-P- phosphate of overlying water (micromolar)  
 NO<sub>3</sub>-N- nitrate of overlying water (micromolar)  
 Silicate- silicate of overlying water (micromolar)  
 NH<sub>4</sub>-N- ammonium of overlying water (micromolar)  
 AOB- abundance of ammonia-oxidizing bacteria (amoA copies per dry gram soil)  
 AOA- abundance of ammonia-oxidizing archaea (amoA copies per dry gram soil)

#### **h) Geology**

The historical data are aligned into the following columns:

1. Reference number (bibliographic – these numbers are referenced to the different rows in the bibliographic data set).
2. Identification of the sites in the published work
3. Latitude
4. Longitude
5. International isotope scale. 0 means no scale noted, 1 is PDB, 2 is NBS-20, and 3 is NBS-1
6. Stable carbon isotope value reported for sedimentary organic material. Nan means no data reported.
7. 1-sigma deviation to reported stable carbon isotope values in column 6. Nan means no data reported.
8. Stable carbon isotope value reported sedimentary carbonate minerals. Nan means no data reported.
9. 1-sigma deviation to reported stable carbon isotope values in column 8. Nan means no data reported.
10. Radiocarbon (14-C) content of sedimentary organic material, reported as in publication or converted to D14C using the formulae  $D14C=(\text{fractionmodern}-1)*1000$  and  $D14C=(\exp(\text{age}/-8033)-1)*1000$
11. 1-sigma deviation to reported stable carbon isotope values in column 10. Nan means no data reported.
12. Radiocarbon (14-C) content of carbonate minerals in sediments, reported as in column 10.
13. 1-sigma deviation to reported stable carbon isotope values in column 12. Nan means no data reported.
14. Reported depth of sample, in meters. Nan means no data reported.
15. Depth of sample within core or sediment sample, in centimeters. Nan means no data reported.
16. Stable carbon isotope of bulk sediment if no effort to separate mineral fraction from organic fractions was reported.
17. Sample type – 1 – terrigenous input likely, 2 – hydrocarbon seep input likely, 3 – ambient sediment
18. GIS depth. These depths are from Gulf of Mexico geo-referenced bathymetry, available from the National Oceanographic and Atmospheric Association's National Geophysical Data Center, coupled with

GIS techniques using the coordinates of the sample to ascertain the accuracy of the reported depth and georeference.

**Tab 3: Keyword Examples**

**Theme Keyword Examples\***:

**a) Ecology**

Elasmobranch, satellite telemetry, aerial survey, Rhinoptera bonasus, Manta birostris, Aetobatus narinari, spotted eagle ray, cownose ray, manta ray

**b) Chemical Engineering**

poly(2-acrylamido-2-methylpropanesulfonate-co-acrylic acid), superparamagnetic nanoparticles, steric stabilization, high salinity, multipoint grafting, magnetite nanoparticles, electromagnetic imaging

**c) Physical Oceanography**

CTD, hydrography, density, temperature, salinity, oxygen, nitrogen, fluorescence

**d) Model/Numerical Model**

Climate Regional Coupled Model, CRCM, Regional Ocean Modeling System, ROMS, Weather Research Forecasting, WRF, model

**e) Sociology**

social impacts, vulnerability, resilience, health, human

**f) Genetics**

red snapper, Lutjanus campechanus, mtDNA control region, hypervariable region, population structure, Lutjanus griseus, gray snapper

**g) Biogeochemistry**

salt marsh, nitrification potential, ammonia-oxidizers

**h) Geology**

carbon isotope, sediment, stable carbon isotope, coretop, organic material, organic matter, carbonate, seep, terrigenous, marine, continental shelf, abyssal

**Place Keywords Examples\***:

**a) Ecology**

Mississippi – Alabama Shelf, Northern Gulf of Mexico

**b) Chemical Engineering**

not applicable

**c) Physical Oceanography**

Gulf of Mexico, DeSoto Canyon

**d) Model/Numerical Model**

Gulf of Mexico

**e) Sociology**

Southeast Louisiana; Plaquemines Parish, Louisiana; Lafourche Parish, Louisiana; Terrebonne Parish, Louisiana; Grand Isle, Louisiana

**f) Genetics**

Cancun, Mexico; Louisiana; Fish Haven-1, Mississippi; Alabama; Atlantic Coast of Florida; Tampa Bay, Florida; Gulf of Mexico

**g) Biogeochemistry**

Terrebonne Bay, Barataria Bay

**h) Geology**

Mississippi River, Gulf of Mexico, Bay of Campeche

**Topic Category Keyword Examples\***:

**a) Ecology**

biota, oceans

**b) Chemical Engineering**

environment

**c) Physical Oceanography**

oceans

**d) Model/Numerical Model**

oceans

**e) Sociology**

economy, health, society

**f) Genetics**

biota

**g) Biogeochemistry**

biota, environment

**h) Geology**

environment, geoscientific information

**Tab 4: Data Extent Examples**

**Example 1: A dataset produced solely in the lab.** For example, a chemical engineering experiment that created a novel dispersant in the laboratory.

1. Select “This dataset does not have any relevant spatial or temporal components”
2. Description Text\*: This dataset does not have any relevant spatial or temporal information. This dataset was created solely in a laboratory, and consists of data about the creation of a novel dispersant.

**Example 2: A dataset produced using samples collected in the field with known latitudes and longitudes.** For example, a cruise that collected samples in the Gulf of Mexico that started on May 15, 2010 and ended December 1, 2010.

1. Provide coordinates in Coordinate List\*
  - Coordinates should form a box/polygon
  - Use decimal degrees
  - Put a space between coordinate pairs (e.g., latitude, longitude latitude, longitude)
  - Repeat first coordinate pair as the last coordinate pair to make a polygon or box

EXAMPLE: 26.037042,-94.218750 25.601902,-86.901855 23.019076,-88.132324 22.796439,-94.592285 26.037042,-94.218750

2. Time Period Description\*: Ground condition  
Since these data reflect the actual conditions of things in the environment during the time period of the cruise, use “ground condition”
3. Start Date\*: 2010-05-15  
These data represent the ground condition during the sampling period. Therefore this date should reflect the beginning of the sampling or collection period.
4. End Date\*: 2010-12-01  
These data represent the ground condition during the sampling period. Therefore this date should reflect the final date of the sampling or collection period.

**Example 3: A dataset generated by a model for an area with known latitudes and longitudes.** For example, a current model for the Gulf of Mexico from July 1, 2000 to December 1, 2000.

1. Provide coordinates in Coordinate List\*
  - Coordinates should form a box/polygon
  - Use decimal degrees
  - Put a space between coordinate pairs (e.g., latitude, longitude latitude, longitude)
  - Repeat first coordinate pair as the last coordinate pair to make a polygon or boxEXAMPLE: 26.037042,-94.218750 25.601902,-86.901855 23.019076,-88.132324 22.796439,-94.592285 26.037042,-94.218750
2. Time Period Description\*: Modeled period  
Since these data are generated from a model for a specific time period, use “modeled period”
3. Start Date\*: 2000-07-01  
These data represent a modeled time period in the past. The first day that the model applies to should be provided as the start date.
4. End Date\*: 2000-12-01  
These data represent a modeled time period in the past. The last day that the model applies to should be provided as the end date.

**Example 4: A dataset produced from samples taken in the field, latitudes and longitudes are unavailable.**

1. Use mapping tool to draw polygon on map
  - Make sure that screen is at 100% zoom or less
  - Left click on “click here for a mapping tool to help find coordinates”

- Left click to place first point on map
  - Continue to use left click to put additional points on map
  - Double left click when you place last point on map to stop drawing
  - Select “submit”
2. Time Period Description\*: Ground condition  
Since these data reflect the actual condition of things in the environment during the sample collection period, indicate “ground condition”
  3. Start Date\*: 2012-06-01
  4. End Date\*: 2012-06-01

### **Tab 5: Distribution Information Examples**

#### **Data Transfer Format Examples**

File types - csv, txt, xlsx, tiff, jpg, pdf

File types with common names – shp – shapefile, mat – matlab datafile, nc – netCDF

Less common or ambiguous file types – opj – Origin Project File -opens with free download from <http://www.originlab.com/viewer/>, hdr, hex, xmlcon – Seabird – opens with most text editors

#### **File Decompression Technique Examples**

zip, tar, gz

#### **Transfer Size Examples\***

1.3, 0.04, 29

#### **URL Examples\***

**Example 1: Dataset will be distributed by GRIIDC**

<http://data.gulfresearchinitiative.org/data/Y1.x000.000:5555>

**Example 2: Dataset is distributed by a national data center**

<http://data.nodc.noaa.gov/DeepwaterHorizon/MooredBuoy/gulfmex/TR5881/hv0090m.csv>

### **Tab 6: Distribution Contact Examples\***

**Example 1:** The data will be or is already submitted to GRIIDC.

1. Select “This data is hosted with GRIIDC”
2. Fields will automatically populate.

**Example 2: The data has been submitted to another national data center that will make the dataset available to the public.**

Distribution Contact Field	Example(s)
Organization*	NODC, NGDC, NCBI NIH Genbank
Telephone Number*	1 555 555 1212
Address*	Example Organization Name Fictional Department 1234 Main Street
City*	Sarasota
State/Province/Region*:	Florida
Zip/Postal Code*:	12345
Country*:	United States
Email Address*:	<a href="mailto:department@organization.com">department@organization.com</a>
URL*:	<a href="https://www.ncei.noaa.gov/">https://www.ncei.noaa.gov/</a>

**Tab 7: Metadata Contact Examples\***

Dataset Contact Field	Examples
Name*	Charles Darwin Marie Curie
Organization*	Auburn University, Department of Biology Smithsonian Institute Mexican Petroleum Institute Deltares
Position*	Assistant Professor Research Scientist
Telephone Number*	1 555-555-1212 ext.0001 47 4555-5555
Address*	XYZ Building 1234 Main Street
City*	Madison Corpus Christi
State/Province/Region*	Wisconsin Texas
Zip/Postal Code*	12345
Country*	USA
Email Address*	<a href="mailto:example@email.edu">example@email.edu</a>