GRIIDC Data Management Plan
Version 1.0

Data Management Plan

Version 1.0 (May 2012)
GULF OF MEXICO RESEARCH INITIATIVE
INFORMATION AND DATA COOPERATIVE: DATA
MANAGEMENT PLAN

Version 1.0
Released May 2012
CONTENTS

List of Acronyms .................................................................................. iv
Revision History .................................................................................. vi

1. Introduction .................................................................................... 1

2. Guiding Principles and General Data Policy ................................. 3
   2.1. Research Consortium DMP Requirement ......................... 4
   2.2. GRIIDC Long-term Commitment .................................... 5
   2.3. Data Formats ...................................................................... 5
   2.4. Very Large Datasets .............................................................. 6
   2.5. Physical Samples ................................................................. 6
   2.6. Human Subjects Data ......................................................... 6

3. GRIIDC Administrative Structure .................................................. 7

4. GRIIDC Data Management .............................................................. 8
   4.1. Data Source Registry ............................................................. 9
   4.2. Data Transfer and Security .................................................. 11
   4.3. Data System Monitoring ...................................................... 14
   4.4. Data Access and Use Policy ................................................. 15

5. Data Provenance and Metadata ..................................................... 16
   5.1. Community Standards .......................................................... 17
   5.2. Metadata Forms ................................................................. 18

6. Communications and Outreach ....................................................... 18
   6.1. Project Web Materials ......................................................... 18
   6.2. Issue Tracking System ....................................................... 18
   6.3. Data Use Analytics .............................................................. 19
   6.4. Sharing Prohibitions and Licensing ................................... 19
   6.5. Training Courses ................................................................. 19
   6.6. Historical Data ................................................................... 19

Appendices
Appendix 1: GRIIDC Database Metadata Forms ............................... 20
Appendix 2: Research Consortium Data Management Plan
   Template .................................................................................. 21
## LIST OF ACRONYMS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCD</td>
<td>Darwin Core and Access to Biological Collections Data</td>
</tr>
<tr>
<td>ASCII</td>
<td>American Standard Code for Information Exchange</td>
</tr>
<tr>
<td>AU</td>
<td>Administrative Unit</td>
</tr>
<tr>
<td>BP</td>
<td>BP (formerly British Petroleum)</td>
</tr>
<tr>
<td>CIM</td>
<td>Common Information Model</td>
</tr>
<tr>
<td>CSDGM</td>
<td>Content Standard for Digital Geospatial Metadata</td>
</tr>
<tr>
<td>DB</td>
<td>Database</td>
</tr>
<tr>
<td>DMP</td>
<td>Data Management Plan</td>
</tr>
<tr>
<td>DOI</td>
<td>Digital Object Identifier</td>
</tr>
<tr>
<td>DSR</td>
<td>Data Source Registry</td>
</tr>
<tr>
<td>DWH</td>
<td>Deepwater Horizon</td>
</tr>
<tr>
<td>EML</td>
<td>Ecological Metadata Language</td>
</tr>
<tr>
<td>FGDC</td>
<td>Federal Geographic Data Committee</td>
</tr>
<tr>
<td>FTP</td>
<td>File Transfer Protocol</td>
</tr>
<tr>
<td>FWRI</td>
<td>Florida Fish and Wildlife Research Institute</td>
</tr>
<tr>
<td>GCOOS</td>
<td>Gulf of Mexico Coastal Ocean Observation System</td>
</tr>
<tr>
<td>GDP</td>
<td>GRIIDC Data Portal</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographic Information System</td>
</tr>
<tr>
<td>GoMRI</td>
<td>Gulf of Mexico Research Initiative</td>
</tr>
<tr>
<td>GRIIDC</td>
<td>Gulf of Mexico Research Initiative Information and Data Cooperative</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphic User Interface</td>
</tr>
<tr>
<td>HTML</td>
<td>HyperText Markup Language</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>IRB</td>
<td>Institutional Review Board</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>MERMAid</td>
<td>Metadata Enterprise Resource Management Aid</td>
</tr>
<tr>
<td>METAFORE</td>
<td>Common Metadata for Climate Modeling Digital Repositories</td>
</tr>
<tr>
<td>MRA</td>
<td>Master Research Agreement</td>
</tr>
<tr>
<td>NCDC</td>
<td>NOAA National Coastal Data Development Center</td>
</tr>
<tr>
<td>netCDF</td>
<td>Network Common Data Format</td>
</tr>
<tr>
<td>NOAA</td>
<td>National Oceanic and Atmospheric Administration</td>
</tr>
<tr>
<td>OGC</td>
<td>Open Geospatial Consortium</td>
</tr>
<tr>
<td>OPeNDAP</td>
<td>Open-source Project for a Network Data Access Protocol</td>
</tr>
<tr>
<td>PI</td>
<td>Principal Investigator</td>
</tr>
<tr>
<td>QA/QC</td>
<td>Quality Assurance/Quality Control</td>
</tr>
<tr>
<td>QC</td>
<td>Quality Control</td>
</tr>
<tr>
<td>-----</td>
<td>--------------------------</td>
</tr>
<tr>
<td>RB</td>
<td>GoMRI Research Board</td>
</tr>
<tr>
<td>RC</td>
<td>GoMRI Research Consortium</td>
</tr>
<tr>
<td>RDb</td>
<td>GoMRI Research Database</td>
</tr>
<tr>
<td>RFP</td>
<td>Request for Proposals</td>
</tr>
<tr>
<td>TAMUCC</td>
<td>Texas A&amp;M University-Corpus Christi, Texas</td>
</tr>
<tr>
<td>URL</td>
<td>Uniform Resource Locator</td>
</tr>
<tr>
<td>VLD</td>
<td>Very Large Datasets</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
</tbody>
</table>
## Revision History

<table>
<thead>
<tr>
<th>Date</th>
<th>Release</th>
<th>Editor(s)</th>
<th>Clauses Modified</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>01/03/12</td>
<td>0.0.1</td>
<td>J. Gibeaut/</td>
<td>Initial draft of the</td>
<td>Initial draft of the Data Management Plan for GRIIDC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S. Holland/</td>
<td>Data Management Plan for GRIIDC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F. Gayanilo</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. Introduction

The impact of the Deepwater Horizon (DWH) oil spill has occurred across a broad range of coastal and marine environments. Measuring and understanding these impacts, therefore, requires a strategy that will sample the environment using a variety of approaches that generate a variety of data types ranging, for example, from acoustic and optical imagery to point measurements of water quality. Furthermore, the Gulf of Mexico Research Initiative (GoMRI) will engage scientists from around the Gulf of Mexico and beyond to acquire and analyze data using different methods and protocols for data acquisition and management to facilitate data sharing and aggregate analyses. This document describes the data management effort of GoMRI with the overarching goal of ensuring a data and information legacy that promotes continual scientific discovery and public awareness of the Gulf of Mexico ecosystem.

The GoMRI Research Database Program is designed as an overarching program to address database issues that arise during implementation of the GoMRI. This program includes: 1) a data management policy framework with guiding principles derived from the Master Research Agreement (MRA) between BP and the Gulf of Mexico Alliance and in consultation with GoMRI Research Board members; 2) components of the GoMRI Research Database (the project information database and the scientific database); and 3) project level data management plans from approved Research Consortia (RC). Updates of each of the plans described within this program are expected as new research projects are awarded and modifications are made during the course of the GoMRI program.

The Gulf of Mexico Research Initiative (GoMRI) is a 10-year, $500 million independent research program established by a MRA between BP and the Gulf of Mexico Alliance to study the impact of the DWH oil spill on the environment and public health in the Gulf of Mexico. The goal of the GoMRI is to improve society’s ability to understand and mitigate the impacts of hydrocarbon pollution and stressors on the marine environment and apply the accrued knowledge to the restoration and improvement of the long-term environmental health of the Gulf of Mexico. The GoMRI-funded research projects and the scientific data generated will be shared openly and results published in peer-reviewed scientific journals.

GoMRI is guided by an independent Research Board comprising 20 marine science and public health experts who provide scientific direction,
make research funding decisions, and ensure intellectual quality, research effectiveness, and academic independence. The Chief Scientific Officer provides coordination of research and communication activities for the Research Board. The Administrative Unit (AU) manages and administers the overall research program and oversees the Grant Unit in the management of grant agreements and distribution of funds for research awards.

**GoMRI Organization Chart**

[Diagram of GoMRI Organization Chart]

**Figure 1.1. GoMRI Organization Chart**

The MRA authorizes the AU to manage a research database as a key means for public access to GoMRI-funded research. This database has two primary components: (1) information about research projects – to include abstracts, people, institutions, funding source, research theme, publications, presentations, reports, and outreach materials; and (2) information about the scientific data generated by GoMRI. These two components are linked, capturing the association between project information and the resulting scientific data. The Northern Gulf Institute maintains the research project information and the Harte Research Institute (HRI) maintains the scientific data. HRI has established the
GoMRI Information and Data Cooperative (GRIIDC) for the registry of scientific data, data transfer and security services, system monitoring and usage analytics, and communicating with researchers for use of this database.

2. Guiding Principles and General Data Policy

The Master Research Agreement states that development of data management practices will involve consultation with the Directors of the Research Consortia and the Administrative Unit, with the Research Board providing approval of data policies. The Research Board Data Management Committee established the GoMRI Data Management Policy for the scientific data generated by GoMRI-funded research. This policy includes the following guiding principles:

- Timely submission of data and model outputs by RC and Principal Investigators (PIs) to the GoMRI “Research Database;”
- Timely submission of data to existing national repositories;
- Data, results, and ancillary information made fully accessible to the public with minimum time delay;
- Use of existing data standards and management systems as much as practicable;
- Ensure data interoperability to the greatest extent possible;
- Employ best practices for data policy and data management as elucidated by the National Science Foundation (NSF) and NOAA or other agency appropriate for the topic of study; and
- A strong commitment to data management by each RC and participating PI.
2.1. Research Consortium DMP Requirement

A schedule for submittal of a draft data management plan will be agreed on between RC and GRIIDC directors within 30 days of execution of the respective grant agreement. GRIIDC will provide a template (Appendix 2) that can be used to draft the plan. Elements of the data management plan should include the following:

- designation of a RC Data Manager and representative on the GRIIDC advisory committee;
- descriptions of types of data and model outputs to be acquired;
- descriptions of products to be included in the RDb;
- national data repositories to be used for each data type;
- path and timeline for data and metadata flow including the following steps:
  - data acquisition planning,
  - data acquisition,
  - data QA/QC,
  - submission of data and products to GRIIDC, and
  - submission of data, model output, and products to national repositories;
- data and metadata standards for format and content; and
- policies and agreements on data sharing, re-use, re-distribution, and the production of derivatives.

**Figure 2.1. GoMRI Research Database**
This framework outlines how the GRIIDC will collect data from each of the RC. The GRIIDC infrastructure (hardware, software, and network) will support a secure and standardized format for public data sharing. This will facilitate data exchange between RC and allow for informed decision making and policy development strategies.

2.2. GRIIDC Long-term Commitment

GoMRI requires that all data collected and generated by RC be submitted to national repositories. However, this may not be available or possible for some data types. The Harte Research Institute of Gulf of Mexico Studies (HRI) at Texas A&M University, Corpus Christi (TAMUCC) is committed to maintain the infrastructure to serve the RDb and its dependencies beyond the current GoMRI contract. To ensure that the RDb remains current and continues to evolve to serve the community with proper foresight, this document and the policies that are contained herewith, will be reviewed by GRIIDC Advisory Committee during their annual meetings.

2.3. Data Formats

The objective of establishing data standards is to provide a common set of terminologies, formats, structuring, protocols, and definitions as it relates to data management and data sharing. While it is not the policy of the GoMRI to dictate what data standards to employ by RCs, it is understood that generally accepted standards for data and metadata format and content are addressed or appropriate documentation presented where such standards do not exist.

It is important to note that GRIIDC will not transform data and supporting documents submitted. All data will be archived as they are submitted and their structure will remain the same when downloading these data. However, in order to ensure the GoMRI data legacy is accessible and useful to future generations, RCs will be encouraged to submit data in open formats where possible. Open formats can be accessed by both proprietary and open source software, and the implementation details are documented and published so that future users can include these formats into their systems and software.

Examples of open formats include netCDF, a self-describing format which is widely used in oceanography, and ASCII text files, used with
various point data sets such as LIDAR (Light Detection and Ranging) or water quality measurements.

2.4. Very Large Datasets

Selected numerical model output judged to be worthy of long term storage (e.g., climatologies, or detailed flows reconstructions of circulation during the DWH event), if any, will be made available for broader use by the public. Details of the model, forcing fields and perhaps model codes may accompany the model output. Model output is often voluminous or often considered Very Large Datasets (VLD; > 1 terabyte). Typical multi-media data (video, sound, etc.), images from remote sensing and ecological survey transects also fall into this category.

Moving these data from the source to the RDb can cause connectivity problems. Issues related to VLDs will be addressed jointly by the researcher and GRIIDC to explore avenues to move the data efficiently. In most instances, it may be more practical to reference the data via persistent URLs that point to RC storage rather than host it locally on the RDb.

2.5. Physical Samples

GRIIDC will not request, accept or archive physical samples such as:

- samples of water, oil, or sediments, or filtrates of these;
- samples of bacteria, plants or animals or parts there from; or
- analog hard-copy data (e.g., hand-written notes, paper copies, analog photographs).

However, GoMRI encourages RCs to explore avenues to preserve these physical samples if resources are available.

2.6. Human Subjects Data

GRIIDC will not handle and archive human subject research data or data requiring Institutional Review Board (IRB) certification. All data requiring such certification will remain in the possession of and be managed by the IRB-certified researcher. However, these data will be recorded in the Data Source Registry (DSR).
3. GRIIDC Administrative Structure

GRIIDC is based at HRI at TAMUCC. Figure 3.1 is the organizational structure of GRIIDC to support the mandate of the organization. A GRIIDC Advisory Committee (Table 3.1) will be formed to consist of two Co-Chairs, the GRIIDC Director, and a RB Data Management Subcommittee representative, one GoMRI AU representative, the Chief Science Officer, one NOAA National Coastal Data Development Center (NCDDC) representative, the data manager from each RC, and a single member to represent the individual/small group researchers. In support of the Director, the Advisory Committee provides the GRIIDC direction and feedback through the evaluation of past activities, and their annual review of the data management policies as stipulated in this document.

In close coordination with the Director, the GRIIDC will be led by the Systems Architect who will orchestrate the design, development, testing, deployment, and documentation of the GRIIDC Data Portal (GDP) that encompasses all the modules necessary to establish the RDb. The team will be supported by a full-time Technical Coordinator to assist the team coordinate actions and training requirements with elements of GoMRI AU, RB and RCs.

A team of technical staff (Data Administrator, Software Engineers, Web Application Developer, Subject Matter Experts, Data Manager, Data Analyst and GIS Analyst) will conduct the technical activities of GRIIDC. These technical activities will be supported by full-time representation from the Gulf of Mexico Coastal Ocean Observing System (GCOOS) and the Florida Fish and Wildlife Research Institute (FWRI) to ensure interoperability with other regional systems, facilitate intra-regional body communication and exchange technologies.
The GRIIDC was organized based on its three primary functions: data management (see Section 4), system integration (services to facilitate search and data exchanges) and communication (see Section 5).

4. GRIIDC Data Management

RCs will collect and generate terabytes of data. These data will be made available to researchers and the general public. To support public access, GRIIDC will implement: (1) a registry of data from previous and current GoMRI projects and selected historical data, (2) data transfer and
data security services to provide access to the data and maintain the integrity of the data, (3) a system to monitor the status of the Research Database and data usage analytics, and (4) a user registration and data access control subsystem. All these features will be contained in a GRIIDC Data Portal (GDP) that will be deployed via the Internet.

4.1. DATA SOURCE REGISTRY

Data sets will be registered in a Data Source Registry (DSR). Data will be registered by GRIIDC staff or by RC data managers. Figure 4.1 is a schematic representation of the registration process. The primary objective of DSR is to document the source of the data, how to access the data, data status (preliminary or final), size of file to download, and conditions associated with data use. The records in the registry will not be limited to current GoMRI projects but will also include records harvested from other repositories and historical data that are relevant to GoMRI research.

GRIIDC will initialize the DSR based on GoMRI records compiled by AU for previous GoMRI projects and related data. However, RC Data Managers are responsible for the creation, validation and updating of their respective records in the registry. The registration or updating of records can be done via user-interactive online forms that will be developed and deployed by GRIIDC. As new data or revised data are made available, the registry will be updated. Data in the GDP will not be updated until the DSR is updated.
Figure 4.1. Schematic representation of the GRIIDC Data Portal and Data Source Registry. The registration of data sources can be done manually or can be done via automated process.

To simplify the installation of automated registry functions, GRIIDC will design, develop and deploy web service toolkits that can be installed on RC computing platforms to effect an automated updating of the registry.

In addition to the GRIIDC project and data metadata forms (see Appendix 2) that will be provided by each RC, the registry will also include the following information:

- Physical location (complete URL) of the data/file and corresponding metadata;
- Authentication parameters if needed to download the files;
- Data type and size of the file;
- Data access protocols;
- Data usage constraints if any;
- If record was revised, version and nature of change; and
- Point-of-Contact for the data or file.
A Graphic User Interface (GUI) will be developed and deployed on the Internet to facilitate the search, browsing and exploration of the data in the registry.

4.2. DATA TRANSFER AND SECURITY

Data and metadata generated by RCs will be submitted to GRIIDC (Figure 4.2). Some of these data will be submitted via online data forms (Appendix 1) or administrative forms (Appendix 2) but the bulk of the data required to build the RDb will be moved during the project period.

4.2.1. PULLING THE DATA

The data collected and/or generated in the program will be 'pulled' in an automated process by GDP using the records in the DSR. There are a number of possible options in downloading RC data. GRIIDC will evaluate each option with care to ensure that all parties involved will be able to support the infrastructure.

DSR provides, among others, the parameters to define what to pull, size of file, when to pull, conditions of data use, and authenticating parameters for an automated data 'pull' onto RDb. GDP will constantly monitor the DSR for changes (Figure 4.1) and these changes will activate a routine to schedule the harvest. Data download will be made during lean periods for large datasets or files, instantaneous download if otherwise, or as dictated by a parameter in the registry record. The determination of what is a large dataset or file will be determined jointly by GRIIDC and the data provider.

4.2.2. RECEIVING THE DATA

An object-relational database engine, such as PostgreSQL, will be used to host the data. GRIIDC will not evaluate the completeness of the data or datasets being submitted and no data level quality control (QC) will be performed. However, the required metadata file will be validated for completeness and compliance. RCs will be notified if the metadata needs to be corrected or completed. Despite errors that may be detected on the metadata forms data download will continue whenever possible. To avoid or minimize the possibility of errors in the metadata file (and remain in compliance with FGDC’s CSDGM and ISO 19115-2 (North American Profile)), GRIIDC will provide materials to facilitate the creation or modification of the required metadata files.
4.2.3. GRIIDC Network

All communications to and from the GDP will be via the internet. To ensure 24/7 services, GRIIDC will maintain a redundant system (high
availability) and evaluate network configurations to ensure that read/write functions are optimized (Figure 4.3 & 4.4). Technology changes rapidly and the network configurations will be evaluated continually to ensure optimal performance. An off-HRI/TAMUCC system backup will be established to supplement the on-site backup procedures that should be inherent in the network configuration and facilitate rapid redeployment if needed.

![Conceptual GRIIDC Data Portal network infrastructure](image)

**Figure 4.3. Conceptual GRIIDC Data Portal network infrastructure.**

### 4.2.4. Getting Data from GRIIDC Data Portal

Data access and other data exchanges will all be executed via the Internet (e.g. HTTP request, FTP, OPeNDAP). GRIIDC will implement a high-speed and high bandwidth connection wherever possible to facilitate exchanges. However, it is given that the bottleneck will be the speed of the internet itself. Handling an unknown number of data requests
presents another challenge. To avoid service denial problems or sudden slow down of data exchanges, especially in cases where multiple requests are made at almost the same time, large data download from the GDP will be done in stages:

1. Users will send data request to GDP via online forms to define the data package;
2. GDP will evaluate the request and package the data;
3. GDP will deposit the data onto an FTP server that will be made available to the requestor for limited time; and
4. Requestor will be notified via emails.

To facilitate data discovery and access, GRIIDC will develop data explorer packages with smart search facilities. The use of image thumbnails, table sample or other ways to present a snap-shot of the data will be employed.

4.3. **Data System Monitoring**

GRIIDC will deploy a monitoring system accessible from the GRIIDC website. This system will support management's need to monitor and report on the progress of GoMRI. The information monitored will include:

- Number of data sources;
- Database size;
- Rate of growth by size and data types;
- Rate of access by size and data types; and
- Number of users.

GoMRI encourages data sharing, not only between researchers within each RC but also between RCs and the general public. To facilitate sharing, the monitoring system will expose the following:

- Data collection plans (parameters, location, time, frequency);
- Expected date of data availability;
- Submission to GRIIDC status (0-not submitted; 1-submitted but no metadata; 2-submitted but incomplete metadata; 3-submitted but not verified; 4-submission verified); and
- Submission to a national repository status (0-not submitted; 1-submitted; 2-not applicable).
4.4. Data Access and Use Policy

Data access will be monitored via a user registration system and subject to data use/reuse agreements. Registration is free and user information will not be used or distributed for any other purposes. The system will have three user classifications (Figure 4.4): General public, GoMRI Researcher, and RC Data Managers.

**General Public:** The ‘general public’ classification will include anyone who is not a member of any of the RCs. Their registration will be evaluated by GRIIDC and registrants will be notified by email for acceptance. Registrants with a profile in GulfBase (http://www.gulfbase.org/) can be approved automatically. Whenever applicable, prospective GDP users will be encouraged to register with GulfBase. These users can browse the repository and view data that are open to the general public. They may also use Web Services to download data.

**GoMRI Researcher:** All researchers of GoMRI will, by default, be classified to this level. They will have access to data and functions available to the general public and to preliminary data that are being shared between RCs.

**RC Data Managers:** Each RC will designate a Data Manager that will be given access to a ‘Builder’ package that will allow the data manager to add and edit data that are already in the repository. Moreover, they will also be given the authority to use the GDP facility to facilitate upload of data onto national repositories (when applicable).
The advent of the internet and ability of researchers to move data easily, and facilities to extract data from documents or databases, has introduced a new problem when defining the origin of the data. In a multi-disciplinary/multi-institutional scenario, it is often difficult to determine the origins of the data collected, compiled and archived in a central repository. To facilitate the process of tracing the origin of data, sensors or process used and transformations (if any) and allow the validation of scientific databases, the use of metadata standards will be imposed. Data provenance is central in the validation of data and as such, no data will be archived in RDb and served to the public unless data provenance is addressed. The RDb will maintain its integrity and reliability when the process used to create them are reproducible and analyzable for defects.

**Figure 4.4.** GRIIDC Data Portal conceptual use-case.

## 5. DATA PROVENANCE AND METADATA
5.1. Community Standards

The supplemental documentation can come in many forms but the use of metadata standards has gained acceptance in most communities. Metadata in digital form will accompany each data set stored in separate companion files. The auxiliary information required (metadata content) and the formatting (metadata schema) will follow published standards. There are a number of metadata editors available in the market but GRIIDC will, in close coordination with NOAA NCDDC, use and recommend the application of MERMAid (Metadata Enterprise Resource Management Aid) whenever possible. Other utilities available, such as ArcCatalog, may also be used as long as they are in compliance to the Federal Geographic Data Committee (FGDC) endorsed Content Standard for Digital Geospatial Metadata (CSDGM). MERMAid was designed as an open source application that allows a secure Web-based platform, custom organization, enhanced validation, and exports in HTML, text, and XML formats. GRIIDC in close coordination with NCDDC will organize training programs to introduce the metadata standards and the editors. More on these training courses can be found at: http://www.ncddc.noaa.gov/metadata-standards/metadata-training/.

Other editors that are in compliance to ISO 19115-2 (North American Profile) are also acceptable. Many federal agencies, including NOAA, are moving towards the adoption of the ISO standards for Geospatial Data. GRIIDC will accept, validate, and process metadata files submitted following any of the two standards.

The ISO 19115-2 and CSDGM standards may not be sufficient to describe the types of data that are collected or generated. Other standards such as Ecological Metadata Language (EML) as established by the Ecological Society of America, Common Information Model (CIM) as established by Common Metadata for Climate Modeling Digital Repositories (METAFOR), Darwin Core and Access to Biological Collections Data (ABCD) as promoted by the Biodiversity Information Standard will be evaluated jointly with RCs.

GoMRI will be collecting many types of data and if a profile or metadata data standard is not available, a working group will be organized to draft one. It is not the intent of GoMRI to process the new profiles for acceptance as a community standard and be endorsed by FGDC and/or Open Geospatial Consortium (OGC). Although it is a nice product that can
emanate from such exercises, the profiles will be designed to only document the data presented and address the origin of the data.

5.2. Metadata Forms

GRIIDC will continue to provide facilities to simplify the creation of metadata. There are metadata that describe the data, as presented above. There are also metadata to describe the project itself. The NSF has provided a data management template that is used by GRIIDC as a template in fulfilling the project-level metadata. These forms and short guide is presented in Appendix 1. More of such utilities are expected to be developed by GRIIDC.

6. Communications and Outreach

One of the primary functions of GRIIDC is on communication and outreach activities. Several facilities will be developed or installed and outreach activities will programmed to ensure that communication will continue to flow, not only within RCs, but to all the general public as well.

6.1. Project Web materials

GRIIDC will maintain web presence via the GoMRI website and via its data portal, GDP, for data management and discovery. In close coordination with other AU departments (Northern Gulf Institute and Consortium for Ocean Leadership), GRIIDC will provide web materials to present new developments in GDP, highlight of activities and media events, and schedule of activities.

6.2. Issue Tracking System

It is a GRIIDC policy that all issues raised, comments, suggestions and other community requests are managed and tracked. GRIIDC will install an issue tracking system like JIRA (http://www.atlassian.com/software/jira/overview) or Request Tracker (http://bestpractical.com/rt/) to document these inputs. In the spirit of sharing and open source development, all inputs from registered users and RC researchers, will be posted in a publicly-accessible issue tracking site and will be managed jointly by all members of GRIIDC. All registered users will be encouraged to participate in online discussions.
6.3. **Data Use Analytics**

All data requested by users of the GDP will be monitored. GoMRI researchers will be regularly notified as to the number of downloads, data citation for all Digital Object Identifier (DOI)-stamped data in the repository, user demographics and other data use analytics to help assess data reusability and importance.

6.4. **Sharing Prohibitions and Licensing**

It is the intent of GoMRI to make all data collected and generated public. However, there are cases where public-sharing introduces legal issues. GRIIDC will, via its DSR, use whatever prohibition or access protocols dictated by the researcher submitting the data. However, RCs will be encouraged to use *Creative Commons by Attribution* (CC-BY 3.0; http://creativecommons.org/licenses/by/3.0/) or something similar.

6.5. **Training Courses**

GRIIDC will organize and hold training courses, seminars and workshops to demonstrate how to use various utilities of GDP. This will include the use of the Data Source Registry (DSR), System Monitoring, data transfer/Exchange utilities, data discovery and other related data management tools. Some of these courses will be held jointly by other agencies, like the training course on the use of MERMAid will be coordinated with NCDDC.

6.6. **Historical Data**

In support of GoMRI researchers, GRIIDC will locate historical data that are relevant to RCs. The DSR will include records that point to these repositories. It will include a link to the data source, and the process required to secure the data. When applicable and legal, these data can be archive in the RDb to facilitate access. RCs may make the request and GRIIDC will evaluate the request and prioritize.
APPENDIX 1. GRIIDC DATASET FORM

I. Project Information
* a. Project Title:
   [SELECT]

   b. Dataset Title:

II. Dataset Steward Information
* a. Primary Contact Name Responsible For This Dataset:
   [SELECT]

   b. Secondary Contact Name Responsible For This Dataset:
   [SELECT]

III. General Dataset Information
* a. Brief description of the purpose for developing dataset:

   b. What order of magnitude is the dataset size? (TB, GB, MB, KB)

   c. What is the geographic area (provide bounding description of the study area or latlon if known)?

   d. What is the dataset type?:
      [SELECT]

      If not on the list, please identify dataset type:

   e. What is the approximate dates for data acquisition?

      Start Date: [MONTH] [YEAR]

      End Date: [MONTH] [YEAR]

Submit & Add Another Dataset  Done
DATA MANAGEMENT PLAN REQUIREMENTS

Each PI receiving funding through a GoMRI RFP will be required to develop a Data Management Plan that will identify the activities of their respective researchers that produce, generate, or even request data. This document should describe how the project will conform to GoMRI policy on the dissemination and sharing of research data. In this document, researchers should present a plan to share (1) analyzed data, (2) metadata that provide origin information, and (3) metadata that describe how the data were generated. The GRIIDC will work with the PIs, and provide them with guidance and documentation to develop their Data Management Plans.

HELP WITH YOUR DATA MANAGEMENT PLAN

To ensure that all projects emerging from GoMRI meet the data management plan requirements, data managers are encouraged to consult with their researchers when drafting this plan and when necessary, data managers are encouraged to consult with staff at the GRIIDC.

To create your Data Management Plan, write a short introduction on the project and rationale for the project, then go through the sections on the next page. Start by answering the questions within each section (numbered under each section description). Modify the answers into prose that makes sense as a paragraph below each Roman numeral header (include the bold text as the header to each of your sections in your Data Management Plan).

Upon completion of this exercise, you will have achieved a Data Management Plan with section headers followed by paragraphs that detail those points. Send a copy of your completed Data Management Plan to Lauren Showalter at the GRIIDC at Lauren.showalter@tamucc.edu. The GRIIDC team will review your plan, make any suggestions or recommendations, and/or help answer questions on items which you may be uncertain.

1 Questions and document template are adapted from the NSF Data Management Plan Template.
To have your Data Management Plan reviewed by the GRIIDC, please forward your document to Lauren.showalter@tamucc.edu.

DATA MANAGEMENT PLAN

I. Data Management Point of Contact

1. Who will be the (lead) PI of this project?
2. Who will be in charge of overall decision making regarding data management? (List name, association, address, phone, and email address)
3. Who will be in charge of day-to-day data management? (If different from above, list all names, associations, addresses, phones, and email addresses)
4. Where will the physical location of the data be? If data will be distributed at separate locations, please list each location and the data or part of the data that they will house.
5. What is the commitment of the organization that will house the data (or part of it) until it is transferred to a national repository or to GRIIDC?

6. Types of data

Samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project:

1. What data will be generated in the research? (Give a short description, including approximate amount or file size—if known—and the content of the data).
2. What data types will you be creating or capturing? (e.g. experimental measures, observational or qualitative, model simulation, processed, etc.)
3. How will you capture or create the data?
4. When will these data be generated?
5. How long will each data be kept by the PI and why?
6. If you will be using historical/existing data, state that fact and include where you will get the required data. What relationship will exist between the data you are collecting and the existing data?

III. Data and Metadata Standards
Standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies).

1. Which database engine and/or file formats will you use for your data, and why?
2. What contextual details (metadata) are needed to make the data you capture or collect meaningful to other researchers? Be sure to include any naming schemes, data dictionaries, taxonomies, or ontologies if applicable.
3. How will you create or capture these details?
4. Which metadata standards will you use (list by tasks and/or data type)?
5. Why have you chosen particular standards and approaches for metadata and contextual documentation? (e.g. recourse to staff expertise, Open Source, accepted domain-local standards, widespread usage).

IV. Policies for access and sharing and provisions for appropriate protection/privacy

1. How will you make the data available? (Resources needed: equipment, network and communication systems, expertise, etc.)
2. When will you make the data available?
3. What is the process for gaining access to the data by other GRIIDC researchers and the general public?

Provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements:

1. Are there ethical and privacy issues?
2. If so, how will these be resolved? (e.g. anonymisation of data, institutional ethical committees, formal consent agreements.)
3. Will the dataset generated by this project be covered by copyright? If so, who owns the copyright and other intellectual property?
V. Policies and provisions for re-use, re-distribution

1. Will any permission restrictions need to be placed on the data?
2. Which bodies/groups are likely to be interested in the data?
3. What and who are the intended or foreseeable uses/users of the data?

VI. Plans for archiving and preservation of access

Plans for archiving data, samples, and other research products, and preservation of access to them:

1. If applicable, which national archive/repository/central database/data center will be identified as a place to deposit data?
2. What transformations will be necessary to prepare data for preservation/data sharing? (e.g. data cleaning/anonymisation where appropriate.)
3. Other than the corresponding metadata documents, what other documentation will be submitted alongside the data or created on deposit/transformation in order to make the data reusable?