GRIIDC: A Tale of One Data Repository's Journey to Become FAIR

Rosalie Rossi*, James Gibeaut & Anthony Reisinger

Harte Research Institute for Gulf of Mexico Studies, Texas A&M University-Corpus Christi, Corpus Christi, Texas

OUR MISSION

To ensure a data and information legacy that promotes continual scientific discovery & public awareness of the Gulf of Mexico ecosystem.

THE BEGINNING: BEFORE FAIR

The Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC) is a multidisciplinary data repository created as a result of the Deepwater Horizon oil spill. GRIIDC developed a repository while researchers were collecting post-spill data with a focus on ingesting data quickly. Almost 10 years later, GRIIDC has improved the data management system implementing FAIR data principles (Figure 1).

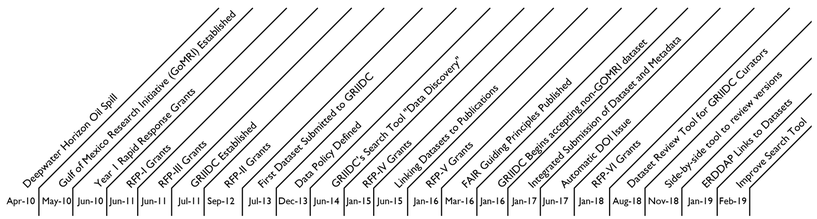


Figure 1: GRIIDC Timeline of Events

GRIIDC FUNDING ORGANIZATIONS

GRIIDC not only houses GoMRI funded data but also stores data from the following organizations.

MBRACE – Mississippi Based RESTORE Act Center of Excellence

TEXAS ONEGULF Center Of Excellence – Science Driving Solutions

FLORACEP – Florida RESTORE Act Centers of Excellence Program

NATIONAL ACADEMY OF SCIENCES

GULF OF MEXICO RESEARCH INITIATIVE

TEXAS A&M UNIVERSITY CORPUS CHRISTI | HARTE RESEARCH INSTITUTE FOR GULF OF MEXICO STUDIES

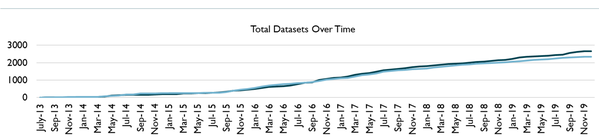
NOAA

TEXAS PARKS & WILDLIFE

GRIIDC DATA TYPES

- Oceanography
- Climatology
- Biology
- Chemistry
- Health and Social Science
- Physics
- Engineering
- Modeling
- Ecology
- Geology
- Meteorology
- Environmental Science

GRIIDC STATS



2548 DATASETS

291 RESEARCH GROUPS

2721 PEOPLE

73.9 TERABYTES

Figure 2: Total Number of Datasets Over Time in GRIIDC Repository

WHAT IS FAIR DATA?

The FAIR guiding principles for scientific data management and stewardship were published in Scientific Data in 2016. These principles provide best practices for data management that ensure data are findable, accessible, interoperable, and reusable.

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016) doi:10.1038/sdata.2016.18

IS GRIIDC FAIR?

GO FAIR

F ☑ FINDABLE

- Data are assigned a digital object identifier (DOI) issued by DataCite
- Data contain rich metadata
- Data can be searched using GRIIDC's data discovery tool, DataONE, Registry of Research Data Repositories (re3registry), and Google Dataset Search
- Oceanographic datasets are sent to National Centers for Environmental Information for long-term archival

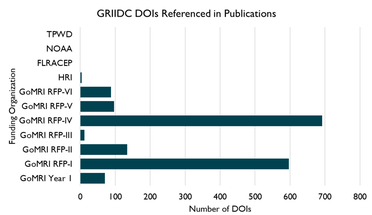


Figure 3: GRIIDC DOIs Acknowledged in Publications

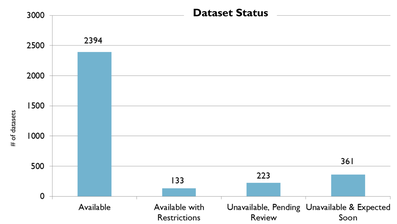


Figure 4: Number and Status of GRIIDC Datasets

A ☑ ACCESSIBLE

- Use standardized communication protocol that allows for authentication
- Metadata are accessible via the dataset landing page after dataset package review process
- If embargo on dataset, metadata still available for download

I ☑ INTEROPERABLE

- Data can be downloaded through the Environmental Research Division's Data Access Program (ERDDAP) server
- Metadata reference DOIs of other related datasets

R ☑ REUSABLE

- Data and metadata are curated by a team of subject matter experts
- Data usage license available when downloading dataset and also located on GRIIDC website
- Data contain point of contact, authors, and some provenance information
- Data include standard ISO 19115-2 metadata

DATA CURATION = FAIR DATA

GRIIDC's team of subject matter experts curate all data and metadata that are submitted to GRIIDC. Data files are reviewed to ensure that they:

- are not corrupt or empty
- do not contain materials subject to copyright
- are in non-proprietary formats
- contain complete datasets
- are well-documented
- include detailed spatial extent

CHALLENGES

- Data management system being developed while researchers were collecting data
- Resources initially focused on ingesting data quickly to comply with data sharing policy
- Data sharing a new concept; data were poorly organized and not well-described
- Data curation lacking in beginning due to limited resources
- Data curation difficult when reviewing a variety of data types
- Hard to interoperate data from different fields of study with different standards, file sizes, and file types
- Larger data stored at Amazon Web Services and take longer to distribute

DEVELOPMENT IMPROVEMENTS - TOWARDS FAIR DATA

- Add DOI attribute to metadata
- Add data usage license to metadata
- Standardized vocabulary
- Detailed provenance
- Utilize Globus for large data distribution

GOT DATA?

In order to ensure a data and information legacy we need data!

- GRIIDC stores many different types of data including data that do not have national data archives available for storage
- The GRIIDC system is user-friendly and provides tools for search and tracking
- Contact us for your data storage needs!!!!

CONTACT

Rosalie Rossi
Program Manager
(361) 825-3719
Rosalie.Rossi@tamucc.edu
Inquiries: griidc@gomri.org

ACKNOWLEDGEMENTS

This work was made possible by a grant from the Gulf of Mexico Research Initiative and support from the Harte Research Institute for Gulf of Mexico Studies at Texas A&M University-Corpus Christi.

GULF OF MEXICO RESEARCH INITIATIVE

TEXAS A&M UNIVERSITY CORPUS CHRISTI | HARTE RESEARCH INSTITUTE FOR GULF OF MEXICO STUDIES