Texas OneGulf Knowledge Base

Strategic Plan 1.0

December 31, 2017

This project was paid for [in part] with federal funding from the Department of the Treasury through the State of Texas under the Resources and Ecosystems Sustainability, Tourist Opportunities, and Revived Economies of the Gulf Coast States Act of 2012 (RESTORE Act). The content, statements, findings, opinions, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the State of Texas or the Treasury.

Introduction

Led by the Harte Research Institute (HRI) for Gulf of Mexico Studies, the Texas OneGulf Center of Excellence is a partnership of nine Texas institutions dedicated to the long-term environmental and economic health of Texas and the Gulf of Mexico. Texas OneGulf was established in 2015 as part of the implementation of the federal RESTORE Act following the Deepwater Horizon oil spill. The consortium is advancing research into ongoing impacts from the oil spill and other long-term issues that threaten the health and sustainability of the Gulf of Mexico. The vision of Texas OneGulf is to become a trusted network of Gulf experts that provides science-driven information to decision makers and resource managers working to ensure the environmental, economic, and human health sustainability of the Texas Gulf coast. A major initiative to realize this vision is the establishment of the Texas OneGulf Knowledge Base (TOKB). This document is the first strategic plan defining and setting forth the initial development of the TOKB.

This strategic plan was developed by data scientists and subject matter experts in coastal and marine science at HRI at Texas A and M University – Corpus Christi (TAMUCC) and at the Department of Oceanography of Texas A and M University (TAMU). Both groups have major data curation programs. HRI is home to the Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC) and the Gulf of Mexico Alliance Geoportal (GOMAportal). TAMU is home to the Gulf of Mexico Coastal Ocean Observing System (GCOOS). GCOOS has much experience in developing products and services for users of the Texas coast in collaboration with their Products and Services Advisory Council. In addition, HRI, through their ongoing roles in helping to lead the development of the Texas Coastal Resiliency Master Plan (TCRMP) and revamping Texas' Resource Management Codes, is experienced collaborating with the state's coastal management community.

Defining the TOKB

The **vision** for the TOKB is a society taking full advantage of scientific research to better live and work on a dynamic and diverse Texas coast. Our **mission** is to improve the usability of existing scientific data to address Texas' Gulf coast issues. The TOKB will do this by making relevant data widely accessible and driving the progression from data to information to knowledge. Data are facts or objective observations and typically lack context or interpretation. Information is derived from data that is processed and placed in context making it useful for specific purposes. Knowledge is subjective and involves expert insight and intuition grounded in reliable data and information.

Developing actionable information requires access to reliable and well-documented scientific data. The TOKB, however, will not be a separate data repository; rather it will rely on the existing GRIIDC, GOMAportal, and GCOOS repositories to archive relevant datasets. The TOKB will focus on seven activities.

Activity 1: Making existing scientific data repositories more known and accessible for Texas Gulf applications.

The TOKB will use GRIIDC, GOMAportal, and GCOOS to archive datasets, but it will also expose Texas-relevant datasets to TOKB users through a search tool on the TOKB website.

Expected outcomes: The Texas Gulf community is aware of major data repositories and accesses them to retrieve or submit data.

Activity 2: Identifying and curating existing underserved datasets.

Underserved datasets are important datasets that are difficult to discover or use for one reason or another. Typical reasons are that the data are in heterogeneous and/or difficult to use formats, they are not accessible through network-enabled software, only a subset is needed out of a large data file or they were orphaned by the individual or institution that created them. Datasets at risk of being lost for future reuse have limited public access or proper documentation, and may only reside in places such as academic research labs, government offices, or with non-governmental organizations. The TOKB will identify priority datasets with help from the research and management communities and conduct the proper curation for submission to one of the TOKB foundational repositories.

Expected outcome: The overall security and use of the Gulf's scientific data and information resources are increased.

Activity 3: Developing and curating new data products.

A data product is a dataset that is derived from an existing dataset or an integration of two or more datasets. It may be a subset of a larger dataset, a standardization of data parameters, a change in format, or a conversation to a different data type, such as a spreadsheet to a GIS file. It may also include additional processing. A product that merges datasets is often desirable to ensure related data values are kept together in their proper relationships and to enable analyses across related datasets. An example is to combine shoreline change rates with shoreline type.

Expected outcome: With the generation of select data products, the management and scientific communities find new and more efficient uses of existing datasets.

Activity 4: Development of information products and tools derived from technical and specialized scientific datasets in collaboration with users.

This activity will take advantage of existing data and data products and present them in a context for addressing particular types of problems, management tasks, or policy development. The need for particular tools or products will be determined in collaboration with the resource management community. Example tools would be an interactive map showing impacts of future sea level rise and a literature atlas.

Expected outcome: Texas Gulf coast stakeholders use more data-driven approaches to help make decisions affecting the natural, socio-economic, and built environments.

Activity 5: Dissemination of user stories, lessons learned, and best practices for applying the data and information products to managing the Texas coast.

This is where users exchange their experiences and insights on how the data and information are being used, should be used, or should not be used in addressing Texas Gulf issues. Insights and knowledge from other experts will also be compiled. This is the activity where the TOKB will attempt to capture the knowledge of the community specific to data-driven approaches of coastal management and policy development.

Expected outcome: Gulf stakeholders more intelligently use data and information products by recognizing shortcomings and proper applications.

Activity 6: Developing and maintaining the TOKB website.

The website will connect users with original datasets in the foundational repositories, data and information products and tools, and knowledge resources.

Expected outcome: The TOKB becomes a trusted and valuable source for addressing Texas Gulf coast issues.

Activity 7: Advancing data sharing in the Gulf of Mexico.

Without scientists and engineers effectively sharing their datasets, a great resource to Texas will be lost. The TOKB will further the development of a data sharing culture by helping to ensure those responsible for generating valuable data are recognized for their work. The TOKB will also seek to collaborate with state agencies to improve the capture of data from consultant studies and funded research programs.

Expected outcome: The effective sharing of scientific data (Open Data) becomes ingrained in the scientific and technical communities.

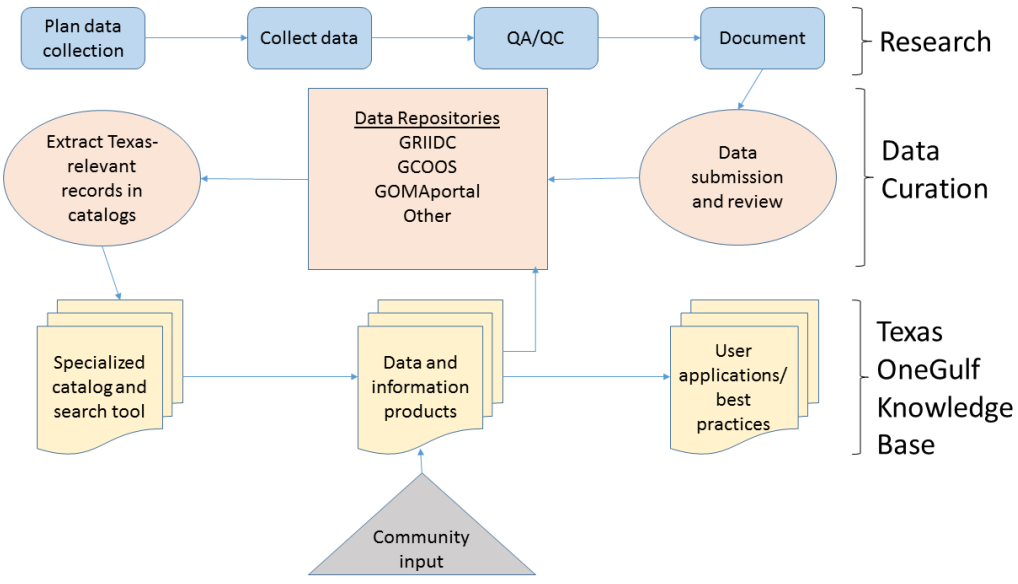


Diagram showing the relationship and activities of the TOKB relative to scientific research that collects data, data curation, and the user community.

Foundational Data Systems of the TOKB

The TOKB will use datasets and make datasets discoverable from established data repositories. It will also submit datasets and data products, which are developed or used by the TOKB, to repositories. The nature and source of the dataset will determine which repository is used for archiving. The use of reliable and existing data repositories will be the most efficient means for developing the vision of the TOKB and will allow resources to focus on developing the needed information and knowledge products. The major underlying repositories are GCOOS, GRIIDC, and GOMAportal.

These repositories and programs have different objectives: (1) GCOOS focuses on providing on-going coastal and oceanographic monitoring data primarily to scientists, but also to those developing value-added data products to the public; (2) the GRIIDC database houses data from discreet and completed scientific research projects for access primarily by scientists; and (3) GOMAportal houses mostly non-federal, coastal geospatial datasets. More in depth descriptions are provided in the appendix.

Website Information Technology

The website for the TOKB is located at http://tkb.geos.tamu.edu/. In time it will be moved to the ".org" domain. The physical host is a Lambda Labs computer housed in the Department of Oceanography, Texas A&M University in College Station, TX. The computer is configured with a pair of NVIDIA 1080Ti General Purpose Graphics Processing Units (GPGPUs) cards, 128 GB of CPU RAM and 4TB of solid state disk in a Raid 5 configuration. It runs a UNIX variant (Ubuntu 16.04) and was pre-configured by the vendor with software for the GPGPU cards. Staff have installed software frameworks for websites, data servers, and accelerated database services. The system will be maintained and backups will be made by project staff located at Texas A&M University (TAMU). TAMU provides 1-Gb network access to the machine and 10Gb network lines to the building. TAMU IT services maintain the campus firewall and scan campus machines regularly to ensure that security patches and system software updates are current. None of the data held by the computer will be confidential, proprietary or secret. No personally identifiable information such as social security numbers or credit card information will be collected or stored on the machine.

Initial Data and Information Priorities

HRI received input during five regional meetings during the fall of 2016 of the TCRMP's Technical Advisory Group (TAC). The TAC is made up of over 100 individuals with knowledge of environmental, economic, infrastructure, or cultural issues facing the Texas coast. Members represented federal and state resource agencies, NGOs, U.S. Army Corps of Engineers, planners, consultants, and local mayors and commissioners. As this multidisciplinary group of experts interacted with geospatial datasets in small groups, they provided us with data and information needs for managing the Texas coast. The table below reflects a subset of unmet needs for managing the coast with the aim of increasing resiliency. In addition the list includes coastal ocean and oceanic datasets that GCOOS has identified as valuable, underserved datasets at risk of being lost or neglected due to lack of proper curation. The TOKB will address these items

during its initialization. Future data and information products will be identified in collaboration with the state's Data Standards Committee (DSC) and other management entities as well as the Texas OneGulf Network of Experts (TONE).

Priority initial datasets/data products for development or curation in the TOKB (MOs: estimated time required to complete in months)

Subject	Description	Purpose	Data Type	Source	Status	MOs
Colonial water birds (coastal plain and nearshore)	Bird counts by species at nesting sites along the Texas coast	Manage birds and as ecosystem health indicator	geospatial, vector, time series	Texas Colonial Waterbird Society	Database available, geospatial processing needed	4
Landcover/use (coastal plain and nearshore)	Classified map describing major environments and land uses	Manage critical habitats for ecosystem and community resiliency	geospatial, raster	Sentinel 2 satellite imagery	Extensive processing needed	9
DEM (coastal plain and nearshore)	Fused optimal DEM of coastal zone	Storm surge/sea level rise vulnerability assessment and geoenvironments classification	geospatial, point and raster	state, federal, university	Extensive methods development and processing needed	4
Literature Atlas (coastal plain and nearshore)	Geographic footprints of studies and links to reports/articles	Aid discovery of prior work and studies, Aid in permitting and planning	geospatial, polygon	Harte Research Institute literature catalog, local, state and federal agencies, universities, consultants, NGOs	~700 studies with footprints available, need to develop GIS application	4
Geoenvironment change index (coastal plain and nearshore)	Measure of past and potential future environmental change	Assess need for restoration or reduction of vulnerability to changing conditions	geospatial, Q4 grid	Texas Coastal Resiliency Master Plan, Harte Research Institute	QA/QC needed	3
Shoreline type and change dataset (coastal plain and nearshore)	Merged historical shoreline change and type for bay and Gulf shoreline	Assess shoreline retreat and impact of hard shoreline structures	geospatial, polyline	Harte Research Institute and Bureau of Economic Geology	Processing required to merge datasets	3
Georeferenced vertical aerial photography of shoreline (coastal plain and nearshore)	Historical photography used for change assessment going back to 1930's	Quantitative/ qualitative assessment of environmental change	geospatial, raster scans of positive prints	Harte Research Institute and Bureau of Economic Geology	Select historical epochs available	6

Subject	Description	Purpose	Data Type	Source	Status	MOs
Oysters (coastal plain and nearshore)	Locations of current and historical subtidal and intertidal oyster reefs	Management of oyster resource and water quality	geospatial, polygon	Texas Parks and Wildlife and legacy datasets	Updates to current dataset needed	4
Sea level rise landscape change (coastal plain and nearshore)	Year 2100 landscape change caused by Sea level rise	Planning for restoration and overall coastal resiliency	geospatial model output, raster	Texas Coastal Resiliency Master Plan, Harte Research Institute	Model output in development	12
Year 2100 storm surge inundation (coastal plain and nearshore)	Reach and depth of hurricane surge on year 2100 landscape following erosion and sea level rise	Planning for restoration, protection, and overall coastal resiliency	geospatial model output, raster	Texas Coastal Resiliency Master Plan, Harte Research Institute	Model output in development	12
LATEX hydrographic dataset (coastal ocean and oceanic)	1,650 hydrographic profiles from the Tx/LA shelf acquired in 1992/94	Baseline oceanographic data	geospatial, points with depth	GCOOS	Data need reformatting and documenting to current standards	9
R/V Manta standard oceanographic data (coastal ocean and oceanic)	Meteorologic and oceanographic data acquired when R/V Manta is underway	Baseline oceanographic data	geospatial, points	NOAA, GCOOS	Data need processing and standardization	9
Autonomous Glider Trajectories along the Texas coast (coastal ocean and oceanic)	6 glider deployments funded by the Texas OneGulf Center of Excellence	Water quality assessment.	geospatial and temporal measurements of temperature, salinity and optical parameters	Deployed by Geochemical and Environmental Research Group of Texas A&M University in 2016-2018.	Some deployments remain. Near real-time data obtained. Delayed-mode data pending.	1
GOM Bibliography of Oceanographic Data Reports (coastal ocean and oceanic)	106 GOM Oceanographic Data Reports published between 1956-2001	Report time, location and parameters measured during various studies.	geospatial, temporal and of station locations and data summaries	Compiled by staff a the Department of Oceanography at Texas A&M University, College Station, TX.	Reports in hand. Boundaries not extracted.	4
GOM Theses and Dissertations (coastal ocean and oceanic)	61 theses and dissertations related to GOM oceanography 1961-2006	Identify ocean studies related to the GOM	various study topics.	Compiled by staff a the Department of Oceanography at Texas A&M University, College Station, TX.	Topics and Titles need to be filtered for studies applicable to Texas	1

Subject	Description	Purpose	Data Type	Source	Status	MOs
GOM Bibliography of 844 published works related to GOM oceanography (coastal ocean and oceanic)	844 citations for GOM oceanography updated in 2006.	Identify scholarly works related to GOM Oceanography	various topics of study.	Compiled by staff a the Department of Oceanography at Texas A&M University, College Station, TX.	Topics and titles need to be filtered for studies	1

Time Frame

The TOKB is an ongoing effort as the data, information, and knowledge needs grow and evolve through time. Following, however, is the outlook and priorities during the initialization of the TOKB.

Near term – January - September 2018

Developing the TOKB website is a priority but is also an ongoing effort as new information products are developed. The priority function of the website is to enable links and searching of the foundational repositories as well as describing the TOKB vision to the community.

Curating existing underserved datasets and developing new data products are also priorities. The initial priority list is presented above. The datasets and products will reside in one of the repositories and be discoverable through the TOKB website.

Midterm – late 2018 - 2019

Developing initial information products and tools will follow the development of the initial datasets/products. The state's Data Standards Committee, which oversees the creation of the Resource Management Codes, will provide input on the information products and tools to be developed.

Work on bringing together the "knowledge" portion of the TOKB will begin in this time frame.

The TOKB will also continue to identify underserved datasets and collaborate with resource agencies to improve data sharing.

Long-term and Ongoing

Once the TOKB is established in all activities, work will continue as directed by GCOOS, HRI, and the data management community.

Appendix – Description of Foundational Data Systems for the TOKB

GCOOS

GCOOS-RA and GCOOS are related but distinct

The Gulf of Mexico Coastal Ocean Observing System Regional Association (GCOOS-RA) is one of 11 Regional Associations, which span the United States and its territories. Its purpose is to organize regional producers and consumers of marine environmental data to serve regional priorities in human health and safety, environmental health, and the economy. It does this by awarding sub-contracts to data providers to maintain their observing systems, by hosting stakeholder meetings, and by conducting outreach activities. GCOOS-RA is funded by the U.S. Integrated Ocean Observing System (IOOS), which is a Federal-regional partnership led by the National Oceanic and Atmospheric Administration (NOAA). GCOOS-RA is currently operating under a third 5-year (2017-2021) grant from the IOOS Program Office. The GCOOS-RA annual grant varies from year to year but is typically in the range of $1.5M/Yr to $2.0M/Yr.

GCOOS is the data stewardship activity of the GCOOS-RA which: aggregates data from Federal and non-federal providers, transforms these data into standard and user-preferred formats and serves these data and data products through Data and Product Portals. The Data Portal provides access to the data by both humans and machines. GCOOS is primarily concerned with oceanographic, marine meteorological and biogeochemical data but in recent years has added biological data (fish and plankton). GCOOS manages near real-time data streams (e.g., automated weather buoys), delayed-mode data (e.g., oceanographic cruise data) and historical data collections. Products include quasi-static information such as bathymetry and climatologies, near real-time maps of current oceanic conditions, and historical collections of nutrients and dissolved oxygen. The GCOOS activity is supported by approximately 20% of the GCOOS-RA budget.

GCOOS-RA Business Website: http://gcoos.org/

GCOOS Data Portal Website: http://data.gcoos.org/

GCOOS Products Portal Website: http://products.gcoos.org/

Overview of data types

Quasi-Static

Bathymetry and coastline data (http://gcoos.tamu.edu/products/topography/Introduction.html) are popular with basemap makers and numerical modelers who want accurate representations of the Gulf in their circulation models. Users can access numerical values, maps, or GIS layers. Average temperature and salinity profiles (http://products.gcoos.org/en/ts-profiles/) are used by engineers to calculate buoyancy for rig transport and placement.

Near Real-Time

GCOOS devotes much of its effort toward maintaining a near real-time data aggregation system. This system monitors data availability and ingests approximately 1500 data streams drawn from

17 providers. Sensor report 1 to 3 measurements per hour. Most of these data streams are of ocean currents measured from fixed and mobile offshore oil-platforms (see http://data.gcoos.org/phenomenonFull.php). The other parameters are air pressure, air temperature, relative humidity, winds, water temperature, salinity, waves, water level, turbidity, chlorophyll and dissolved oxygen.

Historical

Historical collections include river discharge, nutrients, and dissolved oxygen. The river discharge database contains measured discharge values from the first recorded values to the present date for 61 rivers that discharge into the Northern Gulf of Mexico. An interactive map shows the location of the gauging station and the point where the discharge reaches the Gulf of Mexico. Roll-over plots show the discharge for the current year on top of the long-term annual mean cycle. http://products.gcoos.org/en/river-discharge/. All known nutrients and dissolved oxygen data collected around the coastal and offshore Gulf of Mexico through 2014 are available from the GCOOS Hypoxia-Nutrients Database. (https://nutrients.gcoos.org). These data were standardized, both units and nomenclature under a grant from the Gulf of Mexico Alliance.

Biological

Data on fish abundance (species, counts and methods) were transformed from original formats into standards-based formats recommended by the Ocean Biological Information System (OBIS) group as part of an IOOS pilot project. The first dataset transformed was the "Comparative Assessment of Gulf Estuarine ecoSystems (CAGES) Program (http://products.gcoos.org/en/fish-and-marine-mammals/cages-info/) obtained from the NOAA National Marine Fisheries Service (NMFS)-Galveston. Subsequent datasets were processed in support of a Marine Biodiversity Observation Network focused on fishes observed in the Florida Keys. GCOOS hosts data collected from Surfside, TX by an instrument that takes photomicrographs of plankton and passes them to an automated classifier. This system is used to detect Harmful Algal Blooms (HABS), which occur off the Texas coast and affect human and animal health. Pictures, time series plots and data are available through http://products.gcoos.org/en/hab/.

GRIIDC

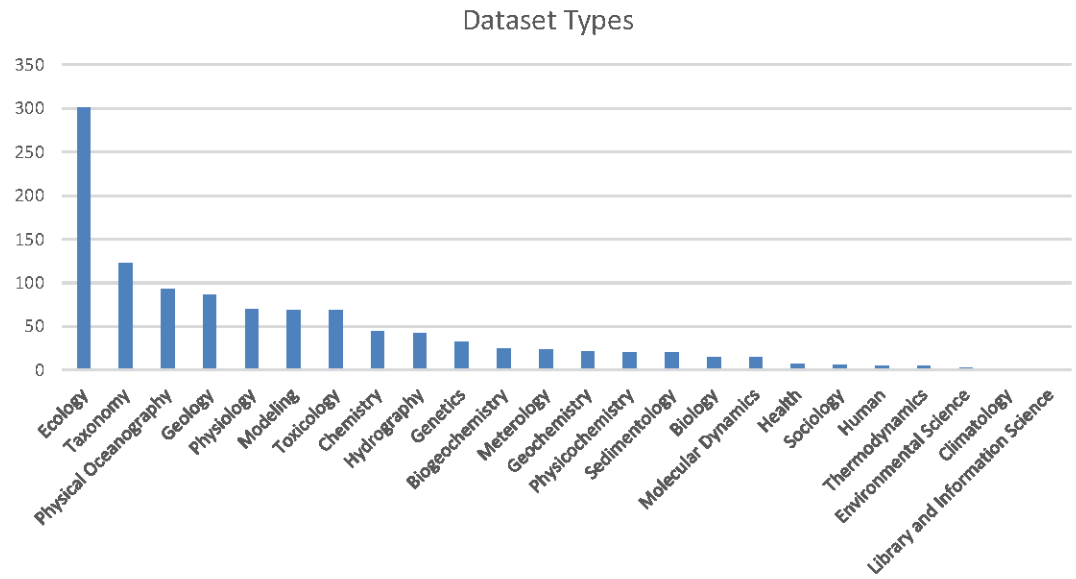
The Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC) is the data management entity responsible for acquiring, cataloging, and curating the full-body of data derived from the 10-year $500 million research program (Gulf of Mexico Research Initiative, GoMRI) that began a few days after the start of the *Deepwater Horizon* event. The GRIIDC network includes some 2,600 researchers and 280 institutions. HRI houses, develops, manages, and maintains GRIIDC for GOMRI and has recently expanded to support the Florida and Mississippi RESTORE Act Centers of Excellence in scientific data curation as well as other individual Gulf of Mexico researchers wanting to share their data. GRIIDC's focus is on managing datasets developed from discreet academic research studies. The goal is to provide a curated data package that, if appropriate, can be cited as a product of the researcher responsible for producing the data, much like a peer-reviewed journal article.

GRIIDC is a team of subject matter experts, data specialists, and software application programmers who have developed and continue to improve and maintain a data management program to store scientific data generated by Gulf of Mexico researchers. The GRIIDC data management system provides researchers, research groups, or funding programs with a variety of tools to help manage and track data throughout the data lifecycle of a project. For example, the GRIIDC Dataset Information Form (DIF) is a resource designed to assist researchers with data management planning. While the system assists researchers with multiple phases of data management, the main functions of the system are storing and sharing data. GRIIDC is also a Member Node of DataONE, a system that allows environmental data searches across multiple member repositories. DataONE has increased the discoverability of GRIIDC datasets.

GRIIDC is supported mostly by GoMRI with an annual budget of $2.0M/Yr. In 2021 the GoMRI funding will reduce, but GoMRI is committed to supporting GRIIDC until 2030. HRI is also committed to continuing and expanding GRIIDC with funding from state and federal research programs, private foundations, and industry to name a few sources.

Overview of data types

As of December 2017, the GRIIDC database had over 30TB of data in 1,890 datasets. Researchers from diverse fields of study, including biology, chemistry, physical oceanography, sociology, political science, and public health, are able to store their data in the GRIIDC system. These datasets are almost exclusively from the Gulf of Mexico, and although most were not acquired in Texas, many pertain to Texas environmental issues. The below graph shows the distribution of the number of datasets by type in GRIIDC.



Numbers of datasets in GRIIDC by type as of July, 2016, as defined by metadata theme keywords. Some datasets are counted more than once because they span multiple categories.

GOMAportal

The Gulf of Mexico Alliance Geoportal (GOMAportal) was established in 2011 by HRI with funding from NOAA in support of the Gulf of Mexico Alliance (GOMA). The goal of GOMAportal is to archive and make available datasets that do not have a clear path for archiving. These datasets are typically originate with state agencies, universities, consultants, or NGOs. GOMA funded HRI and partners in each Gulf state to discover, collect, document, and distribute these 'orphaned' datasets as part of the Governor's Action Plan II. Over 400 datasets were identified in the original effort. These datasets were collected and documented with FGDC-compliant metadata and made discoverable and downloadable in GOMAportal (http://www.gomaportal.org/). Since then, other datasets have been added, including datasets created by the previous GOMA activities, and datasets collected by graduate students funded by the NOAA Environmental Cooperative Science Center. As of fall 2016, GOMAportal housed 900 datasets, all easily searchable, fully documented, freely downloadable with FGDC-compliant metadata.

HRI continues to maintain GOMAportal, and GOMA recently funded upgrades to the system. HRI is installing the latest version of the open source application geoportal-server, adding federated search capability to search other geoportal instances, and adding a metadata creation tool. In addition, select datasets will be published as map services. HRI will continue to house and maintain GOMAportal and has the authority to add additional datasets and map services as needed for the TOKB.