3-2019

Coordination of DWH Long-Term Data Management: The Path Forward Workshop Report

Coastal Response Research Center

COORDINATION OF DWH LONG-TERM DATA MANAGEMENT: THE PATH FORWARD

DECEMBER 4 - 5, 2018

A WORKSHOP REPORT

NOAA'S GULF OF MEXICO DISASTER RESPONSE CENTER
MOBILE, AL

COASTAL RESPONSE RESEARCH CENTER

Table of Contents

I. Acknowledgments

This workshop and report were funded by the National Oceanic and Atmospheric Administration's (NOAA's) Office of Response and Restoration (ORR) and NOAA's National Marine Fisheries Service (NMFS) Restoration Center (RC). Other collaborative partners included the Gulf of Mexico Research Initiative (GOMRI), NOAA National Centers for Environmental Information (NCEI), the National Academy of Sciences, Engineering, and Medicine (NASEM) Gulf Research Program (GRP) and other federal and state natural resource damage assessment and restoration (NRDAR) trustees. The following individuals participated on the Organizing Committee.

- Nancy Kinner, Coastal Response Research Center*
- Benjamin Shorr, NOAA ORR ARD, Spatial Data Branch*
- Marti Goss, NOAA Restoration Center*
- Lauren Showalter, NASEM Gulf Research Program*
- Mike Peccini, NOAA Restoration Center*
- Sharon Mesick, NOAA National Centers for Environmental Information (NCEI)*
- Jessica Henkel, Gulf Coast Ecosystem Restoration Council (RESTORE Council)
- Nicolas Eckhardt, NOAA ORR ARD, Spatial Data Branch
- Jay Coady, NOAA ORR ARD, Spatial Data Branch
- Kyle Wilcox (who replaced outgoing Sandra Ellis, GRIIDC), Axiom Consulting
- William Nichols, Harte Research Institute for Gulf of Mexico Studies

*Denotes Core Team Member

The workshop was facilitated by Dr. Nancy Kinner from the Coastal Response Research Center (CRRC; www.crrc.unh.edu). CRRC is known for its independence and excellence in areas pertaining to environmental engineering, marine science and ocean engineering as they relate to oil spills. The Center has widespread experience working with oil spill-related issues and has conducted 60+ workshops that bring together researchers, practitioners, and scientists of diverse backgrounds (e.g., academia, industry, government, NGOs).

We wish to thank each of the presenters for their participation in the workshop:

- Benjamin Shorr – Overview of Deepwater Horizon Long-Term Data Management
- Jessica Henkel and Kathryn Keating– Overview of DWH LTDM Survey Results
- Jessica Henkel and Nicolas Eckhardt – DWH LTDM Data Management Standards Working Group Outcomes
- Jay Coady and Kyle Wilcox – DWH LTDM Interoperability Working Group Outcomes
- Michael Peccini and William Nichols – DWH LTDM Discovery and Searchability Working Group Outcomes
- Lauren Showalter – DWH LTDM Path Forward

Thank you to our Stakeholders Perspectives Panel:

- Jessica Henkel, Restore Council
- Lauren Showalter, NAS Gulf Research Program
- Natural Resource Damage Assessment (NRDA) – Eric Weissberger and Michele Jacobi

Thank you to our panelists who presented their Perspectives on Data Systems[1]:

- Sharon Mesick – NCEI Overview
- Benjamin Shorr – DIVER
- Rosalie Rossi – GRIID-C
- Craig Conzelmann – CIMS
- Laura Shumway – USEPA WQX
- Shin Kobara – GCOOS Overview
- Debra Hernandez – SECOORA Overview

We would like to give special thanks to: (1) the Breakout Group Leads:
Nicolas Eckhardt, Jessica Henkel, Courtney Arthur, Kathryn Keating; Jay Coady and Kyle Wilcox; Michael Peccini and Benjamin Shorr; Jennifer Bauer; Ann Jones; Steve Delgreco and Michele Jacobi; and (2) Kathy Mandsager, Jesse Ross and Melissa Gloekler for their note-taking during the workshop.

We also appreciate the NOAA staff at the Gulf of Mexico (GOM) Disaster Response Center (DRC) in Mobile, AL for providing an excellent venue for this workshop.

II. Introduction

Following the 2010 DWH Oil Spill a vast amount of environmental data was collected (e.g., 100,000+ environmental samples, 15 million+ publicly available records). The volume of data collected introduced a number of challenges including: data quality assurance, data storage, data integration, and long-term preservation and availability of the data.

An effort to tackle these challenges began in June 2014, at a workshop focused on environmental disaster data management (EDDM) with respect to response and subsequent restoration. The EDDM collaboration improved communication and collaboration among a range of government, industry and NGO entities involved in disaster management. In June 2017, the first DWH Long-Term Data Management (LTDM) workshop focused on reviewing existing data management systems, opportunities to advance integration of these systems, the availability of data for restoration planning, project implementation and restoration monitoring efforts, and providing a platform for increased communication among the various data GOM entities. The June 2017 workshop resulted in the formation of three working groups: Data Management Standards, Interoperability and Discovery/Searchability. The groups spent 2018 coordinating and addressing various complex topics related to DWH LTDM. On December 4th and 5th, 2018 the Coastal Response Research Center (CRRC), NOAA Office of Response and Restoration (ORR) and NOAA National Marine Fisheries Service (NFMS) Restoration Center (RC), co-sponsored a workshop entitled **Deepwater Horizon Oil Spill (DWH) Long-Term Data Management (LTDM): The Path Forward** at the NOAA Gulf of Mexico (GOM) Disaster Response Center (DRC) in Mobile, AL.

The December 2018 workshop had 50 participants (Appendix A) representing a diverse group of organizations and affiliations involved in Gulf of Mexico data coordination. Because a person can join the LTDM effort at any time, some participants attended the first workshop, while others had been involved in the working groups. The objectives of the workshop were to: 1) continue collaboration

[1] N.B. In the context of data management, systems can be applications or integrated applications. There is a blend of these meanings throughout this document depending on the speaker's and groups' perspective.

among Gulf of Mexico partners involved with DWH research, restoration, and monitoring, 2) report on progress and outcomes of the three working groups, and 3) translate the DWH LTDM efforts into deliverables or actionable items.

Questions addressed at the 2018 workshop included:

1. What does collaboration "look like" (e.g., RESTORE Council, GOMRI, GRIID-C, NAS Gulf Research Program, GOMA, NFWF, and Gulf Environmental Benefit Fund)? What does success look like?
2. What progress has been made by the working groups? What has been challenging? What are the recommended best practices and suggestions to address challenges?
3. What concrete steps should be taken to foster collaboration and data integration? (Identify short, medium, and long-term actions, what is their priority? Ease of achievement, cost and organizational mandates/priorities?)
    - Data Systems perspectives: how do we use what we have learned to improve the practice of systems collaboration (this will then make it easier for data providers to understand best management practices)?
    - Funders perspective: how do we simplify/improve workflow for data providers to improve use of standards and best management practices?
4. What is the future direction of the DWH LTDM effort?

The 2018 workshop consisted of a plenary session, a panel discussion with various data management stakeholders, and three breakout group sessions. The workshop agenda (Appendix B) was developed by the organizing committee. Plenary presentation topics included: results from a pre-workshop DWH LTDM stakeholders survey, updates from the three working groups, and presentations on perspectives from GOM data management stakeholders (e.g., funders, users) and data system owners. Participants in the breakout groups were tasked with revising/formulating recommendations put forth by the working groups, prioritizing them (e.g., by assessing ease of achievement or cost), and discussing their feasibility and steps to implementation.

III. Plenary Presentations

Survey Results

Jessica Henkel (RESTORE Council) summarized and compared results from a pre-workshop survey given to participants prior to the June 2017 and December 2018 workshops (see survey questions and results in Appendix C). Forty-four people completed the 2018 pre-workshop survey.

Survey respondents were asked to identify their own role within their agency. A majority defined their role as *data managers/administrator,* and almost half identified their role as *data user*. Other participant roles included: *program manager/funder*, *data generator*, and *decision-maker*, and *other*. Almost half of the survey respondents indicated that their work involved two or more of these roles.

Survey respondents were asked "What do you want from Gulf research/monitoring data 15 years from now?" This question, was based on a previous workshop survey administered in 2017, and asked participants to select from six potential outcomes (see Appendix D). In 2018, most of the respondents said they would like to see: all data follow a common set of standards for documentation and data files, interoperable platforms, freely accessible data, and a long-term data repository. As a follow-up to this question, respondents were asked, "Realistically, what do you think Gulf research/monitoring data will

have achieved in 15 years?" Compared with the 2017 survey, more respondents indicated in 2018 that they believe that no change will realistically take place with respect to GOM-wide research/monitoring data improvements. Compared to 2017, fewer participants in 2018 indicated that they think it is realistic for data to reside in long-term repositories or adhere to a common set of standards. However, more participants believed in 2018 that data will be fully accessible and interoperable allowing users to develop their own systems or synthesize data through existing tools.

These conflicting viewpoints highlight the challenges to reaching long-term data management goals in the Gulf of Mexico. As identified by survey respondents, the biggest challenges for data management in the GOM were identified as variation in data systems and standards, effective means of collaboration and communication, sufficient funding, variation in goals and missions, and leadership and its attitudes about data management. The final survey question asked participants to rank potential workshop outcomes that they deemed most beneficial. Respondents selected "translation of Working Groups' coordination and research into action" to be the highest priority outcome. Other workshop outcomes were ranked in the following order: enhanced collaboration among Gulf partners involved in research, restoration and monitoring; enhanced understanding of LTDM across Gulf partners through summarized Working Group findings; and established future directions for this collaboration.

Data Management Standards Working Group (SWG)

Nicholas Eckhardt (NOAA) and Jessica Henkel (RESTORE Council) were the Team Leads for this working group, with support from Kathryn Keating (RESTORE Council) and Courtney Arthur (IEc). The charge for the SWG was to: 1) identify categories of standards needed (i.e., data acquisition protocols, quality control, data management), and 2) determine the gaps in data management system standards. The group met monthly to gather information across entities on data management systems and standards. The SWG developed an inventory of more than 30 different systems used for data management in the Gulf of Mexico. From this list, 16 "actionable" data systems were selected for a detailed evaluation of data management standards in order to determine commonalities and discrepancies. These evaluated data systems helped to compile and highlight the data management system standards currently being used in the Gulf of Mexico.

The group developed a template to document the data standards used within each data management system (e.g., metadata standards, system purpose, fields used to organize data, associated valid values). These detailed templates were then combined into a "crosswalk" to facilitate cross-system comparisons. This crosswalk includes summary information (e.g., system owner, contact information, system purpose), a review of specific attributes of the fields utilized within the data system (e.g., type, format, definition, valid values or constant, etc.) and a list of all valid values for key fields.

The SWG recommended that after a final review by individual data system managers, documents be made available to the public and shared at appropriate forums, with sufficient detail to document the working group's process. In addition, data system owners and sponsors were encouraged to make their data standards and protocols publicly available, easily accessible, and up-to-date on their websites (See https://crrc.unh.edu/DWH_Long-Term_Data_Management for available downloads).

The SWG recommended the following next steps and associated challenge(s):

(1) Develop a set of recommended data standards
Challenge: Variation in goals across agencies makes this a complicated task.

a. How is this challenge best addressed?

(2) Develop a data exchange format for systems
Challenge: This is a resource-intensive process.
   a. Who has the capacity and resources to develop this and advocate for implementation?

Interoperability Working Group (IWG)

The Working Group Leads were Jay Coady (NOAA), Kyle Wilcox (Axiom Consulting) and Sandra Ellis (until August, 2018, formerly of GRIIDC). The goals of this group were to: 1) determine what could optimize interoperability efficiency between DWH LTDM systems, and drive collaboration among them; 2) compile strategic goals and key features for data warehouses and repositories; and, 3) determine the intended, current and future use of DWH LTDM systems.

The group's first step was to determine which applications/systems should be reviewed. Once those ~40 systems were identified, the group compiled attributes specific to them and documented which attributes pertained to data exchange. Commonalities between systems were identified and a matrix was created to compare the various systems. The attributes used in the commonality matrix were: data types, system functions, metadata types, and data delivery. The commonality matrix showed overlap between systems, their relative size/capacity and levels of interoperability among them (e.g., IOOS → ERMA, GRIIDC ← →DIVER). The commonality matrix and document attributes associated with systems were derived from the common data services and data/metadata standards.

Interoperability between systems is challenging to achieve because it is not typically a mandate for any one system owner/developer, thus organizations are less likely to spend resources on this effort. Machine-to-machine communication would improve interoperability between systems, but again this takes significant time and resources. In general, organizations focus on internal interoperability and internal client needs more than enhancing integrating across data systems. In order to achieve cross-system integration some agreement on common data standards, data services and vocabularies will be required (See https://crrc.unh.edu/DWH_Long-Term_Data_Management for documents).

The IWG recommended:

(1) The development of guidelines for external data system communications (e.g., M2M communication) allowing automation to improve interoperability.
(2) Programs, funding agencies, and data systems should make interoperability a requirement/goal at the onset.
(3) An end-to-end interoperability pilot program should be implemented. This would showcase the ability for different data systems to integrate with one another and could be used as a template.

Discovery/Searchability Working Group (D/SWG)

Michael Peccini (NOAA) and William Nichols (GRIIDC) led this group whose goals were to: 1) evaluate and prioritize DWH data community needs for searchability and discovery, and 2) inventory primary data management systems. The process began by compiling an inventory of the user community for DWH data in the GOM. Once those data users were identified, a survey was sent to the user community at large to better understand their data search needs. Nine organizations responded to the survey, answering questions such as "what types of data are you searching for?" and "what is the purpose of your research?". The D/SWG saw this as a successful effort and recommended conducting the survey at a larger scale to include more users. Survey results identified the types of systems people were using

and vocabularies being used in search tools. Findings included: the need for searchability across systems (i.e., general exploration of data is challenging); tools that better facilitate general exploration of data (i.e., need to know what systems to go to in order to find data); better representation of state and regional levels in Gulf-wide data systems; and a searchable inventory of systems and the kind of data they house.

In addition to the survey, a search log analysis was completed using several existing data systems (e.g., GRIIDC, DIVER) to see what users are looking for, how they are searching, and if they are finding data they need. Many systems do not have free text searches, instead they have categorized data, guided queries and custom queries. Additionally, search logs are not automatically generated for every system and are not easily compared across systems. Despite those challenges, the search logs from GRIIDC, SECOORA, and DIVER were analyzed. GRIIDC's Elastic Search Logs were used as an example; details such as session ID, time stamp, search terms, geofilter used, number of results, 1st and 2nd search scores, and data landing page links were extracted and used in the analysis. Elastic Search generates automatic, custom logs which provided insight into what the users are actually searching. For GRIIDC users are searching data systems using identifiers they know (e.g., digital object identifier (DOI), UDI, authors). The community is using free text searches; geofilters are only used about 4% of the time. The analysis included information on whether users were finding data when searching the GRIIDC data system.

The SECOORA "Pageviews" search log is a unique tool that provides information on what users are searching for, how the users are searching and if they are successful in finding this data. SECOORA users are generally successful when looking for data, primarily data products. They take advantage of the free text search option which has an auto-complete component. However, due to the search being automatically completed, users are only guided to queries where results exist. Hence, results from the analysis may be skewed towards a higher success rate. The D/SWG was unable to identify how often consumers were using spatial, temporal or access method searches because that information was not part of the search logs.

Users generally searched DIVER within their geographic region with a diverse set of terms. The keyword search was used as a primary tool and free text filters were used to refine search results. The analysis did not include the Guided Custom Query offered by DIVER.

Technology advancements such as Google's Dataset Search and the Department of Energy (DOE) National Energy Technology Laboratory's (NETL) SmartSearch are useful tools for improving searchability/discovery of data. The Google Dataset Search requires dataset landing pages and effort on the repository side of data management. DOE NETL SmartSearch is a self-building index of data available on the web, using custom web-scraping algorithms to automatically mine data. The SmartSearch system requires no work on the repository end other than making data accessible and generally findable via search engines. The D/SWG organized a webinar with the SmartSearch team and the Restore Councils CMAP to explore the tool. (See https://crrc.unh.edu/DWH_Long-Term_Data_Management for documents.)

The D/SWG had the following recommendations:

1. Extend the analyses of search logs to a wider variety of repositories.
2. Findings from search log analysis should inform a search term bank (e.g., identify common keywords, proper names).

3. Data should be made available via 'landing pages'
    a. Allow for multiple levels of analysis (e.g., pageviews, referrals)
4. Demonstrate value of search logging mechanisms and define useful improvements
5. Some ancillary search methods (e.g., geofilter, time period) may no longer be needed due to low usage.

IV. Panels

Stakeholders' Perspectives

The first panel included a diverse set of perspectives with respect to data use and data management. Each panelist provided feedback on the Working Groups' progress and how their organization is moving forward with respect to DWH LTDM. The panel members included Jessica Henkel (RESTORE Council), Lauren Showalter (NAS Gulf Research Program), Eric Weissberger (NOAA NMFS Restoration Center) and Michele Jacobi (NOAA ORR ARD Spatial Data Branch).

Jessica Henkel thought the work from all of the working groups was very useful and a lot of progress had been made since the inception. More collaboration between the groups would have improved messaging/recommendations. The RESTORE Council is at the beginning of funding based on Gulf States impact formula (e.g., miles of shoreline oiled, "Bucket 3") and is thinking through future steps regarding RESTORE Council data. The Council does not plan to develop an independent data repository for RESTORE Council funded projects, and will be asking their members to use already existing data repositories.

Eric Weissberger and Michele Jacobi provided perspectives based on NRDAR practice, including data collected during the assessment phase as well as ongoing restoration-related data collection. The NOAA NRDAR process involves working with responsible parties, who sometimes have their proprietary systems, to better integrate their data into NOAA data and visualization systems. DIVER is a NOAA NRDAR data warehouse system and strives to be interoperable at a project and system-to-system level to allow for interactions and data transfers. The data coming into DIVER is diverse and must be available for a diverse set of end users. Integrating restoration monitoring data into DIVER has been challenging because data sets are not always submitted in the appropriate format. Each organization has their own mandate, target questions and tools used to answer specific questions, as well as long-term data collection procedures that often differ from DIVER data formats. This hinders data analysis and the programmatic review process, the process by which the Cross-Trustee Implementation Group Monitoring and Adaptive Management Working Group (Cross-TIG MAM) is reporting NRDAR-funded restoration progress to the public. Data generators should thus be encouraged to follow a defined set of data reporting standards.

By definition, NRDAR evaluates natural resource injuries after a hazardous substance is released to the environment. Baseline conditions must be referenced to evaluate the extent of the injury. This is true for injuries to ecological resources as well as the socio-economic and recreational services they provide. Data for NRDAR cases requires detailed, quality metadata in the proper format. Assessment data should be detailed and include a description of how the data are collected and, given the litigation context, a forensic copy of the stored data sets. Data generators should be aware of the purpose of their data, and standards necessary for compliance prior to collection. The federal government has data compliance standards that are not necessarily shared by other non-federal organizations. This difference can make

certain data unusable in litigation or settlement proceedings. Additionally, NRDAR data sets need to be stored long-term which requires resources and connections to the proper organizations. Regardless of the intent of the data set (e.g., academic, restoration, response, assessment), it should not be lost on hard drives and stored in isolation. One goal of DIVER is to provide public access to data sets now and in the future.

Lauren Showalter represented the perspective of the National Academies of Science, a funding entity. NAS GRP's structure and oversight tends to differ from other funding agencies because their goal is to identify repositories for their projects to store data sets. Their next phase of funding will be to address larger issues surrounding the GOM; these larger projects generally rely on manipulation and integration of data, stressing the importance of data system interoperability. Therefore, internal and external interoperability of systems is important to facilitate multi-faceted projects (e.g., socio-economic, environmental). In order to answer these larger questions, data integration is necessary.

The panelists agreed that data standards would benefit data providers. Progress on data standards could be achieved through raising awareness that the GOM is a shared, unified ecosystem and data should be interoperable and available at a regional level. The panel identified a preliminary list of 'grand' challenges for data management, which include:

- Making data discoverable for data users (e.g., general public)
    - Including multiple user groups with different needs and uses
    - Caveat for proprietary data
- Preserving data sets (i.e., not losing data)
    - Maintain provenance of data
- Using data to inform future science and improve decision-making by resource managers
    - Today's data will become baseline data for future
- Raising awareness of environmental problems/solutions across state boundaries
    - Publicizing and promoting the benefit of exposing data to other communities
    - Developing a two-page document to show benefits of data management and data system integration at all levels and across multiple disciplines (e.g., ecological, economic, human dimensions)
- Maintaining the life cycle of data needs (e.g., description of data, archive and preservation, accessible/discoverable data, reuse and transformation)
- Distinguishing between discrete data and integrated data – both may be valuable
    - Integrating requirements into data collection and management.
    - Not mutually exclusive approaches
- Building consensus, development or adoption, and using standard data formats to preserve data over time
- Documenting data management guidance and enforcing the use of guidance document by funders (e.g., two-pager best management practice for data management)
    - Making standards a funding agency requirement
- Establishing consistent data management plans

Data Systems' Perspectives

Panelists gave a short summary of their system and their response to the working group activities. The participants included: Sharon Mesick (NCEI), Benjamin Shorr (DIVER), Rosalie Rossi (GRIIDC), Craig

Conzelmann (CIMS), Laura Shumway (US EPA, WQX), Shin Kobara (GCOOS), and Debra Hernandez (SECOORA).

NOAA's **NCEI** (National Centers for Environmental Information) is a service resource for stakeholders (e.g., international, federal), and a steward of environmental data. Due to congressional mandates, NCEI is tasked with building a long-term coastal data record. It is also required to provide a central repository to manage data collections, and information services of the various GOM DWH restoration activities. NCEI is a data archive. Once data gets integrated into the system, it remains there and is accessible for the future. Any data that goes into NCEI archive requires a metadata record. The data is categorized into a certain storage location depending on the metadata. The data set can be used for data integration, processes and services. There are typically multiple points of access for the same data set within the archive. The goal is to reduce data complexity, potentially reducing data management costs, facilitating preservation and enabling data integration. The best way to minimize data complexity is to inform generators of standards and best practices and train them how to document data collection to improve discovery and access. By providing best management practices and quality documentation, data can be integrated into a variety of systems, thus increasing the value of data sets. NCEI goals align with those of the DWH LTDM group. Data user workshops are generally scheduled as a form of user assessment. Feedback can also be contributed through comments sent to a standard email address, and the use of customer relationship tools like Salesforce. Salesforce sorts user inquiries into groupings of information that can then be addressed and responded to on a strict time scale. In addition, NCEI is working with NOAA Sea Grant to improve community outreach within regions using surveys.

NOAA's **DIVER** (Data Integration, Visualization, Exploration, and Reporting) application is a data warehouse and query tool used for NRDA-related response, assessment, and restoration data and historical data. DIVER houses multiple types of data (e.g., samples, bioassays, oceanographic, telemetry). Each data category contains core fields as well as fields specific to each individual data category. The core fields are the same for each of the categories and provide overall relationships between data types. The DIVER Environmental Data Specification document describes the fundamental data scheme structures and data exchange methods. Details include: field information, valid values, core fields, and general submission guidelines for structured and unstructured data and metadata. Providing common formats for data categories allows for a smoother transition and efficiency when integrating into the DIVER system. The framework of DIVER is shifting towards increased support for data services for machine readable data, data exchange, and creating discrete data packages; this is likely to incentivize data generators to provide a higher quality of data because the discrete packages can be cited as DOIs. An ongoing challenge has been obtaining and creating high quality metadata records to make data sets applicable to assessment as well as restoration efforts. Focus on data packages and best management practices improves the flow of data sets into NOAA archives.

GRIIDC (Gulf of Mexico Research Initiative Information and Data Cooperative) was initially created as a data repository for all GoMRI-funded projects. The master research agreement between BP and the Gulf of Mexico Alliance that established GoMRI included a requirement that all data must be made publicly available. The GRIIDC data repository was being built as data was deposited into the system with the goal to ingest data and make those data sets readily available for researchers. Standards and descriptions became more important over time when storing, housing and using data sets. The current focus of GRIIDC is to improve searchability, which requires high quality assurance of metadata and

inquiry. The D/SWG also started this process and is a good step towards informing the development of the system structure. Another focus of GRIIDC is training researchers to include descriptive information (e.g., data sets can stand alone). Therefore, prior to data collection, researchers need to be informed of best management practices for data documentation. Currently, GRIIDC does not use surveys as a way to assess user feedback; however, this is something that could be done in the future. In many cases, the people using GRIIDC are GoMRI researchers who know what they are searching for; however, GRIIDC is receiving non-GoMRI data and is working to improve search for a wider audience. GRIIDC receives reports from users to inform internal decision making with respect to download statistics, searching statistics and standard vocabulary. Moving forward, GRIIDC's focus will be on outreach programs to continue informing the public about GRIIDC and data management best practices, and incorporation of Google analytics into their search tool.

The State of Louisiana's **CIMS** (Coastal Information Management System) provides a geospatial, tabular database and document access to the Coastal Protection and Restoration Authority's suite of protection and restoration projects. Following the June 2017 DWH LTDM workshop and the creation of DIVER's Environmental Data Specifications documents, CIMS has been able to convert data sets into a format ingestible by DIVER. The formation of data standards provides a foundation upon which the rest of data management in the future can build upon. Improved data standards will minimize interoperability challenges and help address searchability concerns. Data collection is regimented (SOPs) for how the data is collected and who collects it. Once data sets come into CIMS, state employees are responsible for QA/QC, (i.e., human oversight), making sure data is in the proper format and then approving it for public access. Overall, collaboration and communication between system owners can improve interoperability challenges using common data formats and standards.

Under the Clean Water Act (CWA), state, tribal and federal agencies are required to monitor water quality in lakes, streams and rivers. This system is the largest portal for water quality monitoring data. It uses the **WQX** data format to share data records. WQX is not a primary data steward, therefore, it does not manage QA data nor does it store metadata. WQX was created to improve data accessibility and data reuse. The EPA is currently working on establishing a DOI for every data record. The system tracks data downloads and where they are being pooled, thereby, providing value added services and fulfilling user needs. WQX operates using a machine-to-machine structure which is facilitated by providing data formats/templates to users and best practices for unique user groups. There are resources available on the WQX website such as training videos, user guides and formatting, best practices and metadata guidelines to improve the user's experience.

GCOOS (Gulf of Mexico) Coastal Ocean Observing System data is quality assured and available via data compliance. The data management foundations for GCOOS include community-based vocabularies, data formatting based on IOOS/NCEI requirements, data services direct access via interactive interface, and using an internet-based platform (e.g., http), and an open data policy. The GCOOS data portal is currently undergoing upgrades to improve foundational characteristics to increase site accessibility.

SECOORA (Southeast Coastal Ocean Observing Regional Association) offers their services to members and stakeholders within the four southeastern states (North Carolina, South Carolina, Georgia and Florida) to monitor the coasts and oceans. IOOS (Integrated Ocean Observing System) funding has required SECOORA to introduce standards and interoperability requirements. In addition, data accessibility is currently a focal point to enable users to find the data sets for which they are searching.

SECOORA gets feedback from user assessments to help prioritize where data is collected and how to focus funds for data collection. Most communities are focusing on their specific region, but regions have a wide variety of data, which makes accessing a specific data set challenging. Moving forward, data management and interoperability will continue to be a priority for SECOORA to ensure data is readily available to users. SECOORA regularly performs user assessments; in some instances, stakeholders or Principal Investigators (PI) of projects come forward with specific needs to be addressed. Every five years, a global assessment is performed for product development, providing a forum for stakeholders to give their input and feedback on the data system.

V. Breakout Session 1

The first breakout discussion consisted of breakout groups on the Data Management Standards, Interoperability, and Searchability/Discovery Working Groups.

Each of the three breakout groups filled out a template to discuss the recommendations put forth by the working groups, and provided feedback regarding:

- The significance of their findings.
- What was missed, added, and/or modified?
- What challenges or recommendations were cited by the Working Group?

The completed templates for each of the breakout groups are located in Appendix E.

Data Management Standards

The first recommendation of the Data Management Standards breakout group was to adopt/adapt/develop a data exchange format for data systems. This recommendation was of high significance and posed challenges including: identification of common data fields and narrowing down which fields should be included in the data exchange format. In order to accomplish this, the data exchange format must be independent of existing entities/data systems. A set of common data exchange fields among existing systems should be determined, and the results of this effort be made publicly available. Creation of a data exchange format allows data owners to crosswalk fields. Before moving forward with a data exchange format for all existing systems, the group recommended that a small-scale study be done to compare two existing data systems (e.g., WQX and DIVER).

The second part of this process would include the creation of a consumable, readily accessible document outlining the integration/exchange standards and protocols recommended for use by the public. This guidance document could be available to data generators, in a way, similar to that of NOAA's Field Observation Electronic Data Deliverables (EDD) templates. There was discussion about the order of developing the data exchange format and the data integration exchange standards, or if these efforts could occur simultaneously. Creating a standard set of recommended data integration/exchange standards was of high significance, but should be focused on specific categories of standards. Additionally, further clarification regarding the definition of data integration/exchange standards is required. Because this effort has been started by the Cross-TIG MAM (Cross-Trustee Implementation Group Monitoring and Adaptive Management) working group, consultation with that group is advised in order to avoid overlap in project scopes and to remain consistent with respect to vocabularies and

standards. The group thought parsing out fields vs. interoperability data standards would be a challenge, but if the quality of metadata were improved, then it would be a worthwhile venture.

The third recommendation was to finalize products developed by the SWG and make them available to users. One concern is identifying who the users of these documents would be and if the documents should be prepared with an expansive or refined concept of who the users are (e.g., agencies vs. public). A challenge would be to find ongoing stewardship of the documents in order to maintain their accuracy over time (e.g., annual updates).

Interoperability

The first recommendation of the Interoperability breakout group was to establish standardized data methodologies to maintain consistent processing of data sets (high priority). The second step was to provide standardized data services to ensure consistent data delivery (low priority). Thirdly, if community standards already exist, they should be used as a starting point. If community standards do not exist, a crosswalk should be built to improve integration of data sets to increase value of data (medium priority). The group recommended that if tools are developed, they should be scalable to include data management at all levels (e.g., local, national). Documentation and communication of best management practices are essential to improve integration of data across systems, and this would best be achieved through cross-pollination between working groups in order to increase the overall group perspective/background (high priority). The breakout group recommended that the commonalities matrix, developed by the Interoperability Working Group (Appendix E), be modified to make it a weighted system. Challenges for these recommendations and subsequent steps would be to get buy-in from data system owners, provide funding, maintain an engaged group of participants, identify the correct audience, and find support for this effort at local, state, and federal levels. Due to the turnover of positions within organizations, it is essential to build institutional resilience in order to keep the process moving forward and to produce viable outcomes.

The second recommendation was to advocate that interoperability become a grant/program/system requirement (high priority). This would require communication of the benefits of interoperability between systems at a high level (e.g., risk management, improved decision-making).

Execution of an end-to-end interoperability pilot program was also discussed to show a real-world application of data integration (high priority). The pilot project would require translators to take disparate metrics and make them usable (low priority). The pilot project could begin with two relatively compatible systems (e.g., CIMS and DIVER), and then work to broaden the scope to include multiple sectors and disciplines (medium priority).

Discovery/Searchability

The first recommendation regarding data discovery/searchability was to evaluate smart tools (e.g., machine learning) to see if they can be leveraged. Implementation of this recommendation is dependent on vocabulary, indexes and availability of metadata and all related information, which may pose challenges (high priority). Identification of community driven vocabularies would improve searchability of systems and could be started by leveraging existing vocabularies (e.g., GOMA). This is a high priority, but requires consensus and buy-in which can be challenging when multiple organization are participating.

A directory of systems should also be developed and made publicly available online to assist users when searching for data within a range of repositories. This system could be modeled off the Gulf TREE directory, but requires funding and an understanding of how Gulf TREE works internally (high priority). A data query satisfaction tool should also be created to inform system owners if users are finding information for what they are searching. This is a high priority, but requires development and buy-in from the user community.

To improve searchability, metadata and all related information should be available and indexed readily by advanced searches. This would facilitate exchange of data sets at a high level which would maintain the integrity of data sets over time. This effort is contingent on community-driven vocabularies and was ranked as a medium priority. Metadata should also be exposed by a web search tool (e.g., Google dataset search) that can index and discover data sets (medium-low priority). Additionally, encouraging metadata sharing through open sourced federated systems (e.g., CKAN, NOAA catalogue) will improve user access of data (high priority). Both of these recommendations are dependent upon standardization, which is a challenging task.

The group proposed that both raw and processed data be made available to users and viewed this as a high priority. Data types should also be linked together (e.g., integration of data categories over time, location, ontological network). This is a medium-low priority with many dependences that would evolve depending on user needs.

VI. Breakout Session 2

Part A: Translating Recommendations into Actions

The participants were divided into four breakout groups (A-D) to further refine the recommendations developed with respect to data standards, interoperability and searchability/discovery. Group A discussed recommendations for interoperability, data standards and searchability/discovery. Group B discussed data standards and searchability. Group C provided feedback on searchability recommendations and interoperability. And Group D discussed interoperability, data standards and searchability. Groups, consisting of participants with varying perspectives (e.g., data manager, data user) covered different topics as there was insufficient time to have each group discuss all topics. The four breakout groups developed: project timelines and cost estimates, identified possible project champions, the ease of achievement, and prioritization of recommendations. Each group also noted any dependencies and contingencies. The participants in each group consisted of: data systems representatives, funders, oil spill/emergency and NRDA responders, restoration representatives and participants from each of the three Working Groups. Each group was provided a template with previous recommendations (developed in Breakout Discussion, Group 1). The order of the topics on the template were arranged to ensure full coverage of each topic.

Interoperability Recommendations:

1. Develop standardized data methodology (e.g., consistent processing measurements, recording).
2. Provide standardized data services (use consistent data delivery.)
3. Use existing community standards and build a cross-walk; integrate data so they have meaning together.
4. Enable tools at all levels; local and national (scaling).
5. Document and communicate best management practices.

6. Marketability for future decision-making (i.e., risk management).
7. Advocate to communicate benefits of interoperability across systems.
8. Advocate for interoperability at high levels within organizations.
9. Execute an end-to-end Interoperability Pilot Program
    a. Show real world application (i.e., tell the story)
    b. Take disparate metrics and make them usable together
    c. Build across sectors and disciplines (e.g., energy, clean water, agriculture, future forecasts).

Data Standards Recommendations:

1. Adopt/adapt/develop a data exchange format for systems.
2. Develop a set of recommended data integration/exchange standards.
3. Make data integration/exchange standards & protocols publicly available, easily accessible via websites and up-to-date.
4. Make products developed by working groups readily available to users.
5. Update to system inventory documents annually.

Searchability/Discovery Recommendations:

1. Evaluate smart tools (machine learning, AI) to see if they can be leveraged for this group.
2. Identify community driven vocabularies – leverage existing (e.g. GOMA).
3. Directory of systems published online (finding the right repository) (Similar to Gulf TREE.)
4. Data query satisfaction (Did you find what you were looking for?) Yes/No?
5. Metadata to facilitate exchange at high level – make all related information available (analytical methods etc.) able to be indexed readily by advanced searches.
6. Exposing metadata in a way that a web search tool (e.g. Google dataset search) can index and discover.
7. Encourage metadata sharing through open source federated systems like CKAN.
8. Make both raw and processed data available (reports linked to data).
9. Linking of data types: Integration of data categories by time, location and ontological network.
10. Capture DOI metrics to inform data use (how many times something has been cited).

Group A

Group A discussed recommendations for interoperability, data standards and searchability/discovery. For a summary of Group A's breakout template see Appendix F.

Developing standardized data methodology is a high priority for interoperability and other recommendations are contingent upon this. Many of the recommendations (e.g., 1, 2, 3, 4, 6) have a dynamic timeline depending upon the completion of the other tasks and the quality of the final products. In order to make data standards more marketable and valuable, the tasks explained in steps one through five needed to be completed thoroughly. Execution of an end-to-end interoperability pilot project would be a proof of concept for 1-5. Completion of the pilot project would ideally show value of interoperability using a real-world example.

Group A suggested it was logical to develop a set of recommended data integration/exchange standards prior to creating a data exchange format. Both of these tasks were ranked as high priorities. Upon completing Steps 1 and 2, the data integration/exchange standards and protocols should be made

publicly available on accessible websites and kept up-to-date over time. Steps 1-3 all have a dynamic timeline based upon the skill set of the professionals performing the task and the fact that these efforts need to be maintained over time (e.g., annually). Steps 4 and 5 are based upon completion of the Working Group documents; once completed, the final products should be made publicly available and updated annually. These efforts are contingent upon volunteer time, effort and available resources.

The group noted that identifying community driven vocabularies was a high priority with low cost, medium ease of achievement and a relatively short timeline. This task could be completed prior to the others (e.g., using smart tools) because development of automated querying tools generally needs to be consistent with respect to terminology. Steps 6-8 were grouped together into one category, which was identified as best management practices. Establishing best management practices would encourage metadata sharing and improve transparency of the quality of data sets. Investigation of current practices on the ability for DOI metrics to inform data use was recommended, along with identification of return on investment opportunities using DOI's.

Group B

Group B discussed data standards and searchability (see Appendix F for the template). Group B thought adopting, adapting or developing a data exchange format for systems (including metadata standards) was of high priority, would be challenging to achieve, and could be done within five years, but at a relatively high cost. Step 1 in combination with step 2, developing a set of recommended data integration/exchange standards, could be started at an intensive workshop. The goal of the workshop would be to achieve consensus around the integration/exchange standards and the data exchange format, to be further developed and reviewed by the community after the workshop. It will be challenging to maintain the list of data systems annually.

Of the ten searchability/discovery recommendations, the use of metadata to facilitate data exchange at high levels was the greatest priority and required, in order to make related information available and able to be indexed readily by advanced searches. An example of this is developing/ implementing standards recommendations for metadata. This is a challenging task, with a moderate cost. Group B concurred with Group A that steps 6 and 7 were best management practices, and that DOI's present a potential incentive for data generators and remains a useful tool to improve data discovery.

Group C

This group provided feedback on searchability recommendations and interoperability (see Appendix F). Similar to Group A, identifying community driven vocabularies was a high priority recommendation; with a timeline between 1-5 years with a relatively moderate cost. In addition, Group C identified the use of metadata to facilitate exchange at a high level as a priority that could be initiated by incorporating more training for data generators. Group C identified exposing metadata so that a web search tool can index and discovery it, and encouraging metadata sharing through open source federated systems to be high priorities. These tasks are contingent upon data organization and transparency. Unlike Groups A and B, this group encouraged the use and capture of DOI metrics as a medium priority contingent upon quality metadata from individual data generators.

This group identified developing standardized data methodology, as the foundational interoperability task with a high/medium priority that would be challenging and take a long time to accomplish. Providing standardized data methodology would allow for more consistent processing. Additionally, the

group noted that developing recommendations for interoperable systems (e.g., best management practices), and documenting and communicating findings are a medium priority with a shorter timeline, lower cost and is more easily achievable. Group C noted this to be a future working group topic.

Group D

Group D discussed interoperability, data standards and searchability (Appendix F). The group denoted four recommendations as high priorities and rearranged the recommendations to include some as next steps/contingencies to move projects forward. Their first high priority recommendation was developing standardized data methodology (consistent with Group A) on a medium time scale (1-5 years), with moderate cost and LTDM Working Group Members as possible champions. Group D rated "use of existing community standards or building a cross-walk" as a high priority that could be championed by NOAA, GRIIDC, NAS, NCEI, LTDM, GOMA and some of the Common Operating Pictures (COPs) such as the Environmental Response Management Application (ERMA®). This effort would take a long time to make progress, would have an associated high cost and be relatively hard to accomplish. The steps to complete this effort are: (1) adopt/adapt/develop a data exchange format for systems; (2) develop a set of recommended data integration/exchange standards; and (3) make data integration/exchange standards & protocols publicly available, easily accessible via websites and up-to-date. Facilitated by outreach and presentation at conferences/Trainings. An additional high priority task is documentation and communication of best management practices. Group D thought this easily achievable, low associated cost, which could be accomplished within a year. This would inform previous recommendations and improve data integration across systems. Group D suggested to include "preferred, acceptable, not acceptable" when describing best management practices. The final high priority in interoperability was the development of translators, this would allow disparate metrics and make them usable together. This task would have a short timeline (<1 year), is relatively easy to achieve, low cost and could be championed by DWH LTDM.

Of the five recommendations within the Data Standards section, Group D rated four of them as high priority and a possible champion of the DWH LTDM. Adopt/adapt/develop a data exchange format (step 1) and developing a set of recommended data integration/exchange standards (step 2) were ranked as high priorities by Groups A, B and D (Group C did not complete the data standards section). Step 1 could be accomplished using test cases (e.g., WQX → DIVER), has a short timeline, is moderately expensive with a medium ease of achievement level. Step 2 is easily accomplishable, but dependent upon documentation and data exchange test cases, is moderately expensive and could be accomplished between 1-5 years. Group D noted that making data integration/exchange standards and protocols publicly available, easily accessible via websites and up-do-date is a high priority, but is contingent on Steps 1 and 2. It is easily achievable and has a low associated cost once the integration/exchange standards are developed. The group found that products developed by the working groups should be made available to users as a high priority, at low cost, with a medium level of achievement and could be done within 1-5 years.

The group ranked two recommendations within searchability as high priority: evaluation of smart tools to see if they can be leveraged by data system owners (contingent on common vocabulary). Community driven vocabularies should also be identified. This is a challenging task, with a moderately high cost and medium timeline (e.g., 1-5 years) that could be championed by DWH LTDM, NCEI and/or NETL.

Plenary Report Out: Breakout Group 2

The plenary report out provided an opportunity for all four groups to present their feedback and all participants to see the general themes. In addition to the report outs, the concept of a project tracker for GOM data systems and repositories was discussed in the plenary. The project tracker could be developed so that it captures high level information about projects and directs users to find applicable data. It was noted, however, that the existing DWH Project Tracker, developed by GOMA, is not currently designed to do so. The other major conclusion of the plenary discussion was the importance of adopting/adapting/developing a data exchange format for systems and developing a set of recommended data integration/exchange standards.

VII. Breakout Session 3

Part B: Prioritizing Recommendations/Actions and Next Steps

The goal of Breakout Session 3 was to focus on feasibility and implementation of the identified next recommendations. The Breakout Groups (A-D) were the same as in Session 2. The groups selected the top priority recommendations on searchability, data standards and interoperability.

Once the top three priority recommendations were determined, each group discussed: a) the required collaboration/integration needed to move the effort forward, b) how to encourage stakeholder buy-in, and c) objectives for the proposed work.

Group A

Group A developed three recommendations (Appendix G).

1. Conduct a pilot project to demonstrate successful interoperability between two existing data systems as proposed by the SWG. This would provide proof-of-concept with the goal of leveraging existing data systems to show a real-world application of two interoperable, readily-accessible systems. The LTDM Core Team should identify project ideas, detail the goals for the pilot project and work with system owners who are currently collaborating, or bring together new system owners with the focus on system integration.
2. Develop a framework to document standard interoperable and searchability goals, identifying existing projects, working with system owners and stakeholders whose data are used and testing what is known about the systems' interoperability. The pilot project should: show 'real world' application (i.e., tell a story); develop translators; build across sectors and disciplines; provide feedback on recommendations for standards and interoperability; and leverage community standards/crosswalk, where appropriate.
3. Identify community-driven vocabularies, starting with leveraging existing vocabularies (e.g., GOMA). This would be a volunteer effort to include those entities with a vested interest. Any documents produced could be reviewed and distributed by the LTDM Core Team.

The results of the LTDM workshop, current success stories between interoperable systems, and the future pilot project outcomes should be communicated broadly and used to tell a story about the importance of data accessibility and data integration.

Group B

Group B recommended four actions (see Appendix G).

1. Compile best practices for data systems, including: developing, populating, and exposing metadata, use and evaluation of DOI's, and application of controlled vocabularies. This compilation would be the result of:
    a. A survey of targeted outreach of stakeholders that would result in a work plan. This would be followed by a workshop book-ended by teleconferences. The product would be a Best Practices document.
2. Develop data exchange formats and standards, implementing the following steps:
    a. This group recommended that the format and standards follow the same pattern of survey/outreach, workshops book-ended by teleconferences and a deliverable of a data exchange format and accompanying standards.
3. Identify community-driven vocabularies to build upon previously identified vocabularies (e.g., EDDM). The deliverable for this recommendation is a list of vocabularies and how to apply them.
4. Advocate for interoperability at the funding level and upper management levels. Including, developing talking points (e.g., elevator speech), a two-pager and/or a report.

The collaboration/integration required, stakeholder buy-in and future work are the same for the first three recommendations. Collaboration/integration should include focused working groups with dedicated and committed participants who are representative of stakeholder needs. The group should be empowered to make decisions and consist of diverse and dedicated stakeholders including decision makers, the broader community of providers/system owners and data users. Advocates should include all workshop participants and coordination with existing communication organizations (e.g., Sea Grant).

Group C

Group C's recommendations covered data standards, metadata and advocacy (Appendix G).

1. Develop and/or adapt a data exchange format and development of a set of recommended integration/exchange standards. The group noted the collaboration would be necessary between all data systems, and that representatives from the systems would need to draft the standards. This approach could leverage existing work and start by focusing on a suite of core fields for all data types across the systems. Data systems owners, data generators and the LTDM working groups could contribute to this effort. The effort would advance the ability to integrate data between systems and enhance a suite of products for data blending.

2. Interoperability and searchability are dependent upon metadata standards. Quality would be improved through: (1) identifying community-driven vocabularies; (2) making related metadata available and readily indexed by advanced searches; (3) exposing metadata to allow web search tools to discover them; and (4) encourage metadata sharing through open-source federated systems. Collaboration between system owners, offices, federal and private sectors is necessary to further progress. Additionally, development of common themes and/or messaging for education and outreach purposes would help users when searching for particular data sets. Future work includes creating a list of core metadata information, identifying gaps and standards, and educating stakeholders (e.g., data generators). The recommended actions are to develop a DWH restoration vocabulary and implement it.

3. As with the other groups, Group C stressed advocacy at high levels within organizations regarding the benefits of interoperability across systems. This would include data system owners developing common themes/messaging, leadership buy-in and incorporation of messaging/themes into the agencies' core missions. This could be done using a communication packet describing the need for established community goals (e.g., standards).

Group D

Group D developed four actionable recommendations (Appendix G), these include:

1. Document and communicate best management practices, starting with a compilation of those developed by LTDM SWG. Future work includes crosswalk of best management practices and creation of a guidance document. The guidance document would outline the language for best management practices listing practices that are preferred, accepted and not accepted.
2. Develop and/or adapt a set of recommended data integration/exchange standards and a data exchange format should be developed for systems. System developers and system owners would need to be involved to coordinate test cases. Test cases could be leveraged from existing work (Appendix G cites specific examples) or new avenues of collaboration could be identified. Core fields need to be identified and documentation should be published to encourage data systems to use them. This would allow for integration of data across studies/programs and broader use of data sets for studies/questions.
3. As with the other groups, Group D recommended identification and compilation of community-driven vocabularies, including international vocabularies, where appropriate. The result would be a master list to of core fields detailed to an appropriate level (e.g., optional vs. required).
4. Group D also recommended completing and publishing all of the documents developed by the DWH LTDM working groups, making them available to users and sharing them with the GOM data community.

Plenary Report Out: Breakout Group 2

Feedback from the plenary was positive as each group produced similar recommendations for future steps. During the plenary report, it became apparent that the four groups had arrived at very similar conclusions.

Across all breakout groups, the overarching recommendations were:

- Finish and publish the DWH LTDM Working Group documents (make these broadly available);
- Develop a data exchange format and data integration/exchange standards;
- Identify community vocabularies;
- Use a pilot project to show the importance of system integration; and,
- Communicate/advocate with respect to data management best practices/standards, Working Group documentation and outcomes from DWH LTDM workshops.

VIII. Workshop Outcomes

The workshop concluded with a final presentation by Lauren Showalter (NAS GRP) describing the Working Group outcomes and the plenary sessions' consensus with respect to recommendations for future actions (Figs. 1 and 2). The working group laid the groundwork for productive workshop

discussions, created useful products that can be expanded and improved, and identified actionable items. The future actions that were delineated during the plenary discussions were:

(1) Adoption/adaptation/development of data exchange format/materials;
(2) Identification of vocabularies;
(3) Development of best management practices for data integration;
(4) Pilot project to use as proof of concept of data systems interoperability; and
(5) Communication/advocacy of successes of interoperable systems.

The next steps for the DWH LTDM Core Team should be to develop goals and subsequently write a work plan for each of the actions listed above in the next several months. Outreach to data users, generators and managers should include education/training about available products and show the value of efforts completed by the DWH LTDM group. The participants agreed that action items are important, development of work plans are necessary to outline project scopes, in order to gain support/buy-in from the GOM DWH community and keep tasks moving forward.

Immediate Action Items

- Finalize the DWH LTDM Working Group documents and publish on CRRC website for public accessibility.
- Circulate the document summarizing the federal agency requirements for data management.
- Compile information regarding on-going collaboration between data systems (i.e., use real world examples to show interoperability proof of concept).

Core Team Recommendations for Next Steps

After the workshop adjourned, the DWH LTDM Core Team and working group team leads drafted the next steps listed below.

(1) Data Exchange Framework
    i. Objectives
        1. Allow for integration of data across studies/programs
        2. Allow for broader use of data sets for studies/questions
    ii. Work plan
        1. Test cases (e.g., WQX←→DIVER, CIMS→DIVER, CIMS→WQX, DIVER←→NCEI, GRIIDC←→NCEI)
        2. Identify core fields and advocate for data systems to use them
(2) Vocabularies
    i. Objectives
        1. Achieve identified searchability recommendation
        2. Encourage community adoption of recommended vocabularies
        3. Will inform metadata and environmental data management
    ii. Work plan
        1. Compile existing systems vocabularies (including international vocabularies, where appropriate)
        2. Build a master list of vocabularies including recommendations on required vs. optional adherence
(3) Best Management Practices

i. Objectives
    1. Develop clear language for best data management practices such as: Preferred, Acceptable, Not Acceptable (e.g., pdfs)
    2. Best management practices help to inform previous recommendations, including: metadata, known file types, files sizes, model run guidance
    3. Crosswalk of best management practices and create guidance document
ii. Work plan
    1. Compile existing best management practices using LTDM SWG for each of the participating programs

(4) Advocacy/Communication
i. Develop a two-pager describing the inherent, long-term value of the DWH LTDM effort
ii. Identify audiences for advocacy and primary partners
iii. Use proof of concept from pilot projects/test cases to show value of DWH LTDM efforts

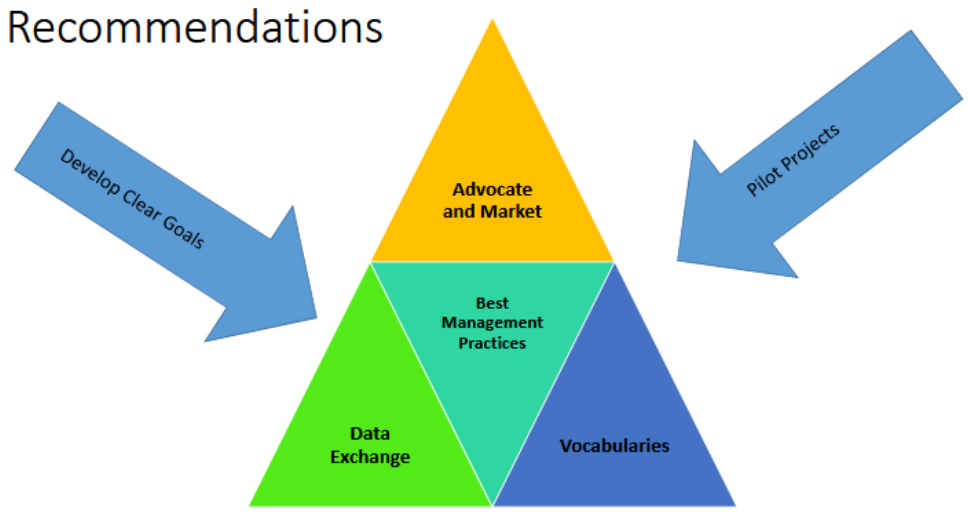


Figure 1. Primary Recommendations from DWH LTDM Workshop

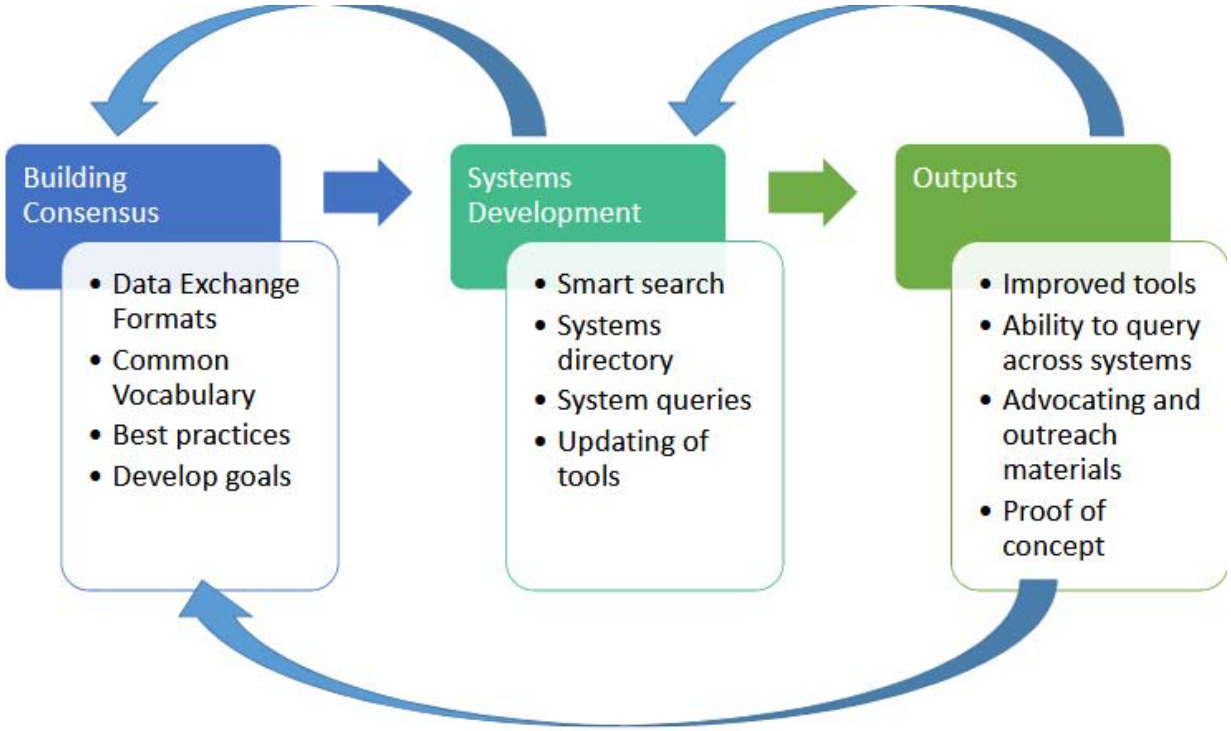


Figure 2. Path Forward with DWH LTDM Coordination

IX. Appendix

Appendix A

Participant List

COORDINATION OF DWH LONG-TERM DATA MANAGEMENT: THE PATH FORWARD

DECEMBER 4 - 5, 2018

PARTICIPANTS (revised post-workshop)

Ursula Arnold
Chevron Upstream IM Champion
ursula@chevron.com

Courtney Arthur
IEc
carthur@indecon.com

Jennifer Bauer
National Energy Technology Laboratory
jennifer.bauer@netl.doe.gov

Jonathan Blythe
BOEM
jonathan.blythe@boem.gov

Julie Bosch
NOAA NCEI
julie.bosch@noaa.gov

Harris Bienn
The Water Institute of the Gulf
hbienn@thewaterinstitute.org

Jay Coady
NOAA ORR ARD, Spatial Data Branch
jay.coady@noaa.gov

Craig Conzelmann
USGS - Wetland and Aquatic Research Center
conzelmannc@usgs.gov

Stephen Delgreco
NOAA NCEI
blackswaninnovations@gmail.com

Nicolas Eckhardt
NOAA ORR ARD, Spatial Data Branch
nicolas.eckhardt@noaa.gov

Melissa Gloekler
CRRC, UNH
mdx52@wildcats.unh.edu

George Graettinger
NOAA ORR, Spatial Data Branch
george.graettinger@noaa.gov

Robert Gruba
Mississippi Dept of Marine Resources
robert.gruba@dmr.ms.gov

Jessica Henkel
Gulf Coast Ecosystem Restoration Council
jessica.henkel@restorethegulf.gov

Debra Hernandez
SECOORA
debra@secoora.org

Lei Hu
Dauphin Island Sea Lab
lhu@disl.org

Christina Hunnicutt
USGS
hunnicuttc@usgs.gov

Michele Jacobi
NOAA ORR ARD, Spatial Data Branch
michele.jacobi@noaa.gov

Ann Jones
IEc
ajones@indecon.com

Kathryn Keating
Gulf Coast Ecosystem Restoration Council
kathryn.keating@restorethegulf.gov

Nancy Kinner
CRRC, UNH
nancy.kinner@unh.edu

Shinichi Kobara
Texas A&M University, GCOOS
shinichi@tamu.edu

Kirsten Larsen
NOAA NCEI
kirsten.larsen@noaa.gov

Matt Love
Gulf Coast Ecosystem Restoration Council
matt.love@restorethegulf.gov

Kathy Mandsager
CRRC, UNH
kathy.mandsager@unh.edu

Kathy Martinolich
NOAA NCEI
kathy.martinolich@noaa.gov

Sharon Mesick
NOAA NCEI
sharon.mesick@noaa.gov

Mike Peccini
NOAA Restoration Center
mike.peccii@noaa.gov

Steven Ramsey
Social & Scientific Systems, NIH GuLF Study
sramsey@s-3.com

Richard Raynie
Coastal Protection and Restoration Authority of Louisiana
richard.raynie@la.gov

Ali Robertson
GOMA, Data & Monitoring Community of Practice (CMAP)
ali.robertson@gomxa.org

Jesse Ross
CRRC, UNH
jjo94@wildcats.unh.edu

Rosalie Rossi
GRIIDC
rosalie.rossi@tamucc.edu

Shelby Servais
U.S. Fish & Wildlife Service
shelby_servais@fws.gov

Benjamin Shorr
NOAA ORR ARD, Spatial Data Branch
benjamin.shorr@noaa.gov

Lauren Showalter
Gulf Research Program, NAS
lshowalter@nas.edu

Laura Shumway
USEPA (WQX)
shumway.laura@epa.gov

Thomas Strange
Mississippi Restoration Team
tstrange@cce.ns

Kevin Suir
USGS Wetland & Aquatic Research Center
suirk@usgs.gov

Dan Van-Nostrand
NOAA NMFS Restoration Center
dan.van-nostrand@noaa.gov

Zhankun Wang
NOAA NCEI
zhankun.wang@noaa.gov

Eric Weissberger
NOAA NMFS Restoration Center
eric.weissberger@noaa.gov

Kyle Wilcox
Axiom Consulting AOOS Team
kyle@axiomdatascience.com

Caitlin Young
NOAA RESTORE Science Program
caitlin.young@noaa.gov

COORDINATION OF DWH LONG-TERM DATA MANAGEMENT: THE PATH FORWARD

DECEMBER 4 - 5, 2018

Appendix B

Workshop Agenda

COORDINATION OF DWH LONG-TERM DATA MANAGEMENT: THE PATH FORWARD

DECEMBER 4 - 5, 2018

AGENDA

0800 Registration

0830 Welcome/Logistics – *Nancy Kinner, Coastal Response Research Center*

0845 Background of DWH LTDM Effort - *Benjamin Shorr, NOAA ORR*

0900 Workshop Objectives and Format

0915 Self-Introductions of Workshop Participants

0930 Overview of Survey Results - *Jessica Henkel, Restore Council*

0945 Reports from Teams/Working Groups

- DWH LTDM Data Management Standards Working Group

1015 *Break*

1030 Reports from Teams/Working Groups

- DWH LTDM Interoperability Working Group
- DWH LTDM Discovery/Searchability Working Group

1130 Perspectives from Stakeholders (Panel)

- RESTORE Council - *Jessica Henkel*
- STATES - *Libby Fetherston, Robert Gruba & Steve Jones*
- NAS GRP (Gulf Research Program) - *Lauren Showalter*
- Emergency Response - *Charlie Henry, NOAA Disaster Response Center*
- NRDA (Natural Resource Damage Assessment) - *Eric Weissberger & Michele Jacobi, NOAA*

1230 *Lunch*

1315 Perspectives from Data Systems (Panel) - Response to working group activities

- NCEI – *Sharon Mesick*
- DIVER – *Benjamin Shorr*
- GRIID-C - *Rosalie Rossi*
- CIMS - *Craig Conzelman*
- USEPA WQX - *Laura Shumway*
- SECOORA & GCOOS - *Shin Kobara*

1430 *Break*

1445 Breakout Discussion Group 1: Input on Working Group Findings/Recommendations

- Significance of findings
- What was missed, additions, modifications?
- Challenges

Breakout Discussion Group 1 will consist of 3 breakout groups:

- Data Management Standards
- Interoperability
- Discovery/Searchability

1600 Plenary Report outs

1700 Adjourn

Day 2

0830 Recap and Recalibration – *Nancy Kinner, Coastal Response Research Center*

0900 Breakout Discussion Group 2: Translating Recommendations into Action

A. Working Group Outcomes and Recommendations: Next Steps: Short, Medium and Long Term

- Priorities
- Ease of achievement, cost, organizational mandates needed for:
    - Data Management Standards
    - Interoperability
    - Discovery/Searchability

1045 *Break*

1100 Plenary Report Outs

1200 *Lunch*

1245 Breakout Discussion Group 3: Translating Recommendations into Action

B. Feasibility and Implementation of Next Steps and Recommendations

- Collaboration/integration needed
- Stakeholder buy-in
- Future work/objectives

Breakout Group 2 will consist of 4 breakout groups each with a mix of:

- Data Systems Representatives
- Funders
- Oil Spill, Emergency & NRDA Responders
- Restoration Representatives
- Representatives of each of the 3 Working Groups

1415 *Break*

1430 Plenary Report Outs

1530 Path Forward - *Lauren Showalter, LTDM Core Team*

1615 Closing Remarks - *Mike Peccini and Benjamin Shorr, LTDM Core Team*

1630 Adjourn

2018 Workshop Objectives:

1. Continue to foster collaboration among the Gulf of Mexico partners involved in restoration planning, implementation and monitoring (e.g., RESTORE Council, GOMRI, GRIID-C, NAS Gulf Research Program, GOMA, NFWF, and Gulf Environmental Benefit Fund).
    a. What does collaboration "look like"? What does success look like?
2. Report out on progress towards goals of 3 Working Groups (see below).
    a. What progress has been made and summarize findings;
    b. What has been challenging;
    c. Were these good working groups and goals? (Did your group goals shift and if so, why?)
    d. Recommendations for best practices / suggestions to address challenges
3. How to translate Working Groups coordination and research into action.
    a. What concrete steps should be taken to foster collaboration and data integration? (Identify short, medium, and long-term actions, what is their priority? Ease of achievement, cost, and organizational mandates/priorities?)
        i. Data Systems perspective: how do we use what we have learned to improve the practice of systems working together (this will then make it easier for data providers to understand best management practices)
        ii. Funders perspective: how to simplify/improve workflow for data providers to improve use of standards and best management practices
    b. What is the future direction of this Working Group or other forum?

Appendix C

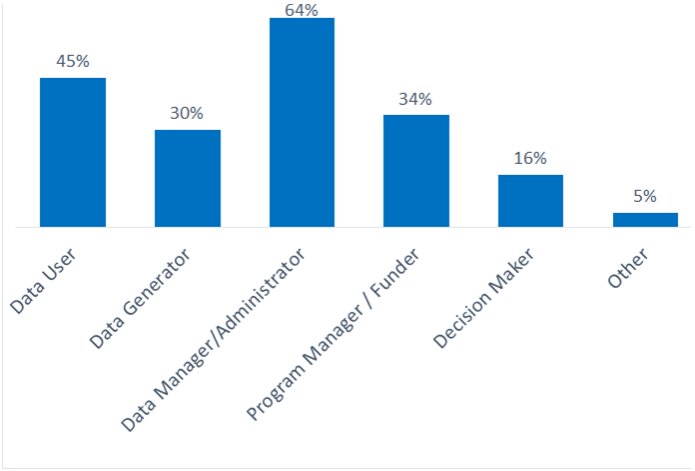
Pre-Workshop Survey

Looking Forward:
Long Term Data Management in the Gulf

DWH Long Term Data Management Workshop Participant Survey Responses

Survey Responses

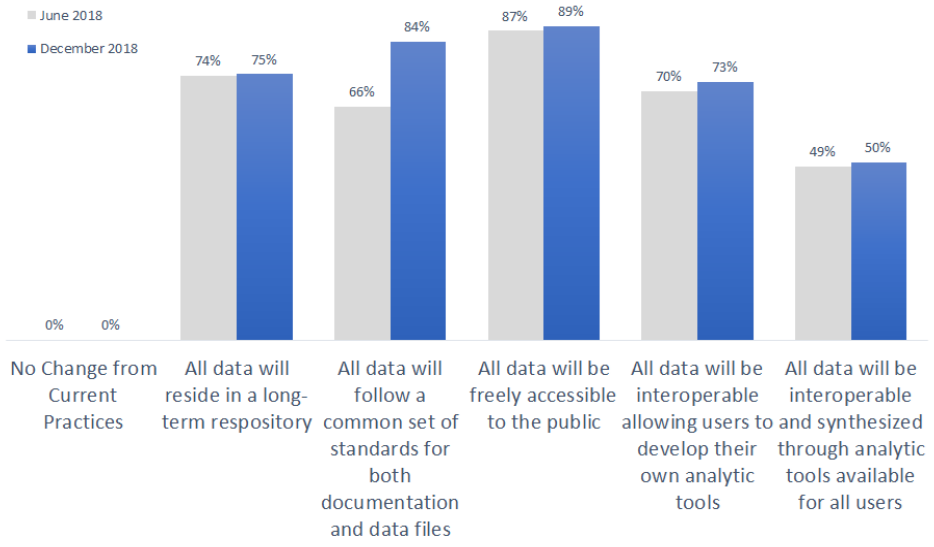
How would you best describe Yourself?
(Select all that apply)



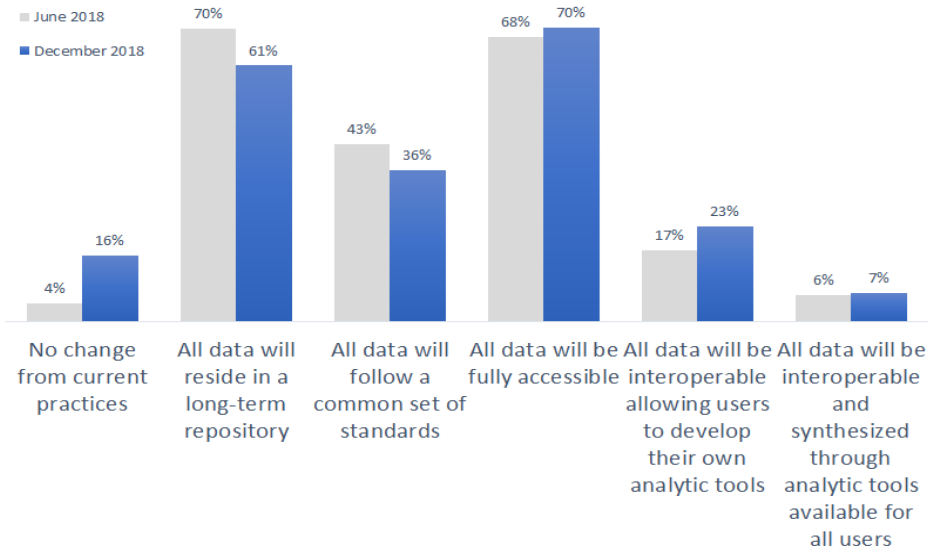
45% of survey respondents chose two or more roles to describe their work.

2

What do you want from Gulf research/monitoring data 15 years from now? (Select all that apply)



Realistically, what do you think Gulf research/monitoring data will have achieved in 15 years? (Select all that apply)



What do you see as the biggest challenge to data management in the Gulf?

Key Themes

Variation in Data Systems & Standards	Collaboration & Communication	Funding	Variation in Goals & Missions	Leadership	Attitudes about Data Management

5

What do you see as the biggest challenge to data management in the Gulf?

Variation in Data Systems & Standards	"There are so many different entities collecting, storing, and using data. Therefore, there is a lot of redundancies, inefficiencies, and barriers to synthesis and comprehensive analysis of restoration progress." "Agreeing to a set of common standards and then addressing how to transition from historic methods to recommended."
Funding	"Variety of governance and legacy (pre-existing) programs with resistance to change. Lack of incentive &/or funding to change." "Sustainability of data management efforts with projects and funding winding down" "Lack of funding. Lack of leadership from funding agencies".

6

What do you see as the biggest challenge to data management in the Gulf?

Collaboration & Communication	"Coordination across different offices and sectors to work on data management to establish standards, interoperability, data discovery etc.." "Coordination and acceptance of developed standards, data warehousing for access and funding support for longterm data management.
Variation in Entity Goals / Missions	"Widely varying scale, resources, and goals among Gulf-focused projects" "....The fact that the majority of data repositories/providers have different missions and there is not money to encourage collaboration and allow those working on them the time to do the work needed for collaboration"

7

What do you see as the biggest challenge to data management in the Gulf?

Leadership	"No single entity responsible for uniform data management practices" "There is no clear decision maker. We need to decide where we want to go and work to get there, rather than seeing where things already overlap and hope that people might make use of that information."
Attitudes toward Long Term Data Management	"I think data managers are convinced that data interoperability won't or can't work. We are limiting ourselves with this perspective." "Data management is not a high priority task for data generators, and other demands on their time win over, even if they philosophically agree with data sharing."

8

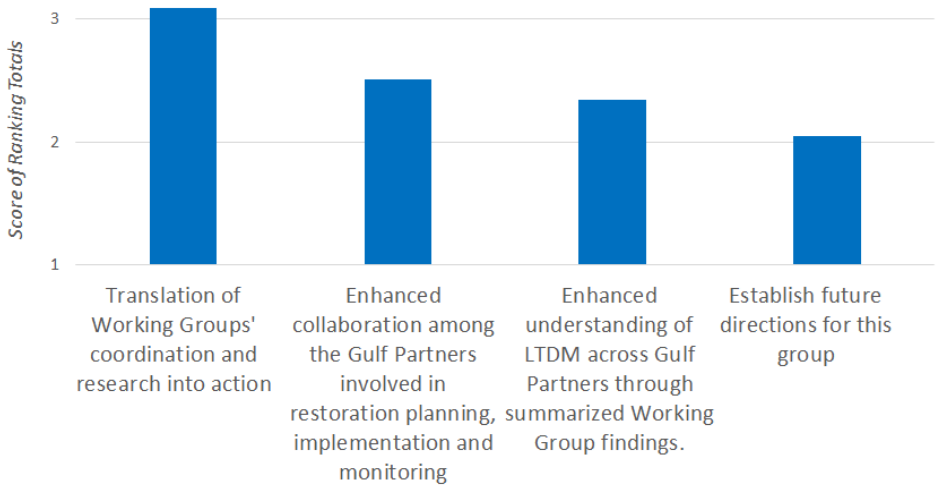
Other Survey Questions

"How has the work that your Working Group has completed over this past year contributed to addressing challenges related to data management in the Gulf?"

"Please share any additional suggestions/ ideas for the future direction and goals of the Long Term Data Management Working Group"

9

What would be the most beneficial outcome of this workshop? (Rank 1-4)



Thank you!!

Up Next:

Reports from Teams/Working Groups

Appendix D

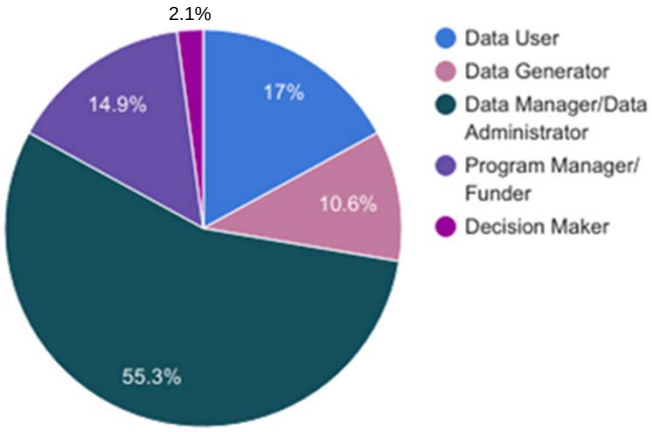
2017 Pre-Workshop Survey Results

Visions for Long Term Data Management in the Gulf

DWH Long Term Data Management Workshop Participant Survey Responses

Survey Responses

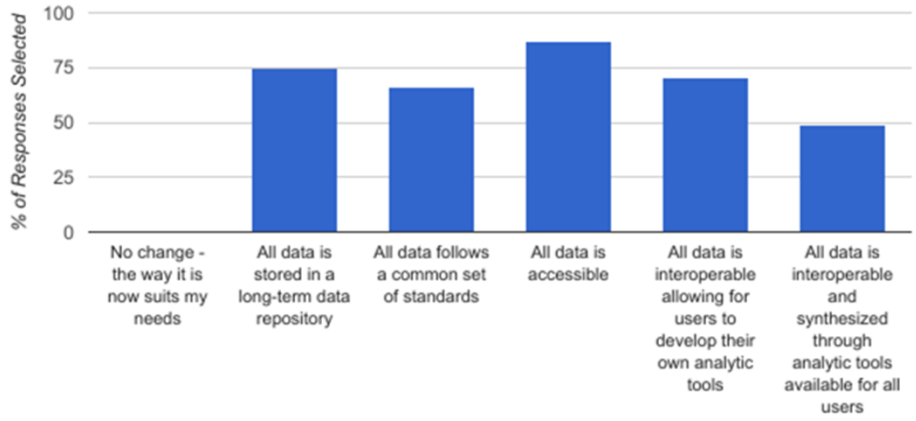
How would you best describe yourself? (Pick one)



2

Survey Responses

What do you want from Gulf research/monitoring data 15 years from now? (Pick all that apply)



3

Survey Responses

Realistically, what do you think Gulf research/monitoring data will have achieved in 15 years? (Pick all that apply)

4

Survey Responses

What do you see as the biggest challenge to data management in the Gulf?

5

Survey Responses

What do you see as the biggest challenge to data management in the Gulf?

"Developing and adhering to a common set of data standards across all data generators."

"..data exchange needs more than common acceptance of need. There needs to be momentum in the form of funding contingent or leadership from organizations."

"The flexibility of a framework for data, so users can upload their data for the repository as well as driving analytics and visualization, where the burden is off, or at least lessened, for the user to meet specific standards, formats, etc."

6

Survey Responses

What do you see as the biggest challenge to data management in the Gulf?

"One group driving the bus!"

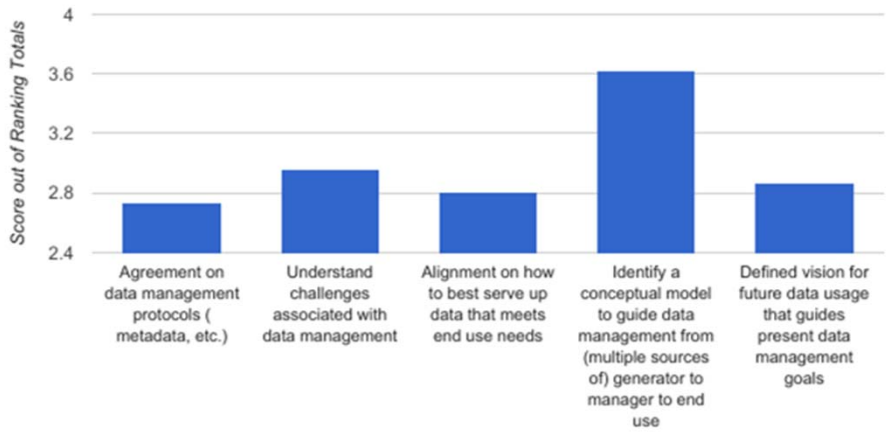
"Identifying how we can tailor data management towards the eventual use of the data on decision making through models, synthesis, etc. Connecting data management and data utilization."

"Move forward with collaboration despite remaining uncertainty. Take a calculated risk that existing data systems can expand to encompass common goals, and will be improved with greater engagement."

7

Survey Responses

What would be the most beneficial outcome of this workshop? (Rank 1-5)



8

Thank you!!

Up Next:

9:45 ***Break***

10:00 PLENARY: Overview of Data Generation [*Michele Jacobi*]

9

Appendix E

Session 1: Breakout Group Results

Name of Group: Data Management Standards

Comment: the definition of exchange is how you achieve integration

Recommendations	Significance of findings (High, Med, Low)	What was missed? What should be added? What should be modified?	Challenges
Step 2: Develop a set of recommended data integration/exchange standards	High	-Add scope – categories of standards -Cross walk/Consistency with Cross TIG-MAM (control vocabularies for parameters) -Clarify what data integration/exchange standards are -Guidance for data generators- data management plan (e.g., standard of metadata). Provide detailed metadata field descriptions to meet system needs.	-Parsing out fields vs. interoperability data standards -Data and metadata curation
Step 1: Adopt/adapt/develop a data exchange format for systems	High	-Determine common data exchange fields among existing systems, -Compare systems (do bake off of stable/vetted exchange format) -Make publically available when complete, -Data owners cross walk fields to “data exchange format” -Independent of any current data system	-What are the common fields? -Limit number of fields (minimally viable product)
Make data integration/exchange standards & protocols publicly available, easily accessible via websites and up-to-date	High	-Make guidance document available to data generators -Field Observation template (NOAA DIVER)- to create potential exchange document	
Products developed by working group should be made available users.		-Who are the users (agencies, public) expansive vs. constricted	-Ongoing steward ships of documents
Annual updates to system inventory documents			-Funding labor hours

Name of Group: Interoperability Group

Recommendations	Significance of findings (High, Med, Low)	What was missed? What should be added? What should be modified?	Challenges
Develop recommendations for Interoperable data systems (best management practices) 1. Develop standardized data methodology (e.g., consistent processing, measurements, recording) 2. Provide standardized data services (use consistent data delivery) 3. Use existing community standards or build a cross-walk; integrate data so they have meaning together 4. Enable tools at all levels; local and national (scaling). 5. Documentation and communication of best management practices.	1. High 2. Low 3. Med 4. Low 5. High	Commonalities Matrix could be weighted (modified) Coordinate with standards group (added) Identify the target audience (missed) Working Group representation (need broader affiliation representation) (missed)	Get buy-in; no stick. Mandate to do; but no funding. Human capital changeover. Right message to the right people. Build in institutional resilience. Distribute responsibility. Get people excited to participate.
Advocate Interoperability to become a grant/program/system requirement 1. Market ability for future decision-making (risk management)	High		

2. How to communicate benefits of interoperability across systems. 3. Sell at high level			
Execute an end-to-end Interoperability Pilot Program 1. Translators; take disparate metrics and make them usable together 2. Show real world application (tell the story) 3. Build across sectors and disciplines, energy, clean water, agriculture, future forecast etc.	1. Low 2. High 3. Med		

Name of Group: Searchability

Recommendations	Significance of findings (High, Med, Low)	Challenges
Evaluate smart tools (machine learning, AI…) to see if they can be leveraged for this group	High	- Implementation dependent on vocabulary, indexes and availability
Identify community driven vocabularies – leverage existing e.g. GOMA	High	- Consensus - Adoption
Directory of systems published online (finding the right repository) (Similar to Gulf TREE)	High	- Funding - Understanding of how it works internally
Data query satisfaction (Did you find what you were looking for?) Y/N?	High	- Development - User response opportunities
Metadata to facilitate exchange at high level – make all related information available (analytical methods etc.) able to be indexed readily by advanced searches	Med	- Dependent upon community driven vocabularies
Exposing metadata in a way that a web search tool (e.g. Google dataset search) can index and discover	Med/Low	- Standardization
Encourage metadata sharing through open source federated systems like CKAN	High	- Standardization
Make both raw and processed data available (reports linked to data)	High	- Recording data provenance
Linking of data types: Integration of data categories by time, location and ontological network	Med/Low	- Many dependencies - Evolution of user needs
Capture DOI metrics to inform data use (how many times something has been cited)	Med	- Requires broader

Appendix F

Session 2: Breakout Group Results

DWH LTDM Workshop: Day 2 AM – December 5

Name of Group: Group A

Breakout Discussion 2: Translating Recommendations into Action, Part I

Interoperability	Standards	Searchability

Recommendation	**Ease of Achievement** (Easy, Med, Hard)	**Cost** (Low, Moderate, High)	**Possible Champions**	**Timeline Yrs.** (short < 1 yr, medium 1-5, long term >5)	**Priority** (High, Med, Low)	**Dependent/contingent on other recommendations?**
1.Develop standardized data methodology (e.g., consistent processing measurements, recording)	Med	High	Funding Agency & system owner	Depends on scope and scale	High	-Contingent upon data workflow for data generator -Steps inform Pilot Project (9)
2.Provide standardized data services (use consistent data delivery)	Easy	Moderate to High b/c based on technology and skill set	Regional/federal entities (e.g., ERMA, GOMA, DIVER)	Variable depending on skill set and will	Med	-Document in 5 -Steps inform Pilot Project (9)
3.Use existing community standards or build a cross-walk; integrate data so they have meaning together	Hard	Moderate b/c of building cross-walk and integrating data	NCEI, IOOS, others	Depends on quality of #1 as well as systems for integration (depends on many small steps)	High	-Contingent on 1 -Demonstrate upon test data set and system(s) -Possible tie in with standard vocabulary -Steps inform Pilot Project (9)

4.Enable tools at all levels; local and national (scaling).	Hard	Variable, dependent upon data (e.g., low cost for sensor data)	Federal partners, funding agencies	Dynamic – needs/tools evolve	Low	-Dependent upon the data system - Steps inform Pilot Project (9)
5.Documentation and communication of best management practices.	Easy	Moderate b/c time sink	Core Team oversees/pilot project responsible for documentation (see number 5)	Medium	High	-Contingent on 1 -3 could inform 5 - Steps inform Pilot Project (9)
6.Market ability for future decision-making (risk management)	Easy if right people are involved (communicating science). Tasks are easy, but hard sell for decision makers to buy	Initial work – Low	SEA Grant, GRIID-C, Core Team, disseminated to individual system, Data and Monitoring priority issue team (GOMA), NCEI, State advocate	Dynamic/on-going effort	High	-State advocate may have more impact than federal gov. -Flush out 1-5 to make 6-8 more valuable -Use 9a as another example -where to store
7.Advocate - How to communicate benefits of interoperability across systems.						
8.Advocate for interoperability at high levels within organization						
9.Execute an end-to-end Interoperability Pilot Program	Dependent upon example/application –pick something easy	Dependent upon pilot selected	Driven by community need (volunteer), Funding entities with big picture views (e.g., ORR, NCEI)	Short (ideally)	High	-Steps 1-5 inform pilot project
9a. Show real world application (tell the story)	Dependent upon example/application –pick something easy	Moderate	SAV pilot (maybe, if they are thinking	Short time (ideally)	High	Falls in with 5

			towards data standards beyond monitoring parameters)			
9b. Translators; take disparate metrics and make them usable together						Note: metrics considered as data conversions (e.g., elevation vs. depth) -quick and dirty for pilot project
9c. Build across sectors and disciplines, energy, clean water, agriculture, future forecast etc.						
Recommendation	**Ease of Achievement** (Easy, Med, Hard)	**Cost** (Low, Moderate, High)	**Possible Champions**	**Timeline Yrs.** (short < 1 yr, medium 1-5, long term >5)	**Priority** (High, Med, Low)	**Dependent/contingent on other recommendations?**
Step 1: Develop a set of recommended data integration/exchange standards	Medium	Dependent upon consensus	NCEI, Core Team	Dependent upon consensus	High	-dependent upon consensus
Step 2: Adopt/adapt/develop a data exchange format for systems	Vary depends on adopt/adapt/develop	Professional ability/skill set – variable depending on adopt/adapt/develop	ORR (work with core team/EPA)	Professional ability/skill set – variable depending on adopt/adapt/develop	High	-non system specific -foundational to other steps -post recommended format through system like OGC
Make data integration/exchange standards & protocols publicly available, easily accessible via websites and up-to-date	Easy	Low	Regional IOOS associations, Possible Council	Dynamic – maintain over time	High after 1 and 2	-contingent upon step 1 and 2

Products developed by working group should be made available to users.	Easy	Low	Core Team/Working Groups	Short	High	-NEK time
Annual updates to system inventory documents	Easy-Medium	Low	Core Team/ Working Group, Individual systems support	Annual/dynamic	Med	-contingent on volunteer time/might change timeline for update -reminders for updated content from individual systems -where to store
Recommendation	**Ease of Achievement** (Easy, Med, Hard)	**Cost** (Low, Moderate, High)	**Possible Champions**	**Timeline Yrs.** (short < 1 yr, medium 1-5, long term >5)	**Priority** (High, Med, Low)	**Dependent/contingent on other recommendations?**
Evaluate smart tools (machine learning, AI…) to see if they can be leveraged for this group	Easy	Low	NETL, NCEI, DIVER (workshop participant input)	Short	Med	-depends upon common vocabulary
Identify community driven vocabularies – leverage existing e.g. GOMA	Medium b/c involves consensus	Low	NCEI	Short (ideally)	High	-identify/inventory vocabulary -come up with first draft
Directory of systems published online (finding the right repository) (Similar to Gulf TREE)	Easy-Medium	Low-Moderate	Core Team/Working group	Short/dynamic b/c of maintenance	Med	-where to post publically
Data query satisfaction (Did you find what you were looking for?) Y/N?	Easy-Medium	Low cost, but interoperability may be mod-high	System owner	Dynamic	Low	-important to get user metrics -success relates to interoperability 1-5
Metadata to facilitate exchange at high level – make all related information available	Hard	Moderate-High (depends on systems)	System owner within working groups, core team oversees,	Medium	Med	-relates to interoperability #3

(analytical methods etc.) able to be indexed readily by advanced searches			input from funding entities			
Exposing metadata in a way that a web search tool (e.g. Google dataset search) can index and discover						-Best management practices
Encourage metadata sharing through open source federated systems like CKAN						
Make both raw and processed data available (reports linked to data)						
Linking of data types: Integration of data categories by time, location and ontological network	Ties back to interoperability #3					
Capture DOI metrics to inform data use (how many times something has been cited)			E.g., NCEI, NETL, others			-investigative, current research going on -shows return on investment -Demonstrate capability at next meeting, could be next steps

DWH LTDM Workshop: Day 2 AM – December 5

Name of Group: Group B

Breakout Discussion 2: Translating Recommendations into Action, Part I

Interoperability	Standards	Searchability

Recommendation	**Ease of Achievement** (Easy, Med, Hard)	**Cost** (Low, Moderate, High)	**Possible Champions**	**Timeline Yrs.** (short < 1 yr, medium 1-5, long term >5)	**Priority** (High, Med, Low)	**Comments & Dependent/contingent on other recommendations?**
Step 1: Adopt/adapt/develop a data exchange format for systems *Group assumes this includes metadata standards*	Hard	High	- RESTORE - NAS - NOAA (*1)	Short/Medium (*2)	High	(*1) Possible champions listed, other groups necessary for success (*2) Intensive workshop for consensus building then review by community
Step 2: Develop a set of recommended data integration/exchange standards	Med	High	- RESTORE - NAS - NOAA (*1)	Short/Medium (*2)	High	Paired with Step 1 workshop
(a) Make data integration/exchange standards & protocols publicly available, easily	(a) Easy (b) Hard	(a) Low (b) Moderate	- One group from above (Steps 1 &2)	(a) Short (b) Long term	(a) Med (b) Med	Championed identified at workshop Dependent on Steps 1 & 2

accessible via websites and (b) up-to-date						
Products developed by working group should be made available users.	Easy	Low	- CRRC (workshop report)*	Short	High	*Static List
Annual updates to system inventory documents	Hard	Low	- NOAA - RESTORE - NAS	Long term	Low/Med	Depends on item above / response to item above (Products developed by working groups – usefulness)
Recommendation	**Ease of Achievement** (Easy, Med, Hard)	**Cost** (Low, Moderate, High)	**Possible Champions**	**Timeline Yrs.** (short < 1 yr, medium 1-5, long term >5)	**Priority** (High, Med, Low)	**Comments & Dependent/contingent on other recommendations?**
Evaluate smart tools (machine learning, AI…) to see if they can be leveraged for this group E.g. Evaluating tools to improve searchability and discoverability	Easy	Low	- GRIIDC - DIVER - NETL/DOE	Short	Med	Other
Identify community driven vocabularies – leverage existing e.g. GOMA	Med/Hard	Low/Moderate	- NOAA (esp. NCEI) - RESTORE - NAS	Short	Med	Discoverability Related to but not dependent on Step 1 & 2 Standards
Directory of systems published online (finding the right repository) (e.g. Gulf TREE)	Med/Hard				Low/Med	Searchability Decision support tool Publicizing spreadsheet of systems (system inventory documents in standards)

Data query satisfaction (Did you find what you were looking for?) Y/N? e.g. Implement user feedback tools to understand what works well and doesn't work well regarding data, metadata and query results.	Easy	Low	- Each system	Short	Low	Searchability Best practice suggestion for individual systems
Recommendation	**Ease of Achievement** (Easy, Med, Hard)	**Cost** (Low, Moderate, High)	**Possible Champions**	**Timeline Yrs.** (short < 1 yr, medium 1-5, long term >5)	**Priority** (High, Med, Low)	**Comments & Dependent/contingent on other recommendations?**
Metadata to facilitate exchange at high level – make all related information available (analytical methods etc.) able to be indexed readily by advanced searches e.g. implement data standards recommendation for metadata	Hard	Moderate	-Each individual system*	Short/Med Depends on system	High	Discoverability *Best practice Dependent on data standards recommendation for metadata (Step 1 &2 in Standards)
Exposing metadata in a way that a web search tool (e.g. Google dataset search) can index and discover	Depends on system NCEI, GRIIDC = Easy DIVER = Hard Others = unknown	System specific	- Each individual system	Short/med Depends on system	Depends on goals of system	Discoverability Best Practice
Encourage metadata sharing through open	Easy*	Low	Each system	Short*	Low	Discoverability *Depends on implementing data standards (two rows above)

source federated systems like CKAN						Best Practice
Recommendation	**Ease of Achievement** (Easy, Med, Hard)	**Cost** (Low, Moderate, High)	**Possible Champions**	**Timeline Yrs.** (short < 1 yr, medium 1-5, long term >5)	**Priority** (High, Med, Low)	**Comments & Dependent/contingent on other recommendations?**
Make both raw and processed data available (reports linked to data)	Needs more information					Other
Linking of data types: Integration of data categories by time, location and ontological network e.g. Amazon search model… You may also be interested in (based on time & location)	Hard / Needs more clarification					Searchability (a) place and time easier (b) relationship development more difficult
Capture DOI metrics to inform data use (how many times something has been cited)						Searchability - Integrate with user satisfaction metric - Best Practice (higher level) - Could be recommendation to funding groups - Potential incentive
Develop standardized data methodology (e.g., consistent processing,						

Provide standardized data services (use consistent data delivery)						
Use existing community standards or build a cross-walk; integrate data so they have meaning together						Related to data exchange
Enable tools at all levels; local and national (scaling).						
Documentation and communication of best management practices.						Related to BMPs recommendation in other fields
Market ability for future decision-making (risk management)						
Advocate - How to communicate benefits of interoperability across systems.						
Advocate for interoperability at high levels within organization						
Translators; take disparate metrics and make them usable together						

Show real world application (tell the story)						
Build across sectors and disciplines, energy, clean water, agriculture, future forecast etc.						

Name of Group: Group C

Breakout Discussion 2: Translating Recommendations into Action, Part I

Interoperability	Standards	Searchability

Recommendation	Ease of Achievement (Easy, Med, Hard)	Cost (Low, Moderate, High)	Possible Champions	Timeline Yrs. (short < 1 yr, medium 1-5, long term >5)	Priority (High, Med, Low)	Dependent/contingent on other recommendations?	Group Notes
Evaluate smart tools (machine learning, AI…) to see if they can be leveraged for this group	Medium	Moderate	Research fellow; funded by Restore Science	Medium	Low	Evaluate the tools is a research project; but to use is dependent on metadata and other priorities	web crawling; research project; structure metadata and data to be searchability
Identify community driven vocabularies – leverage existing e.g. GOMA	Med/Hard (not that the task is hard, but it is not a priority)	Moderate	NOAA NCEI	Medium	High	Order #1	
Directory of systems published online (finding the right repository) (Similar to Gulf TREE)	Medium	Low	RESTORE Council through GOMA Project Tracker	short	medium	technically not hard, but need GOMA to take lead	

Data query satisfaction (Did you find what you were looking for?) Y/N?	Easy	Low	data systems with query tools (GRIID-C, DIVER, etc)	Short	Low	N/A	
Metadata to facilitate exchange at high level – make all related information available (analytical methods etc.) able to be indexed readily by advanced searches	Hard	High	Every organization (funders make it a priority, data generators must submit metadata, etc.)	start now and ongoing	High (begin by incorporating more training etc.)	Order #2.a	
Exposing metadata in a way that a web search tool (e.g. Google dataset search) can index and discover	Easy/Med	Moderate	Every organization (leveraging GCOOS work)	short	high	Order #2.b; contingent on data organized and exposed	
Encourage metadata sharing through open source federated systems like CKAN	Easy/Med	Moderate	Every organization (leveraging GCOOS work)	short	high	Order #2.c; contingent on data organized and exposed	

Make both raw and processed data available (reports linked to data)							NOTE: processed data is different than reports; not related to searchability; this is the end point
Linking of data types: Integration of data categories by time, location and ontological network	Hard	Moderate/High (leveraging other attributes)	Every organization	Medium	medium	related to data exchange format from standards group	ex. Linking different data sets, lat/long; require query tools to capture this
Encourage use and Capture DOI metrics to inform data use (how many times something has been cited)	Easy	Moderate (because you have to purchase DOI)	GRIID-C; funders as a requirement in data management plans	Short (start now); Medium to capture metrics	medium	final step to capture DOI, but need good metadata from individual	some aspects may already be done; but inconsistently done
Develop standardized data methodology (e.g., consistent processing)	Hard	Moderate	National Data Systems, NOAA NCEI	Long term	High/medium	This is a foundational recommendation	If data in systems are consistent then translators can be used
Provide standardized data services (use consistent data delivery)							

Use existing community standards or build a cross-walk; integrate data so they have meaning together	Easy due to technology access; Hard due to human will	Low/Moderate	National Data Systems, NOAA NCEI, Scientific community, Funders	Medium	High	N/A	
Enable tools at all levels; local and national (scaling).							
Develop recommendations for Interoperable data systems (best management practices): Documentation and communication of best management practices.	Medium	Low	All data systems; NOAA DIVER, NOAA NCEI	Short/Medium	medium		future sub-group
Market ability for future decision-making (risk management)							

Advocate - How to communicate benefits of interoperability across systems.	Easy	Low	Leadership	Medium	medium		
Advocate for interoperability at high levels within organization	Easy	Low	Leadership	Medium	medium	doesn't mean it will be accomplished	
Translators; take disparate metrics and make them usable together							
Show real world application (tell the story)							
Build across sectors and disciplines, energy, clean water, agriculture, future forecast etc.							

Step 1: Adopt/adapt/develop a data exchange format for systems							
Step 2: Develop a set of recommended data integration/exchange standards							
Make data integration/exchange standards & protocols publicly available, easily accessible via websites and up-to-date							
Products developed by working group should be made available users.							
Annual updates to system inventory documents							

DWH LTDM Workshop: Day 2 AM – December 5

Name of Group: Group D

Breakout Discussion 2: Translating Recommendations into Action, Part I

Interoperability	Standards	Searchability

Recommendation	**Ease of Achievement** (Easy, Med, Hard)	**Cost** (Low, Moderate, High)	**Possible Champions**	**Timeline Yrs.** (short < 1 yr, medium 1-5, long term >5)	**Priority** (High, Med, Low)	**Dependent/contingent on other recommendations?**
Develop standardized data methodology (e.g., consistent processing/handling)	Medium	Moderate (Savings in long run)	LTDM Working Group Members, End users (e.g. TIGs)	Medium	High	Consistency within program methodology (depends on program definition)
Provide standardized data services (use consistent data delivery)	Medium	Moderate	OGC (Spatial only), NAS, NOAA RESTORE, LTDM	Medium→Long	Medium	Consistency within project/program Costs dependent on size/diversity of program
Use existing community standards or build a cross-walk; integrate data so they have meaning together	Medium/Hard	High	GRIID-C, NOAA RESTORE, NAS, NCEI, NGOs (TNC), LTDM, GOMA, COPs	Long	High (Would take a long time and effort, so need to begin ASAP)	Step 1: Adopt/adapt/develop a data exchange format for systems Step 2: Develop a set of recommended data integration/exchange standards Step 3: Make data integration/exchange standards & protocols publicly available, easily accessible via websites and up-to-date Facilitated by Outreach and presentation at Conferences/Training

DWH LTDM Workshop: Day 2 AM – December 5

Recommendation	**Ease of Achievement** (Easy, Med, Hard)	**Cost** (Low, Moderate, High)	**Possible Champions**	**Timeline Yrs.** (short < 1 yr, medium 1-5, long term >5)	**Priority** (High, Med, Low)	**Dependent/contingent on other recommendations?**
Enable tools at all levels; local and national (scaling).						
Documentation and communication of best management practices.	Easy	Low	SeaGrant, GRIID-C, Feds, LTDM	Short	High (Do before GoMOSES in February – Ben in charge)	Using language such as: Preferred, Acceptable, Not Acceptable Helps inform previous Recommendations
Marketability for future decision-making (risk management)	Medium (Varies)	Low	SeaGrant, NGOs, GOMA, NCEI	Long	Medium (But need to start now)	Dependent on knowing the right narrative for the appropriate decision-maker 1) Advocate - How to communicate benefits of interoperability across systems. 2) Advocate for interoperability at high levels within organization 3) Show real world application (tell the story)
Translators; take disparate metrics and make them usable together	Easy	Low	LTDM	Short	High (helps inform documentation)	Note: This is analytic data metrics (can we revise the recommendation to reflect?) 1) Helps to identify what is important to data users and developers
Build across sectors and disciplines, energy, clean water, agriculture, future forecast etc.	**HARD**	High $$$$	NCEI, CEQ	Long	Low (for this group)	Dependent on marketability recommendation

DWH LTDM Workshop: Day 2 AM – December 5

Recommendation	**Ease of Achievement** (Easy, Med, Hard)	**Cost** (Low, Moderate, High)	**Possible Champions**	**Timeline Yrs.** (short < 1 yr, medium 1-5, long term >5)	**Priority** (High, Med, Low)	**Dependent/contingent on other recommendations?**
Step 1: Adopt/adapt/develop a data exchange format for systems	Medium	Moderate	LTDM, NCEI	Short	High	Test cases (e.g. WQX→DIVER, CIMS→DIVER, CIMS→WQX)
Step 2: Develop a set of recommended data integration/exchange standards	Easy (but dependent on documentation and data exchange test cases)	Moderate	LTDM, NAS, NCEI, FEDs (TIGs, RESTORE)	Medium	High	
Make data integration/exchange standards & protocols publicly available, easily accessible via websites and up-to-date	Easy	Low	LTDM, NAS, NCEI, FEDs (TIGs, RESTORE)	Short (once developed)	High	Contingent on Step 1 and 2 above
Products developed by working group should be made available users.	Medium	Low	LTDM (USGS, RESTORE)	Medium	High	Final approval and process (Kathryn?) Important sites for outreach and distribution: LTDM websites Needs to be hosted in a single location (possibly restorethegulf.gov)
Annual updates to system inventory documents	Medium	Low	LTDM	Long (on-going)	Medium	Dependent on primary POC staying up to date
Evaluate smart tools (machine learning, AI…) to see if they can be leveraged for this group	Medium	Moderate	NETL, NCEI	Medium	High	Contingent on agreed upon common vocabulary (see row below)

Recommendation	**Ease of Achievement** (Easy, Med, Hard)	**Cost** (Low, Moderate, High)	**Possible Champions**	**Timeline Yrs.** (short < 1 yr, medium 1-5, long term >5)	**Priority** (High, Med, Low)	**Dependent/contingent on other recommendations?**
Identify community driven vocabularies – leverage existing e.g. GOMA	Hard	Moderate	LTDM, NCEI, NETL	Medium	High(est)	
Directory of systems published online (finding the right repository) (Similar to Gulf TREE)						Preliminarily addressed by existing Data Systems Spreadsheet?
Data query satisfaction (Did you find what you were looking for?) Y/N?	Med	Moderate	System providers	Medium	Med (but very cool)	Dependent on end-user need Privacy constraints
Exposing metadata in a way that a web search tool (e.g. Google dataset search) can index and discover	Med	Moderate	System providers	Medium	Med	Suggest Lessons Learned Metadata Team to address the steps: 1) Metadata to facilitate exchange at high level – make all related information available (analytical methods etc.) able to be indexed readily by advanced searches 2) Encourage metadata sharing through open source federated systems like CKAN
Make both raw and processed data available (reports linked to data)					Low	
Linking of data types: Integration of data categories by time,					Low	

location and ontological network						
Capture DOI metrics to inform data use (how many times something has been cited)					Low	

Appendix G

Session 3: Breakout Group Results

DWH LTDM Workshop: Day 2 AM – December 5

Name of Group: Group B

Breakout Discussion 3: Translating Recommendations into Action, Part II

Select top 3-5 (Maximum of 10) recommendations from list above

Recommendation (not listed in order of priority)	**Collaboration/integration needed**	**Stakeholder buy-in**	**Future work**	**Objectives**
1. Compile best practices for data systems including: - developing, populating and exposing metadata - use and evaluation of DOIs - applying controlled vocabularies	- Focused working groups - Dedicated and committed members representative of stakeholder needs - Empowered to make decisions - Diverse, dedicated, doers and decision-makers	- Decision makers based on advocacy (4) - Broader community of providers/system owners - Data users awareness	- Developing communication and outreach strategies	- Pre-workshop survey and targeted outreach to develop workplan - Intensive workshop bookended by teleconferences - Deliverable: o Best practices "document"
2. Data exchange format and standards				- Pre-workshop survey and workshop as above - Deliverable: data exchange format and standards
3. Identify community driven vocabularies				- Build upon previously identified vocabularies (EDDM) - Deliverable: List of vocabularies and how to apply them

4. Advocate for interoperability to funders at high levels	- Everyone needs to advocate	- Fundamental	- Further outreach with SeaGrant	- Deliverable: talking points (elevator speech), 2-pager and report

DWH LTDM Workshop: Day 2 AM – December 5

Name of Group: Group C

Breakout Discussion 3: Translating Recommendations into Action, Part II

Select top 3-5 (Maximum of 10) recommendations from list above

Recommendation	Collaboration/integration needed	Stakeholder buy-in	Future work	Objectives
1. Data Standards Step 1: Adopt/adapt/develop a data exchange format for systems Step 2: Develop a set of recommended data integration/exchange standards	- Data standards inventory –identify suite of core fields for all data types across systems - Representatives from all data systems draft recommended standards - Leverage existing work	- Owners of data systems - Data generators	- System owners verify compiled field data spreadsheet - Working groups continue	- Advance the ability to integrate data between systems - Enhance a suite of products for data blending
2. Metadata Identify community driven vocabularies – leverage existing e.g. GOMA	- System owners; line offices; federal and private sectors - Develop common themes/message for education/outreach	- All data generators	- Create list of core metadata information - Identify gaps and standards - SeaGrant and/or o NAS Fellow to build outreach communications	- Develop a restoration vocabulary - Implement a restoration vocabulary

Metadata to facilitate exchange at high level – make all related information available (analytical methods etc.) able to be indexed readily by advanced searches Exposing metadata in a way that a web search tool (e.g. Google dataset search) can index and discover Encourage metadata sharing through open source federated systems like CKAN				
3. Advocacy Advocate - How to communicate benefits of interoperability across systems.	- Data system owners to develop common themes/message	- Leadership buy-in - Incorporate into agency core mission	- Develop communication packet	- Create awareness and leadership support for community goals (e.g. standards)

Advocate for interoperability at high levels within organization				

DWH LTDM Workshop: Day 2 AM – December 5

Name of Group: Group D

Breakout Discussion 3: Translating Recommendations into Action, Part II

Select top 3-5 (Maximum of 10) recommendations from list above

Recommendation	**Collaboration/integration needed**	**Stakeholder buy-in**	**Future work (after today)**	**Objectives/Goals**
1. Documentation and communication of best management practices.	Compile existing best management practices in LTDM Data Standards Working Group Folder from each of the participating programs	Outstanding stakeholders need to buy-in	Crosswalk of BMPs and creation of guidance document	Develop clear language for BMPs such as: Preferred, Acceptable, Not Acceptable (e.g., pdfs) Helps inform previous Recommendations Includes: Metadata, known file types, files sizes, model run guidance
2. Develop a set of recommended data integration/exchange standards a) Develop a data exchange format for systems b) Develop/adopt integration/exchange standards from consistent fields	Systems developers to develop test cases	System owners	1) Test cases (e.g. WQX→DIVER, CIMS→DIVER, CIMS→WQX, DIVER→NCEI GRIIDC→NCEI 2) Identify core fields and advocate for data systems to use them	1) Allow for integration of data across studies/programs 2) Allow for broader use of data sets for studies/questions
3. Identify community driven vocabularies – leverage existing e.g. GOMA, NCEI.	Compile existing systems vocabularies in Working Group folder (including International vocabularies where appropriate)	Current stakeholders with current data vocabularies	Building a master list – to the appropriate level (required vs. optional)	1) Achieve identified searchability recommendation 2) Community adoption of recommended vocabularies 3) Will inform metadata and environmental data management.

4. Products developed by working group should be made available users	Bandwidth and hosting availability	LTDM	Final product review and workshop report	Share products with the Gulf data community, and have them used

COORDINATION OF DWH LONG-TERM DATA MANAGEMENT: THE PATH FORWARD

DECEMBER 4 - 5, 2018

Appendix H

Workshop Presentations

WELCOME

NOAA's GOM Disaster Response Center

DWH LONG-TERM DATA MANAGEMENT

WORKSHOP LOGISTICS

- Emergency Exits
- Restrooms
- Cell phones / laptops
- Breaks (coffee, tea, snacks)
- Meals
  - $15/day for special lunch delivery
  - Dinners on your own
  - See restaurant map in packet
- Logistical questions – see Kathy Mandsager or me

DWH LONG-TERM DATA MANAGEMENT

Coastal Response Research Center (CRRC)

- **Partnership between NOAA's Office of Response and Restoration and the University of New Hampshire**
  - **Emergency Response Division (ERD)**
  - **Assessment and Restoration Division (ARD)**
  - **Marine Debris**
- **Since 2004**
- **Co-Directors:**
  - **UNH – Nancy Kinner**
  - **NOAA – Ben Shorr**

DWH LONG-TERM DATA MANAGEMENT 3

Overall CRRC Mission

- **Conduct and oversee basic and applied research and outreach on spill & environmental hazard response and restoration**
- **Transform research results into practice**
- **Serve as hub for spill /environmental hazards R&D**
- **Facilitate workshops bringing together ALL STAKEHOLDERS to discuss spill/hazards issues and concerns**

DWH LONG-TERM DATA MANAGEMENT 4

FACILITATION PLEDGE

- I will recognize and encourage everyone to speak
- I will discourage side conversations
- I commit to:
  - Being engaged in meeting
  - Keeping us on task and time
- Tell me if I am not doing this!

DWH LONG-TERM DATA MANAGEMENT 5

PARTICIPANT PLEDGE

- Be Engaged
  - Turn off cell phones & laptops(except at breaks)
- Listen to Others
- Contribute
- Speak Clearly; Use Microphones
- Learn from Others
- Avoid Side Conversations

DWH LONG-TERM DATA MANAGEMENT 6

Core Team

- Benjamin Shorr, NOAA ORR ARD & Co-Director of CRRC
- Marti Goss, NOAA Restoration Center
- Mike Peccini, NOAA Restoration Center
- Lauren Showalter, NAS Gulf Research Program, NAS
- Sharon Mesick, NOAA National Centers for Environmental Information

Working Groups

- Data Management Standards
  - Jessica Henkel, RESTORE Council (Kathryn Keating)
  - Nicolas Eckhardt, NOAA ORR ARD Spatial Data Branch
- Interoperability
  - Jay Coady, NOAA ORR ARD Spatial Data Branch
  - Kyle Wilcox, Axiom Consulting, AOOS Team
- Discovery/Searchability
  - William Nichols, GRIIDC
  - Mike Peccini, NOAA Restoration Center

AGENDA – DAY 1

0830	Welcome/Logistics – *Nancy Kinner, Coastal Response Research Center*
0845	Background of DWH LTDM Effort - *Benjamin Shorr, NOAA ORR*
0900	Workshop Objectives and Format
0915	Self-Introductions of Workshop Participants

DWH LONG-TERM DATA MANAGEMENT

9

Background of DWH LTDM Effort

Ben Shorr

NOAA ORR Spatial Data Branch

CRRC Co-Director

DWH LONG-TERM DATA MANAGEMENT

10

Workshop Goals

1. Continue to foster collaboration among the Gulf of Mexico partners involved in restoration planning, implementation and monitoring (e.g., RESTORE Council, GOMRI, GRIID-C, NAS Gulf Research Program, GOMA, NFWF, and Gulf Environmental Benefit Fund).
    a. What does collaboration "look like"? What does success look like?
2. Report out on progress towards goals of 3 Working Groups (see below).
    a. What progress has been made and summarize findings;
    b. What has been challenging;
    c. Were these good working groups and goals? (Did your group goals shift and if so, why?)
    d. Recommendations for best practices / suggestions to address challenges

DWH LONG-TERM DATA MANAGEMENT

11

Workshop Goals

3. How to translate Working Groups coordination and research into action.
    a. What concrete steps should be taken to foster collaboration and data integration? (Identify short, medium, and long-term actions, what is their priority? Ease of achievement, cost, and organizational mandates/priorities?)
        i. Data Systems perspective: how do we use what we have learned to improve the practice of systems working together (this will then make it easier for data providers to understand best management practices)
        ii. Funders perspective: how to simplify/improve workflow for data providers to improve use of standards and best management practices
    b. What is the future direction of this Working Group or other forum?

DWH LONG-TERM DATA MANAGEMENT

12

AGENDA – DAY 1

0830	Welcome/Logistics – *Nancy Kinner, Coastal Response Research Center*
0845	Background of DWH LTDM Effort - *Benjamin Shorr, NOAA ORR*
0900	Workshop Objectives and Format
0915	Self-Introductions of Workshop Participants

DWH LONG-TERM DATA MANAGEMENT

13

PARTICIPANT INTRODUCTIONS

- **Name**
- **Affiliation**
- **Work related to DWH Long Term Data Management**
- **Attendance at June 2017 DWH LTDM Workshop?**
- **Member DWH LTDM Working Group or Core Team?**
  - **Data Management Standards**
  - **Interoperability**
  - **Searchability/Data Discovery**

DWH LONG-TERM DATA MANAGEMENT

14

AGENDA – DAY 1

0930 Overview of Survey Results - *Jessica Henkel, Restore Council*

0945 Reports from Teams/Working Groups

- DWH LTDM Data Management Standards Working Group

1015 *Break*

1030 Reports from Teams/Working Groups

- DWH LTDM Interoperability Working Group
- DWH LTDM Discovery/Searchability Working Group

DWH LONG-TERM DATA MANAGEMENT

15

AGENDA – DAY 1

0930 Overview of Survey Results - *Jessica Henkel, Restore Council*

0945 Reports from Teams/Working Groups

- DWH LTDM Data Management Standards Working Group

1015 *Break*

1030 Reports from Teams/Working Groups

- DWH LTDM Interoperability Working Group
- DWH LTDM Discovery/Searchability Working Group

DWH LONG-TERM DATA MANAGEMENT

16

AGENDA – DAY 1

1130 Perspectives from Stakeholders (Panel)

- RESTORE Council - *Jessica Henkel*
- STATES - *Libby Fetherston, Robert Gruba & Steve Jones*
- NAS GRP (Gulf Research Program) - *Lauren Showalter*
- Emergency Response - *Charlie Henry, NOAA Disaster Response Center*
- NRDA (Natural Resource Damage Assessment) - *Eric Weissberger & Michele Jacobi, NOAA*

1230 *Lunch*

DWH LONG-TERM DATA MANAGEMENT

17

AGENDA – Day 1 continued

1315 Perspectives from Data Systems (Panel) - Response to working group activities

- NCEI – *Sharon Mesick*
- DIVER – *Benjamin Shorr*
- GRIID-C - *Rosalie Rossi*
- CIMS - *Craig Conzelmann*
- USEPA WQX - *Laura Shumway*
- SECOORA & GCOOS - *Shin Kobara*

1430 *Break*

1445 Breakout Discussion Group 1: Input on Working Group Findings/Recommendations

- Significance of findings
- What was missed, additions, modifications?
- Challenges

Breakout Discussion Group 1 will consist of 3 breakout groups:

- Data Management Standards
- Interoperability
- Discovery/Searchability

1600 Plenary Report outs

1700 Adjourn

DWH LONG-TERM DATA MANAGEMENT

18

AGENDA – Day 2

Day 2

0830 Recap and Recalibration – *Nancy Kinner, Coastal Response Research Center*

0900 Breakout Discussion Group 2: Translating Recommendations into Action

A. Working Group Outcomes and Recommendations: Next Steps: Short, Medium and Long Term
- Priorities
- Ease of achievement, cost, organizational mandates needed for:
  - Data Management Standards
  - Interoperability
  - Discovery/Searchability

Breakout Group 2 & 3 will consist of 4 breakout groups each with a mix of:

- Data Systems Representatives
- Funders
- Oil Spill, Emergency & NRDA Responders
- Restoration Representatives
- Representatives of each of the 3 Working Groups

1045 *Break*

1100 Plenary Report Outs

1200 *Lunch*

1245 Breakout Discussion Group 3: Translating Recommendations into Action

B. Feasibility and Implementation of Next Steps and Recommendations
- Collaboration/integration needed
- Stakeholder buy-in
- Future work/objectives

1415 *Break*

1430 Plenary Report Outs

1530 Path Forward - *Lauren Showalter, LTDM Core Team*

1615 Closing Remarks - *Mike Peccini and Benjamin Shorr, LTDM Core Team*

1630 Adjourn

DWH LONG-TERM DATA MANAGEMENT

19

Session I Breakout Groups:

DATA MANAGEMENT STANDARDS	INTER-OPERABILITY	SEARCHABILITY	Choose A Group
Courtney Arthur	**Jay Coady**	**Mike Peccini**	Harris Bien
Julie Bosch	**Kyle Wilcox**	Ann Jones	Mark Defley
Nicolas Eckhardt	Zhankun Wang	**Ben Shorr**	George Graettinger
Jessica Henkel	Rosalie Rossi	Caitlin Young	Robert Gruba
Lei Hu	Craig Conzelmann	Jennifer Bauer	Steve Jones
Kathryn Keating	Steve Delgreco	Kathy Martinolich	Jonathan Blythe
Steve Ramsey	Jason Weick	Shin Kobara	Kirsten Larsen
Michele Jacobi		Kevin Suir	Matt Love
			Ali Robertson
			Lauren Showalter
			Laura Shumway
			Dan Van-Nostrand
			Eric Weissberger
			Ursula Arnold
			Richard Raynie
			Tom Strange - day 1 only
			Sharon Mesick
			Kenneth Halanych
			Christina Hunnicutt
			Deb Hernandez
			Shelby Servais

DWH LONG-TERM DATA MANAGEMENT

20

Session I Breakout Groups - Questions

DWH LTDM workshop: Day 1 – December 4

Name of Group: _____

Breakout Discussion 1

Recommendations	Significance of findings (High, Med, Low)	What was missed? What should be added? What should be modified?	Challenges

DWH LONG-TERM DATA MANAGEMENT 21

Session 2 & 3 Breakout Groups:

Group A	Group B	Group C	Group D
Courtney Arthur	Ursula Arnold	Nicolas Eckhardt	Harris Bien
Jennifer Bauer	Julie Bosch	**Steve Delgreco**	Jessica Henkel
Jonathan Blythe	George Graettinger	Christina Hunnicutt	**Michele Jacobi**
Jay Coady	Kenneth Halanych	Shin Kobara	Steve Jones
Craig Conzelmann	**Ann Jones**	Ben Shorr	Matt Love
Mark Defley	Kathryn Keating	Laura Shumway	Richard Raynie
Lei Hu	Kathy Martinolich	Steve Ramsey	Rosalie Rossi
Sharon Mesick	Lauren Showalter	Zhankun Wang	Kevin Suir
Mike Peccini	Kyle Wilcox	Jason Weick	Dan Van-Nostrand
Ali Robertson	Robert Gruba	Eric Weissberger	Caitlin Young
		Kirsten Larsen	

DWH LONG-TERM DATA MANAGEMENT 22

Session 2 Breakout Groups - Questions

DWH LTDM Workshop: Day 2 – December 5

Name of Group: _____

Breakout Discussion 2: Translating Recommendations into Action, Part I

Recommendation	**Priorities** (High, Med, Low)	**Ease of Achievement** (Easy, Med, Hard)	**Cost** (Low, Moderate, High)	**Organizational mandate**	**Timeline Yrs.** (short < 1 yr, medium 1-5, long term >5)
1.					
2.					
3.					
4.					

DWH LONG-TERM DATA MANAGEMENT

Session 3 Breakout Groups - Questions

Breakout Discussion 3: Translating Recommendations into Action, Part II

Select top 3 recommendations from list above

Recommendation	Collaboration/integration needed	Stakeholder buy-in	Future work	Objectives
1.				
2.				
3.				

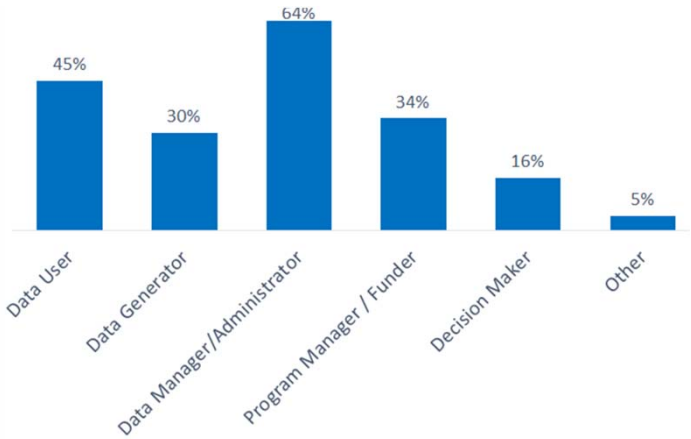
DWH LONG-TERM DATA MANAGEMENT

Looking Forward:
Long Term Data Management in the Gulf

DWH Long Term Data Management Workshop Participant Survey Responses

Survey Responses

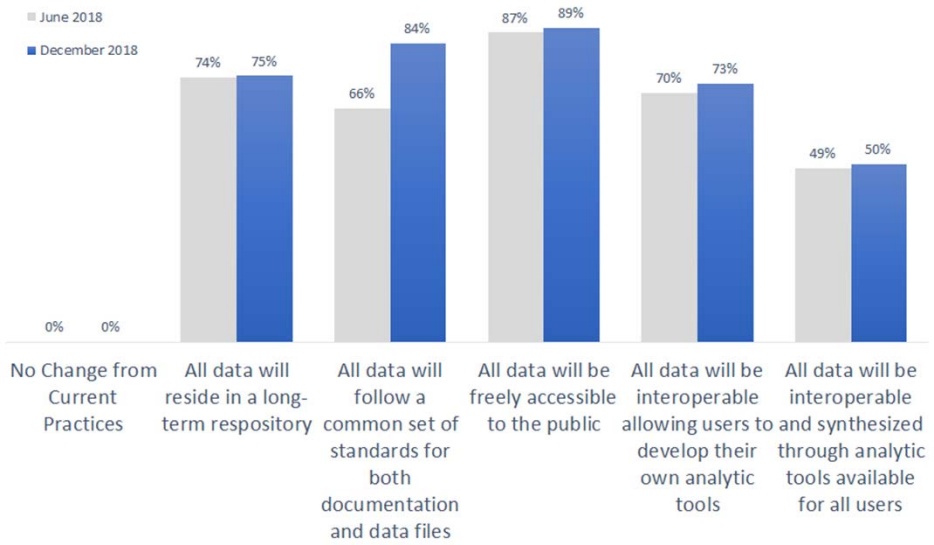
How would you best describe Yourself?
(Select all that apply)



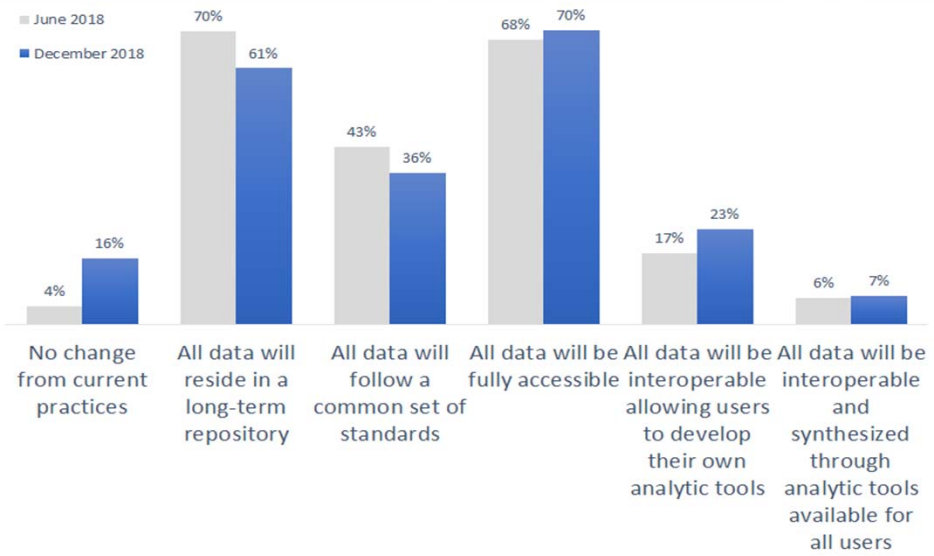
45% of survey respondents chose two or more roles to describe their work.

2

What do you want from Gulf research/monitoring data 15 years from now? (Select all that apply)



Realistically, what do you think Gulf research/monitoring data will have achieved in 15 years? (Select all that apply)



What do you see as the biggest challenge to data management in the Gulf?

Key Themes

Variation in Data Systems & Standards	Collaboration & Communication	Funding	Variation in Goals & Missions	Leadership	Attitudes about Data Management

5

What do you see as the biggest challenge to data management in the Gulf?

Variation in Data Systems & Standards	"There are so many different entities collecting, storing, and using data. Therefore, there is a lot of redundancies, inefficiencies, and barriers to synthesis and comprehensive analysis of restoration progress." "Agreeing to a set of common standards and then addressing how to transition from historic methods to recommended."
Funding	"Variety of governance and legacy (pre-existing) programs with resistance to change. Lack of incentive &/or funding to change." "Sustainability of data management efforts with projects and funding winding down" "Lack of funding. Lack of leadership from funding agencies".

6

What do you see as the biggest challenge to data management in the Gulf?

Collaboration & Communication	"Coordination across different offices and sectors to work on data management to establish standards, interoperability, data discovery etc.."
	"Coordination and acceptance of developed standards, data warehousing for access and funding support for longterm data management.
Variation in Entity Goals / Missions	"Widely varying scale, resources, and goals among Gulf-focused projects"
	"....The fact that the majority of data repositories/providers have different missions and there is not money to encourage collaboration and allow those working on them the time to do the work needed for collaboration"

7

What do you see as the biggest challenge to data management in the Gulf?

Leadership	"No single entity responsible for uniform data management practices"
	"There is no clear decision maker. We need to decide where we want to go and work to get there, rather than seeing where things already overlap and hope that people might make use of that information."
Attitudes toward Long Term Data Management	"I think data managers are convinced that data interoperability won't or can't work. We are limiting ourselves with this perspective."
	"Data management is not a high priority task for data generators, and other demands on their time win over, even if they philosophically agree with data sharing."

8

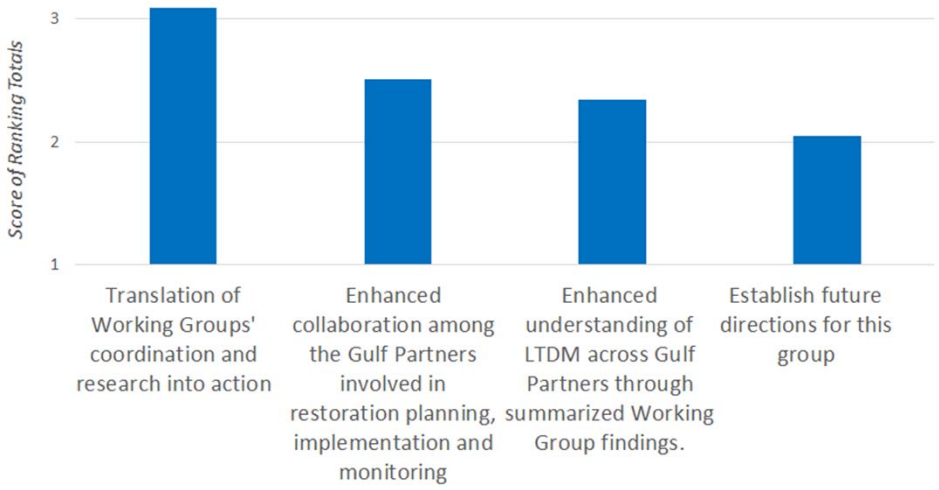
Other Survey Questions

"How has the work that your Working Group has completed over this past year contributed to addressing challenges related to data management in the Gulf?"

"Please share any additional suggestions/ ideas for the future direction and goals of the Long Term Data Management Working Group"

9

What would be the most beneficial outcome of this workshop? (Rank 1-4)



Thank you!!

Up Next:

Reports from Teams/Working Groups

11

Data Standards Group Update

2018 DWH Long Term Data Management Workshop

Charge

- Identify categories of standards needed (i.e. data acquisition including sampling protocols and quality control, data management).
- Determine what gaps need to be filled for data management standards. This gap analysis will inform the list of standards that need to be established.

“ ‘Data Standard’ is a very abstract and general term. The... information collected by the Data-Standard group would allow other groups to see what “data standards” actually entails, and how it is related to searchability and interoperability of scientific data.”

Process

- 6 Month Timeline
- Team Leads
  - Jessica Henkel (RESTORE)
  - Nick Eckhardt (NOAA)
- Working Group Support
  - Kathryn Keating (RESTORE)
  - Courtney Arthur (IEC)
- Monthly Meetings with Working Group
  - Coordination across entities to gather information on data management systems and standards

Products: Inventory of Existing Systems

- Data Management Systems Inventory
  - Collaborative overview of 30 different systems used for data management in the Gulf.
  - Reduced to 16 “actionable” systems for detailed evaluation of standards

“...just knowing all the data systems that are out there is a big help.”

“Having a list of all the data systems and their standards is helpful for those trying to figure out the best place to store their data.”

Products: Detailed System Templates

- Data Systems Templates for majority of systems
  16 Data Systems Templates completed
  - Example:
    - Data Integration Visualization Exploration and Reporting - DIVER
    - LA Coastal Information Management System - CIMS
    - Dauphin Island Sea Lab - MyMobileBay
  - Templates will allow for development of system "crosswalks" to identify consistent data fields, that increase interoperability of data across systems

"This detailed information at the attribute level is crucial for any effort in the future to make two or more data systems interconnect."

Products: Compiled Fields Document

- Compiled Data System Information Spreadsheet
  - This document contains a summary of systems reviewed. Contact and system overview information is included.
  - All the data system fields are compiled on one tab to allow for analysis

"Developing data standards across multiple agencies with different missions is almost impossible. Instead, to the extent practicable, it would be good to at least identify "common" data fields/attributes, and try to get agreement on standards or formats for these fields."

"This could lead to improvements in both data collection and tool development.. [and] provide benefit to those groups/agencies/entities that are just starting up their own monitoring and assessment programs allowing them to learn from existing programs."

Working Group Recommendations for Sharing Data System Information With Public

- Sharing information about *Data Systems*:
  - Easy-to-understand documentation and explanations
    - Summaries
    - Charts & graphs
    - Links to detailed spreadsheets
  - Offered via scalable method of exchange
    - Appropriate for various levels of expertise
- Sharing information about *Working Group processes*:
  - Summary document
    - "Read Me" element in spreadsheets

Recommendations

- Teams involved with data management are encouraged to make their system's data standards & protocols publicly available, easily accessible via websites and up-to-date.
  - 10/16 systems info available
  - Ex. NCEI World Ocean Database
- Products developed by working group should made be available users.
- Next Steps

Possible Next Steps?

- Develop a set of recommended data standards
  - *Challenge:* Variation in goals across agencies makes this a complicated task. How best to address this challenge?
- Develop a data exchange format for systems
  - *Challenge:* This is a resource-intensive process. Who has the capacity to develop this?

Thank you!!

10

Next Steps? - Reference

Go a step higher by plotting the data to see who is collecting what and on what temporal space

A set of recommended data standards

A summary document explaining out process and the outcome.

I'd like to see the creation of documentation about how the data standards working group came up with the data systems review spreadsheet, and why. Such documentation would be a helpful guide, informing researchers, scientists, and data managers, about the necessary steps to take to have an accurate and precise understanding of a data system. Without this understanding, Long-Term Data Management, Data Synthesis, and other data related tasks will be rendered less achievable.

Identify "common" data fields/attributes, and try to get agreement on standards or formats for these fields. It would also be helpful to have a comprehensive gulf-wide data dictionary that provides responsible agency, contact person, etc., for situational awareness and for information exchange. For example, if someone from Florida would like to start collecting data that Louisiana has been collecting for years, a comprehensive data dictionary with contacts would allow for better consistency and the ability to share lessons learned.

It would also allow for transparency and would help those involved with research (e.g., Universities) to know what type of data and information are available to help leverage resources and support research. This could lead to improvements in both data collection and tool development that might help the collecting agency. It would provide benefit to those groups/agencies/entities that are just starting up their own monitoring and assessment programs allowing them to learn from existing programs.

Interoperability Group Update

2018 DWH Long Term Data Management Workshop

Group Charge

1. Determine what could optimize interoperability efficiency between DWH long-term data management systems, and drive collaboration among them.
1. Compile strategic goals and key features for data warehouses and repositories.
1. Determine the intended, current and future use of DWH long-term data management systems.

2

Process

- Determine applications/systems to consider/review
  - Data Management Systems Inventory
- Compile attributes for those systems and identify POCs
- Narrow attributes of systems to those that pertain to data exchange
- Compile commonalities of systems and create a matrix.

Spreadsheet Analysis Slides

Metadata Types	System Counts
ISO 19115-2	22
FGDC	11
CSDGM FGDC	6
ISO 19115-1	2
ISO 19110	1
ISO 19139	1
DDI	1
Dublin Core	1
N/A	10

System Function	System Counts
Mapping	33
Search	24
Export/Data Download	11
Query	2

Data Types	System Counts
Spatial	32
Tabular	19
Metadata	5

Data Delivery	System Counts
Direct Download/Query	22
Map Service	17
API	9
FTP	6
ERDDAP	5
THREDDS	4
Open DAP	3
HTTPS	3
Live Access Server	3
TDS (Tabular Data Stream)	2
OGC SOS	1
WAF (Web Accessible Folder)	1
Other	1
N/A	6

Geographic Range	System Counts
National	15
State	11
Global	9
Regional (Gulf)	5
Local	2

4

Spreadsheet Data Delivery Example

System	API	ERDDAP	Map Service	WAF (Web Accessible Folder)	TDS (Tabular Data	OGC SOS	Open DAP	Direct Download/ Query/FTP/HTTPS	THREDDS	Live Access Server
TOTALS	**9**	**5**	**17**	**1**	**2**	**1**	**3**	**30**	**4**	**3**
SECOORA	X	X	X		X		X	X	X	
DIVER	X	X	X					X		
ERMA	X		X					X		
Texas Natural Resources Information System	X		X					X		
National Water Quality Portal USGS & EPA	X		X					X		
GRIIDC	X	X						X		
State of Florida Geodata	X							X		
NCEI Severe Weather Data Inventory	X							X		
GCOOS Data Portal	X	X		X	X	X				
FL GIS Open Data Portal			X					X		
LA Coastal Information Management System (CIMS)			X					X		
Digital Coast			X					X		
National Estuary Research Reserve System			X					X		
Texas Coastal Sediments Database			X					X		
TCEQ SQWMIS Database			X					X		
NCEI - Marine Geology Data			X					X		
NCEI - Deep-sea Corals Data		X	X							
Coastal Marine Planning Tool			X							
DWH Project Tracker			X							
FL Water-Cat			X							
Mississippi Coastal Improvements Program (MsCIP)			X							
Dauphin Island Sea Lab Dataset Archive								X		
EPA Scribe								X		
Mississippi Beach Monitoring Program								X		
Mymobilebay								X		
Open ICPSR								X		
NCEI - Climate Resilience Toolkit								X		
NCEI - Bathymetry and Global Relief								X		

Spreadsheet Metadata Example

System	ISO 19115-2	ISO 19115-1	ISO 19110	ISO 19139	FGDC	CSDGM FGDC	DDI	Dublin Core	N/A
Totals	**22**	**2**	**1**	**1**	**11**	**6**	**1**	**1**	**10**
Dauphin Island Sea Lab Dataset Archive	X				X				
InPORT	X				X	X		X	
Mymobilebay	X				X				
National Estuary Research Reserve System	X				X				
NCEI - Gulf of Mexico Data Atlas	X				X	X			
NCEI - Bathymetry and Global Relief	X				X	X			
NCEI - Marine Geophysics Data	X				X	X			
DIVER	X								
State of Florida Geodata	X								
GRIIDC	X								
Digital Coast	X								
SECOORA	X		X						
NCEI - Severe Weather Data Inventory	X								
NCEI - Historical Hurricane Tracks: (IBTrACS)	X								
NCEI - Storm Events Database	X								
NCEI - Billion-Dollar Weather and Climate Disasters	X								
NCEI - Geoportal - archive query tool	X								
NCEI - World Ocean Database	X								
NCEI - Gulf of Mexico Regional Climatology	X								
NCEI - Water Column Sonar Data	X								
NCEI - Marine Geology Data	X								
NCEI - Deep-sea Corals Data	X								
Mississippi Coastal Improvements Program (MsCIP)					X				
Texas Natural Resources Information System					X				
TCEQ SQWMIS Database					X				
National Water Quality Portal USGS & EPA					X				
DWH Project Tracker									X
EPA Scribe									X
FL GIS Open Data Portal									X
FL Water-Cat									X

Many Commonalities

Metadata

				Commonalities
ISO 19115-2	- DIVER - ERMA - State of Florida Geodata - GRIIDC - InPORT - Mymobilebay	- Digital Coast - National Estuary Research Reserve System - SECOORA - NCEI Severe Weather Data Inventory - NCEI Historical Hurricane Tracks: (IBTrACS) - NCEI Storm Events Database	- NCEI Billion-Dollar Weather and Climate Disasters - NCEI Geoportal - archive query tool - NCEI World Ocean Database - NCEI Gulf of Mexico Regional Climatology - NCEI Gulf of Mexico Data Atlas - NCEI Bathymetry and Global Relief	- NCEI Water Column Sonar Data - NCEI Marine Geophysics Data - NCEI Marine Geology Data - NCEI Deep-sea Corals Data
ISO 19115-1	- NCEI Climate Resilience Toolkit - NCEI Archive Management System S2N and ATRAC			
ISO 19110	- SECOORA			
ISO 19139	- NCEI Archive Management System S2N and ATRAC			
FGDC	- Dauphin Island Sea Lab Dataset Archive - InPORT -Mississippi Coastal Improvements Program (MsCIP) - Mymobilebay - National Estuary Research Reserve System	- Texas Natural Resources Information System - TCEQ SQWMIS Database - National Water Quality Portal USGS & EPA - NCEI Gulf of Mexico Data Atlas - NCEI Bathymetry and Global Relief - NCEI Marine Geophysics Data		
DDI	- Open ICPSR			
CSDGM FGDC	- Alabama Coastal Marine Planning Tool - Coastal Marine Planning Tool - InPORT - NCEI Gulf of Mexico Data Atlas - NCEI Bathymetry and Global Relief - NCEI - Marine Geophysics Data			
Dublin Core	InPORT			
N/A	DWH Project Tracker EPA Scribe FL GIS Open Data Portal FL Water-Cat GCOOS Data Portal LA Coastal Information Management System (CIMS)	Mississippi Automated Resource Information System (MARIS) Mississippi Beach Monitoring Program National Library of Medicine Texas Coastal Sediments Database		

Many Commonalities

DIVER

Mapping Function
Alabama Coastal Marine Planning Tool
Coastal Marine Planning Tool
DWH Project Tracker
ERMA
FL GIS Open Data Portal
State of Florida Geodata
FL Water-Cat
GCOOS Data Portal
GRIIDC
LA Coastal Information Management System (CIMS)
Mississippi Automated Resource Information System (MARIS)
Mississippi Beach Monitoring Program
Mississippi Coastal Improvements Program (MsCIP)
Mymobilebay
Digital Coast
National Estuary Research Reserve System
SECOORA
Texas Coastal Sediments Database
Texas Natural Resources Information System
TCEQ SQWMIS Database
National Water Quality Portal USGS & EPA
NCEI - Climate Resilience Toolkit
NCEI - Billion-Dollar Weather and Climate Disasters
NCEI - Geoportal - archive query tool
NCEI - World Ocean Database
NCEI - Gulf of Mexico Regional Climatology
NCEI - Gulf of Mexico Data Atlas
NCEI - Bathymetry and Global Relief
NCEI - Water Column Sonar Data
NCEI - Marine Geophysics Data
NCEI - Marine Geology Data
NCEI - Deep-sea Corals Data

Query Function
EPA Scribe

Data Delivery Map Service
Coastal Marine Planning Tool
DWH Project Tracker
ERMA
FL GIS Open Data Portal
FL Water-Cat
LA Coastal Information Management System (CIMS)
Mississippi Coastal Improvements Program (MsCIP)
Digital Coast
National Estuary Research Reserve System
SECOORA
Texas Coastal Sediments Database
Texas Natural Resources Information System
TCEQ SQWMIS Database
National Water Quality Portal USGS & EPA
NCEI - Marine Geology Data
NCEI - Deep-sea Corals Data

Search Function
Dauphin Island Sea Lab Dataset Archive
FL GIS Open Data Portal
State of Florida Geodata
FL Water-Cat
GCOOS Data Portal
GRIIDC
InPORT
LA Coastal Information Management System (CIMS)
Mymobilebay
Digital Coast
National Estuary Research Reserve System
SECOORA
Texas Natural Resources Information System
TCEQ SQWMIS Database
NCEI Severe Weather Data Inventory
NCEI - Historical Hurricane Tracks: (IBTrACS)
NCEI - Storm Events Database
NCEI - Billion-Dollar Weather and Climate Disasters
NCEI - Archive Management System S2N and ATRAC
NCEI - Geoportal - archive query tool
NCEI - World Ocean Database
NCEI - Gulf of Mexico Regional Climatology
NCEI - Deep-sea Corals Data

Data Delivery Export/Download
DWH Project Tracker
EPA Scribe
GRIIDC
LA Coastal Information Management System (CIMS)
Mississippi Beach Monitoring Program
Mymobilebay
Digital Coast
National Estuary Research Reserve System
SECOORA
NCEI - Deep-sea Corals Data

Data Delivery API
Dauphin Island Sea Lab Dataset Archive
ERMA
State of Florida Geodata
GCOOS Data Portal
GRIIDC
SECOORA
Texas Natural Resources Information System
National Water Quality Portal USGS & EPA
NCEI Severe Weather Data Inventory

Data Delivery ERDDAPI
GCOOS Data Portal
GRIIDC
SECOORA
NCEI - Deep-sea Corals Data

Data Delivery Direct Download
Dauphin Island Sea Lab Dataset Archive
EPA Scribe
ERMA
FL GIS Open Data Portal
State of Florida Geodata
GRIIDC
LA Coastal Information Management System (CIMS)
Mississippi Beach Monitoring Program
Mymobilebay
Digital Coast
National Estuary Research Reserve System
Open ICPSR
SECOORA
Texas Coastal Sediments Database
Texas Natural Resources Information System
TCEQ SQWMIS Database
National Water Quality Portal USGS & EPA
NCEI - Climate Resilience Toolkit
NCEI Severe Weather Data Inventory
NCEI - Bathymetry and Global Relief
NCEI - Marine Geology Data

Tabular Data Types
DWH Project Tracker
EPA Scribe
GCOOS Data Portal
GRIIDC
LA Coastal Information Management System (CIMS)
Mississippi Beach Monitoring Program
Mymobilebay
Digital Coast
National Estuary Research Reserve System
SECOORA
TCEQ SQWMIS Database
National Water Quality Portal USGS & EPA
NCEI - Climate Resilience Toolkit
NCEI Severe Weather Data Inventory
NCEI - Historical Hurricane Tracks: (IBTrACS)
NCEI - Storm Events Database
NCEI - World Ocean Database
NCEI - Gulf of Mexico Regional Climatology

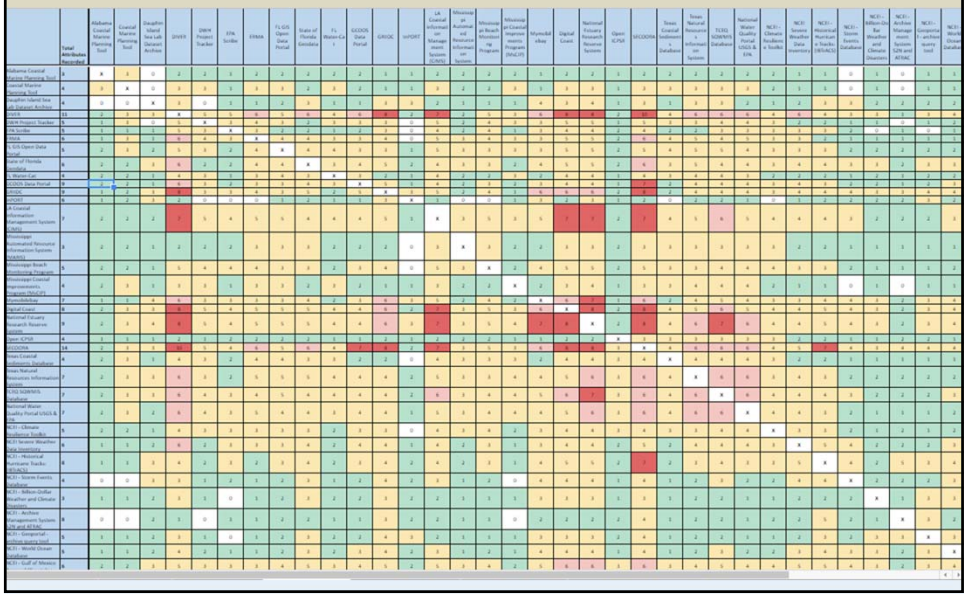
Metadata ISO 19115-2
State of Florida Geodata
GRIIDC
InPORT
Mymobilebay
Digital Coast
National Estuary Research Reserve System
SECOORA
NCEI Severe Weather Data Inventory
NCEI - Historical Hurricane Tracks: (IBTrACS)
NCEI - Storm Events Database
NCEI - Billion-Dollar Weather and Climate Disasters
NCEI - Geoportal - archive query tool
NCEI - World Ocean Database
NCEI - Gulf of Mexico Regional Climatology
NCEI - Gulf of Mexico Data Atlas
NCEI - Bathymetry and Global Relief
NCEI - Water Column Sonar Data
NCEI - Marine Geophysics Data
NCEI - Marine Geology Data
NCEI - Deep-sea Corals Data

Spatial Data Types
Alabama Coastal Marine Planning Tool
Coastal Marine Planning Tool
DWH Project Tracker
EPA Scribe
ERMA
FL GIS Open Data Portal
State of Florida Geodata
FL Water-Cat
GCOOS Data Portal
LA Coastal Information Management System (CIMS)
Mississippi Automated Resource Information System (MARIS)
Mississippi Beach Monitoring Program
Mississippi Coastal Improvements Program (MsCIP)
Digital Coast
National Estuary Research Reserve System
Open ICPSR
SECOORA
Texas Coastal Sediments Database
Texas Natural Resources Information System
TCEQ SQWMIS Database
National Water Quality Portal USGS & EPA
NCEI - Climate Resilience Toolkit
NCEI Severe Weather Data Inventory
NCEI - Historical Hurricane Tracks: (IBTrACS)
NCEI - Gulf of Mexico Regional Climatology
NCEI - Gulf of Mexico Data Atlas
NCEI - Bathymetry and Global Relief
NCEI - Water Column Sonar Data
NCEI - Marine Geophysics Data
NCEI - Marine Geology Data
NCEI - Deep-sea Corals Data

Spreadsheets

- Pared down list
  - Interoperability Commonalities Spreadsheet
    - Geographic Scope
    - Data Types
    - System Functions
    - Metadata Types
    - Data Delivery
    - Commonality Matrix
      - Used data types, system functions, metadata types, and data delivery. NOT Geographic scope.

Commonality Matrix



Current State

- Levels of Interoperability already exist or are in the works
  - IOOS -> ERMA
  - GRIIDC -> DIVER
- Common **Data Services** and **Data/Metadata Standards** make this possible

11

Challenges

- Interoperability is typically **not a mandate**
- Automated M2M communication between data systems is a **real challenge**.
- Enhancing **legacy systems** with Interoperability can be a considerable effort

12

Next Steps

- **Develop** recommendations for Interoperable data systems
- **Advocate** Interoperability to become a grant/program/system requirement
- **Execute** an end-to-end Interoperability Pilot Program

13

Feedback / Questions

- What topics/challenges did we miss?
- Are Data Systems represented correctly in the matrix?
- Do you consider the Data Systems you work with Interoperable?
- Are you surprised by any of the findings of this group?
- What Next Steps would you like to see out of the Interoperability group?
- How should this group communicate their findings?

14

Group Members

Lead: Jay Coady
Lead: Kyle Wilcox

Allison Fischer
Brie Bernik
Craig Conzelmann
Dan Hudgens
Dwane Young
Jason Weick
Jessica Morgan
Kelly Rose
Michele Jacobi
Mike Peccini
Stephen Del Greco
Zhankun Wang
Vanessa Lazar
Rosalie Rossi

Plus many more and of course the **CORE TEAM**

15

Discoverability/Searchability Search Log Analysis

William Nichols

DWH LTDM Workshop

Mobile, AL

Dec 4-5, 2018

Task

- Task 1: Evaluate and prioritize DWH data community needs for search and discovery
  - Subtasks:
    - Identify major data user communities working on DWH-related activities in the Gulf of Mexico
    - Describe data user needs to the extent possible by:
      - Data and/or resource types of interest
      - How are they using the data
      - How do their efforts relate to overall DWH assessment, restoration, or research needs
      - What are their search and discovery needs? What's working?
      - What could be improved? (e.g. Better tools? Specific search functions? Better integration between systems etc)
      - What searches are currently being conducted with existing tools?
        - Pull search logs from data systems and analyze

Search Log Analysis Questions

- What are users actually searching for?
- How are users searching?
- Are users finding data?

Obtaining System Search Logs

- Many systems don't have free text search
  - Categorized Data / Static Links
  - Guided Queries
  - Custom Queries
- Logs are not automatically generated for every system
  - GRIIDC – Elasticsearch and custom code
  - GOMAportal(esri Geoportal) – depends on log settings
- Logs are not easily compared across systems
- Obtained
  - GRIIDC – Elastic Search Logs
  - SECOORA – Pageviews
  - DIVER – Keyword Search and Custom Queries

Example of Search logs (GRIIDC)

- Session ID – allows grouping of user actions without personally identifying
- Timestamp
- Search Terms
- Geofilter Used? (and WKT of geometry if Y)
- Number of Results
- 1st and 2nd Search Scores
- 1st and 2nd GRIIDC ID, Title, Data Landing Page Link

GRIIDC

- Custom Logs generated from Elasticsearch
- What are users actually searching for?
  - Things they already know about
    - DOI
    - UDI
  - Authors
- How are users searching?
  - Free text
  - Geofilter use very low – 4% of searches
- Are users finding data?
  - 0 results – 13%
  - 0 – 4 results – 50%
  - 35 results – 75%

SECOORA

- Search Result Pageviews - Total and Unique
  - https://portal.secoora.org/#search?q=**buoy**&page=1
- What are users actually searching for?
  - Data Products
  - Search is partially auto complete which could skew results
- How are users searching?
  - Free Text with auto complete
  - Spatial, Temporal, Access Method – but not part of logs
- Are users finding data?
  - Yes, but unable to determine success rate from logs
  - Auto complete search gives higher successful search rate as user is guided to queries where results exist

DIVER

- Region, Query Summary, # of results
- What are users actually searching for?
  - Very diverse terms
- How are users searching?
  - Guided Custom Query – did not analyze
  - Keyword Search – free text with follow up filters
- Are users finding data?
  - 0 results – 42%
  - 0 – 12 results – 75%

Technology Advancements

- Discoverability Advancements
  - Google Dataset Search
    - Requires Dataset Landing Pages
    - Requires work on repository end
  - DOE NETL SmartSearch
    - Self building index of data available on web
    - Leverages Bing with automated data mining
    - Requires no work on repository end other than making data accessible and generally findable via search engines
- Searchability Advancements
  - Search tool maturity
    - Lucene -> Solr -> Elasticsearch
  - Autocomplete of search terms

Future Work

- Continue Search Log Analysis
  - Obtain more logs from more repositories
  - Develop search term bank
    - Common keywords
    - Proper Names
  - More cross-repository analysis
- Pageview Log Analysis
  - Capture interest in datasets found from not using repository search engines
- Webserver Log Analysis
  - Determine where users are finding datasets and coming from

Recommendations

- Recommendations
  - Initial Recommendations
    - Make data available via landing pages
      - Silos without landing pages may be largely excluded from future 3rd party dataset search methods
      - Allows for multiple levels of analysis (pageviews, referrals)
    - Invest in improving search logging mechanisms
      - Spend valuable time on areas which need improvement or are heavily used
  - Requires Further Analysis
    - Ancillary search methods may not be needed due to low use
      - Geofilter, Time Period, etc

NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION

NCEI Overview

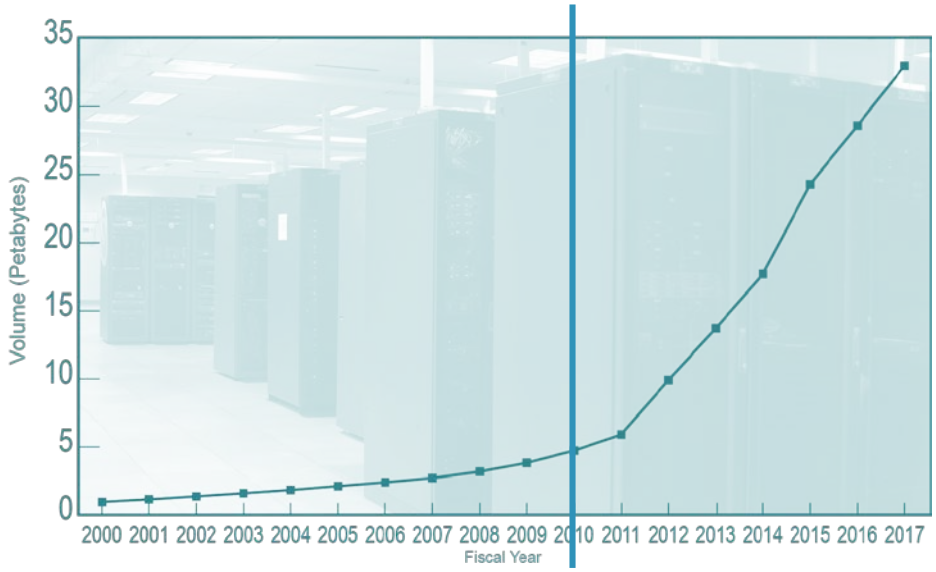
Coordination of DWH Long-Term Data Management: The Path Forward

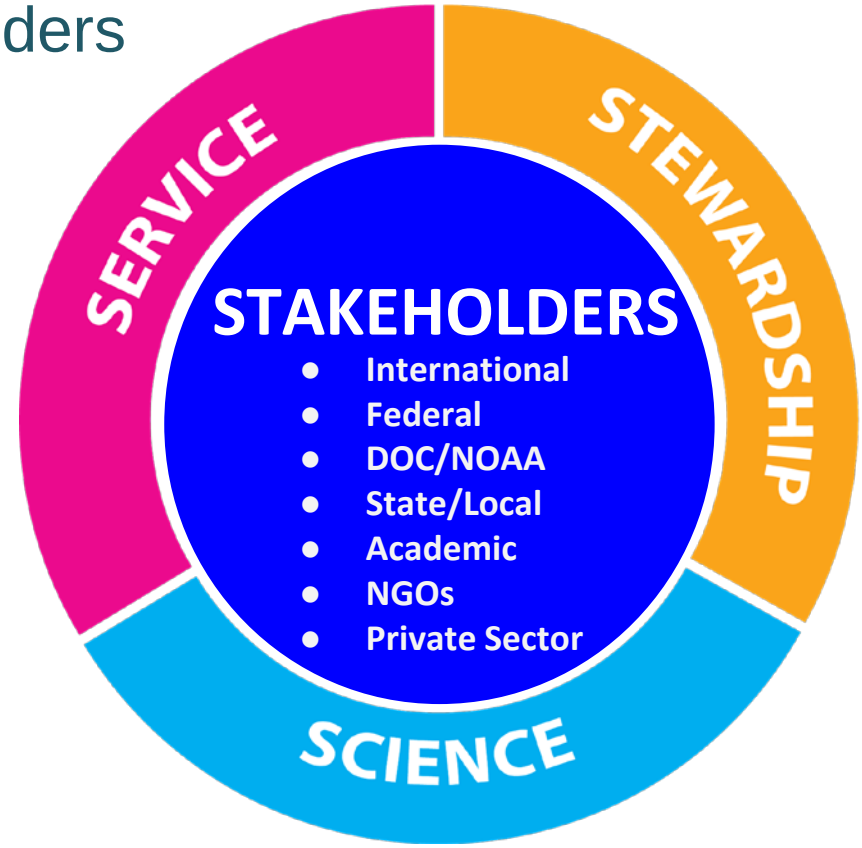
Sharon Mesick, Information Services Chief
Coastal Data Development Program Manager

December 4, 2018

National Centers for Environmental Information

- A **Service** resource for Stakeholders
- **Stewarding** environmental data
- Adding **Scientific Value** to data





NCEI Alignment with DWH-LTDM

. Mandate

. Mission

. Methods

Congressional Mandate

- FY2017 Appropriations Language:

  "The Committee supports NCEI's long term efforts in coastal data development to better understand historical trends, anomalies, and the frequency of event occurrences. CDD is tasked to build the long-term coastal data record to support environmental prediction, scientific analyses, and formulation of public policy."

- FY2018 and 2019 Appropriations Language:

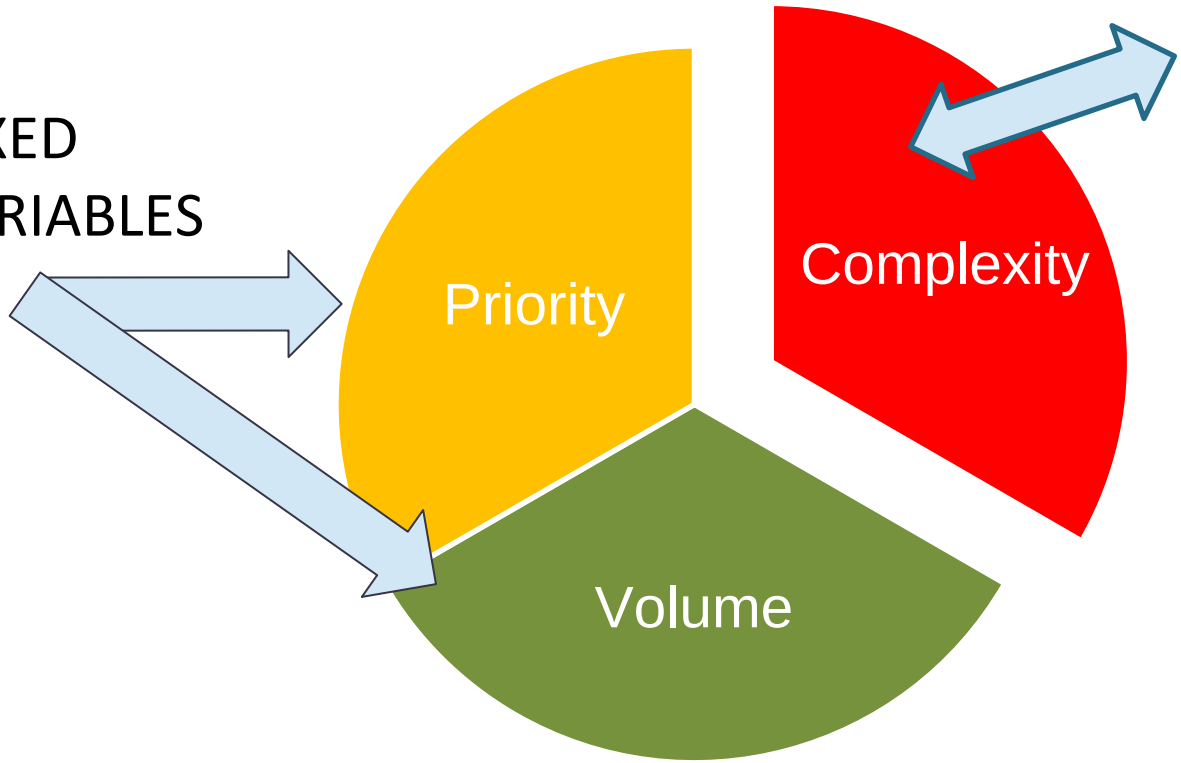
  "NOAA shall consider the Coastal Data Development program as the central repository to manage data collections and information services of the various Gulf of Mexico Restoration activities funded in response to the 2010 Deepwater Horizon oil spill for scientific stewardship."

Mission: Data in the Context of History

1768 **2018->**

Methods: Reducing Data Complexity

FIXED VARIABLES



IMPACTFUL

- Alignment with DWH LTDM objectives / working groups

NCEI Business Model

DWH LTDM – a $hared Responsibility

- ✓ Legally regulated funding model(s)
- ✓ Federal funding source(s) = mandatory compliance with Federal Open Data Policy*
- ✓ No uniform funding model for DWH data management
- ✓ Reduced Complexity increases efficiency

* Some exceptions noted for non-federally funded data collections

Reducing Complexity Increases Efficiency

Helping data providers reduce data complexity

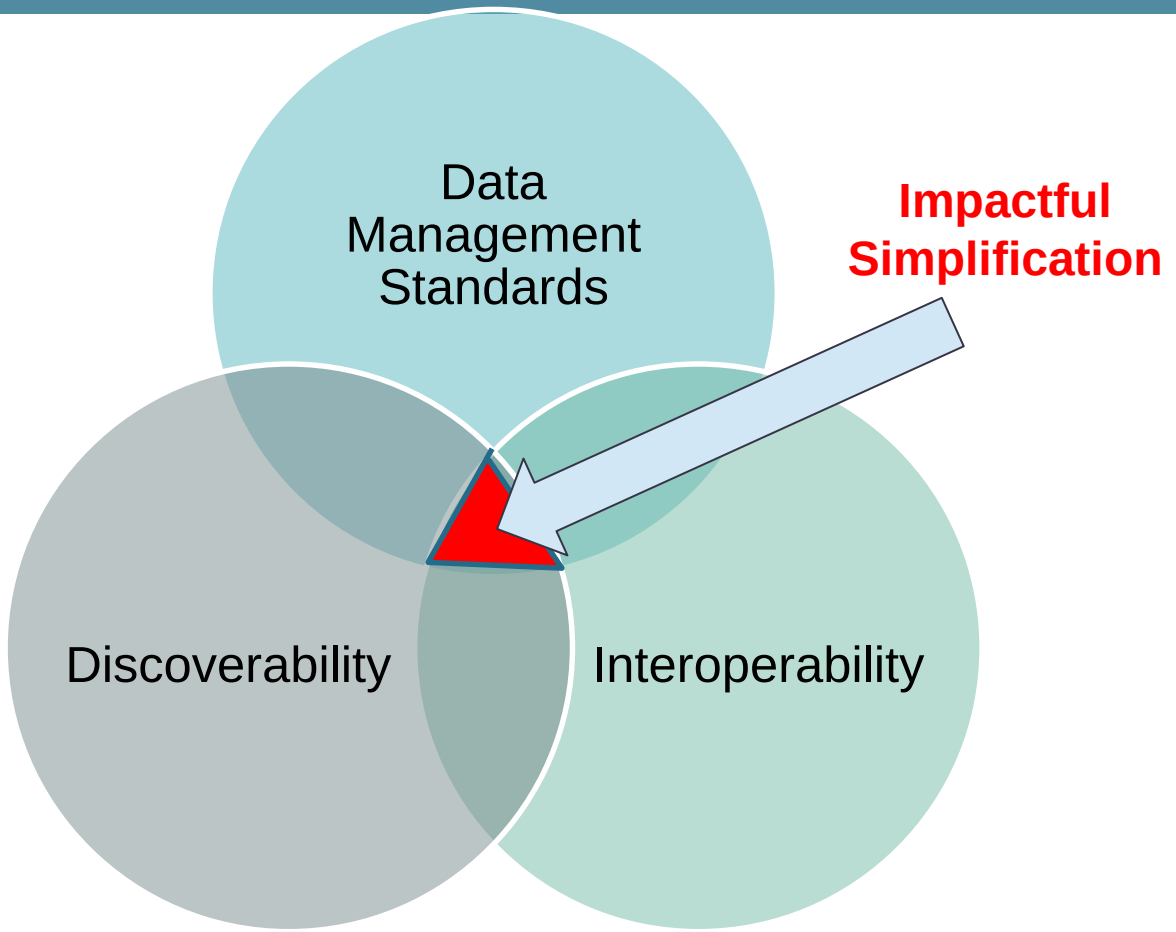
- Reduces data management costs
- Facilitates preservation
- Enables data integration

Methods:

- Informing: Standards and best practices
- Training: Documentation for discovery and access
- Integrating: Broad interoperability and value added
- Partnering: Broad partnership model

In Plain Language *Please…*

- Leverage your process to build information over time
- Do things the same way, one 'like' collection to the next



Proposal – Award – Interactive data collection & documentation – Publication - Archive

Data Stewardship Partnership Examples

IOOS Glider Data Assembly Center

- NCEI receives QA/QC data
- Standards-based preservation and integration

OMAO Data Assembly Center

- Ensuring data integrity, quality, & documentation
- Developing tools to aid data management

Ocean Exploration – End-to-end

- Innovative video data management
- Direct access to integrated environmental data

Data Stewardship Partnership Examples

IOOS Glider Data As[illegible]embly Center

- NCEI receives [illegible]A/Q[illegible]ata
- Sta[illegible]dards-bas[illegible]re[illegible] and integration

OMAO Da[illegible]

- Ensur[illegible] [illegible]ocumentation
- Developing [illegible]ement

Ocean Explora[illegible]on [illegible]End[illegible]o-end

- Innovative video data [illegible]anagement
- Direct access to integrated environmental data

DIVER

World Ocean Database: Benefit to the Nation

World Ocean Database

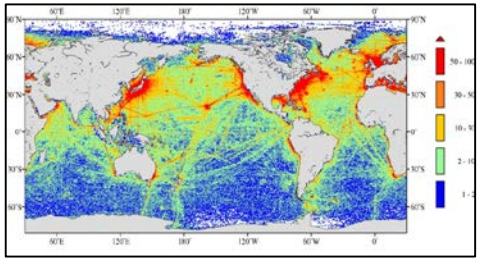
ENHANCE

ASSESS

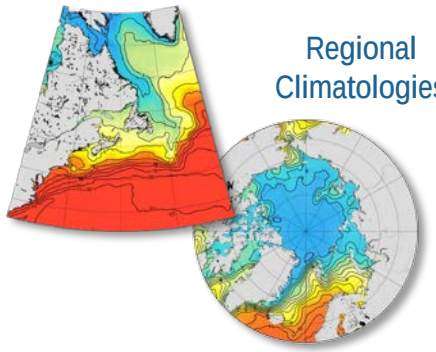
INTEGRATE

Ocean Observing Community

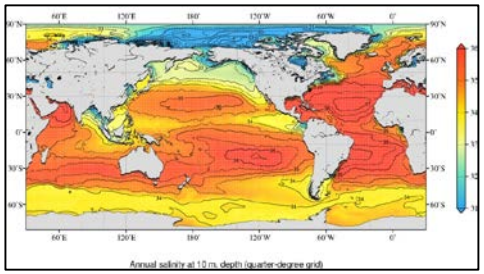
PRODUCTS / SERVICES



World Ocean Atlas



Regional Climatologies

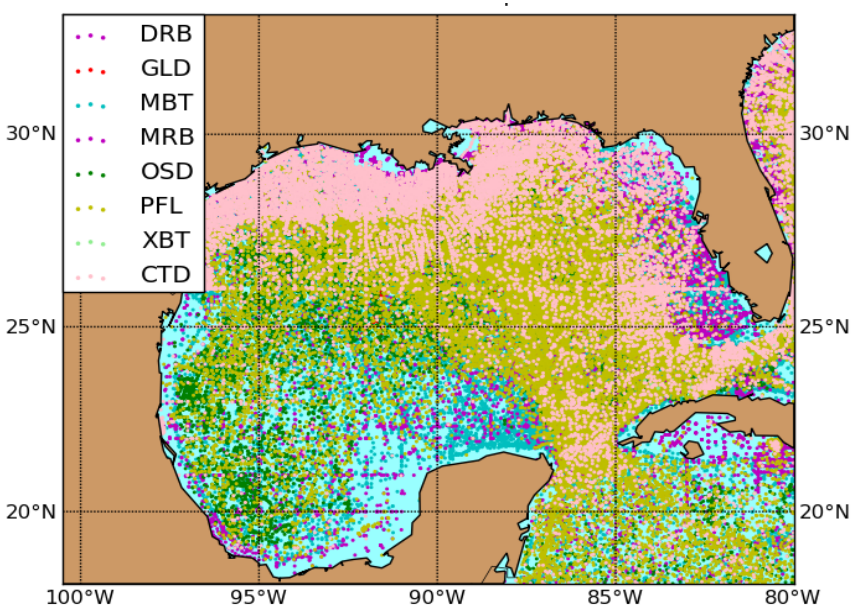


Integrated Ocean Climate Information

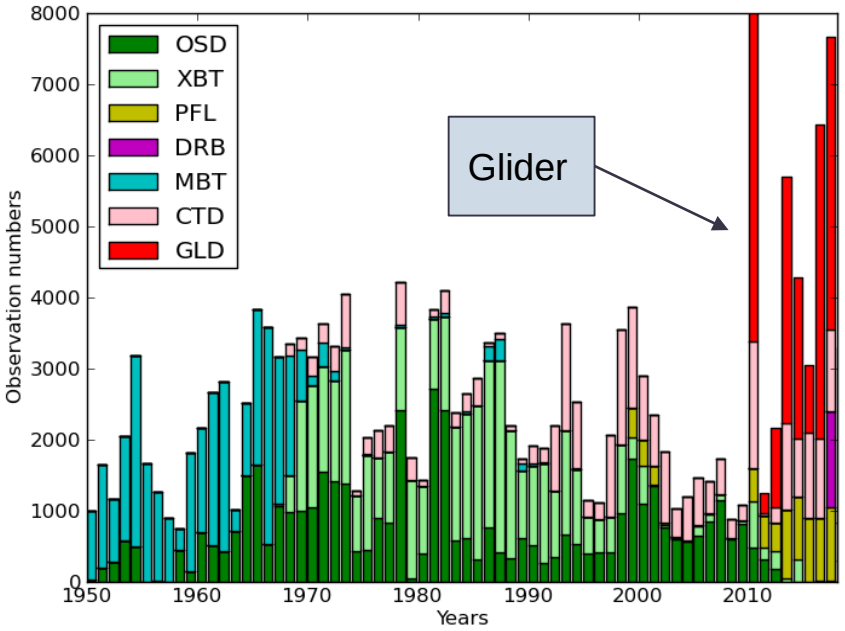
BENEFIT TO THE NATION

- Range of spatial and temporal environments
- ~ 16m profiles
- Assess change over time

Gulf of Mexico World Ocean Database Profiles



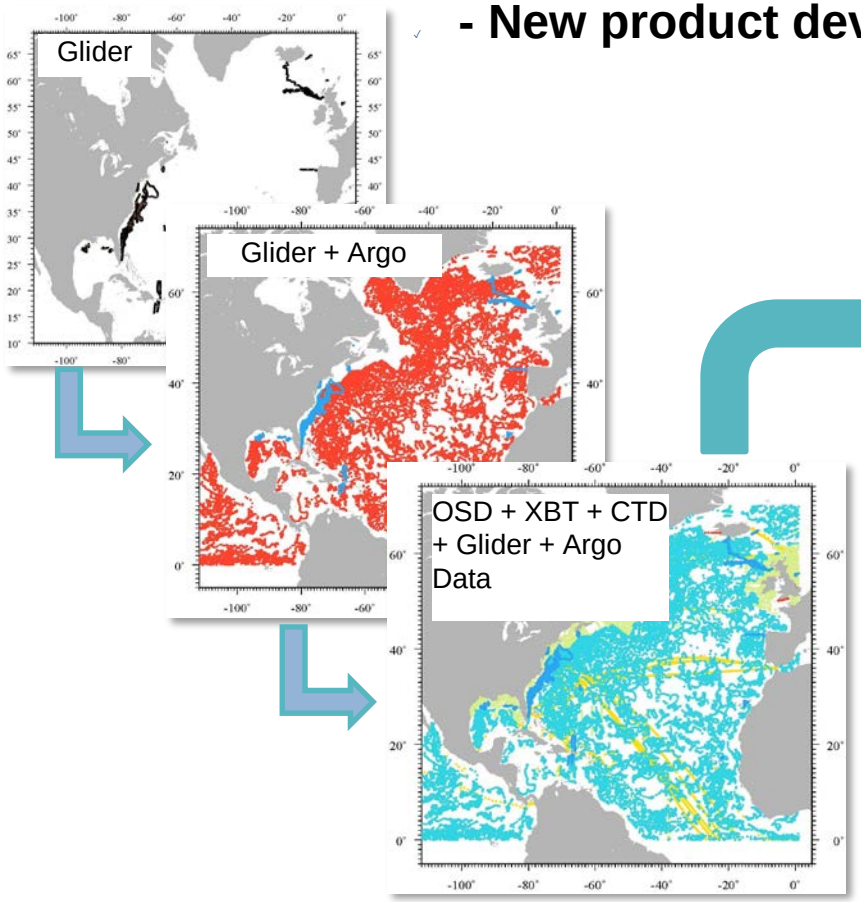
Color-coded WOD profiles in the Gulf of Mexico

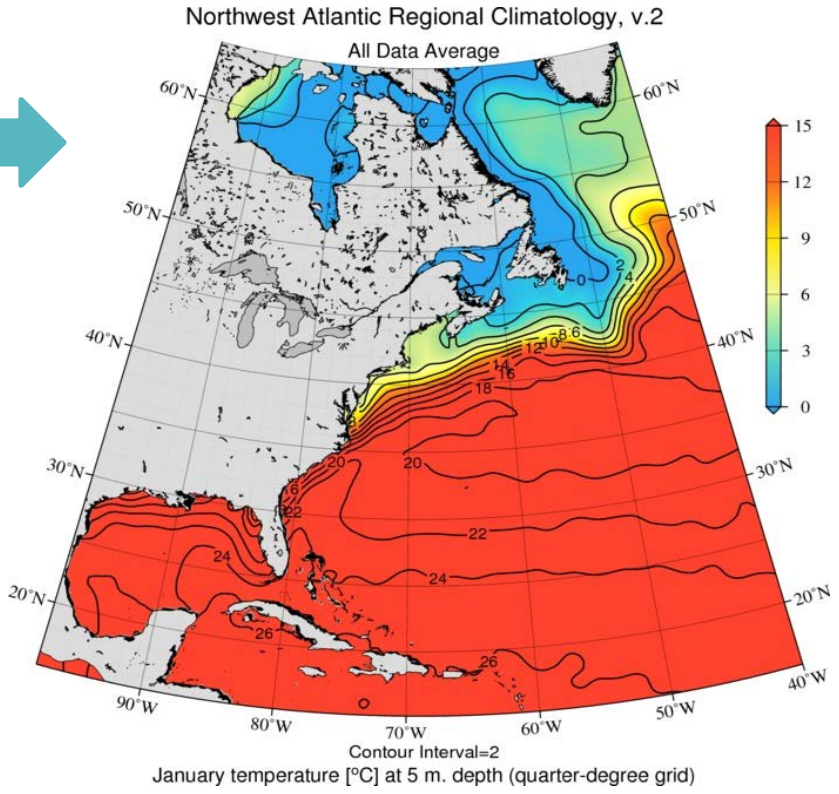


Number of WOD profiles over years in the Gulf of Mexico

Interoperability: Data Synthesis & Analysis

- **Combines data from different sources and platforms**
- **Enables comprehensive scientific data analyses**
- **New product development and publications**

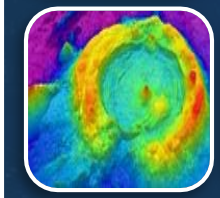




NCEI: Adding Value to Data

Earth Observing Systems



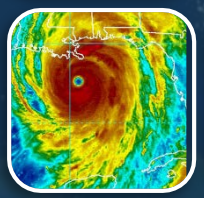
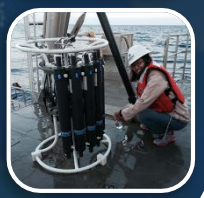











National Centers for Environmental Information

Scientific Data Stewardship

Research-quality products for decision-making

- Gulf of Mexico Data Atlas
- U.S. Extended Continental Shelf
- World Ocean Database
- U.S. Drought Monitor
- Ocean & Coastal Mapping & Products
- Coastal Digital Elevation Models
- Coastal Water Temperature Guide
- Global & U.S. Climate Summaries
- Temperature & Precipitation Climatologies
- Water Level Data
- Billion $ Disasters, Climate Extremes Index
- Harmful Algal Blooms Observing System

Learn More About NCEI

www.ncei.noaa.gov

The Nation's Trusted Authority for Environmental Information

IOOS Integrated Ocean Observing System

GCOOS GULF OF MEXICO COASTAL OCEAN OBSERVING SYSTEM

GCOOS DATA MANAGEMENT

Data Management Foundations:

- Community-based Vocabulary: Use of Climate-Forecast (v59) of CF Standard Names, and IOOS-OBIS Marine Biogeography terms (v2.1)
- Data format: CSV+ISO 19115-2 Metadata, and netCDF4 compliant to IOOS/NCEI requirements (CF, GCMD Science Keywords v8.6 rel., ACDD 1.3)
- Data Services: Direct Access (interactive interface available), OGC SOS (Sensor Observation Service), ERDDAP/TDS (supports many data endpoints and formats), and CKAN
- Data Access: Internet-based (HTTP/HTTPS, SFTP, RESTful, WMS/WCS)
- Data Policy: Employs QARTOD QA/QC recommendations, archive to NCEI and Open Data policies

IOOS Integrated Ocean Observing System

GCOOS GULF OF MEXICO COASTAL OCEAN OBSERVING SYSTEM

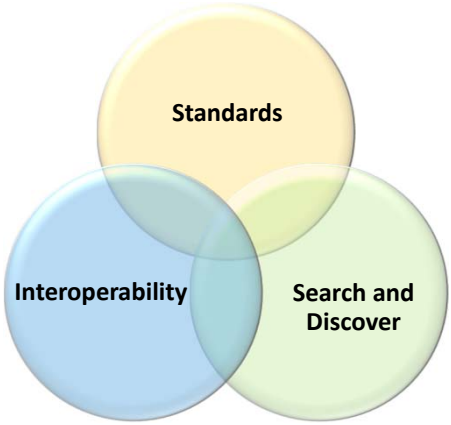
DWH LTDM Path Forward

Open Science by Design

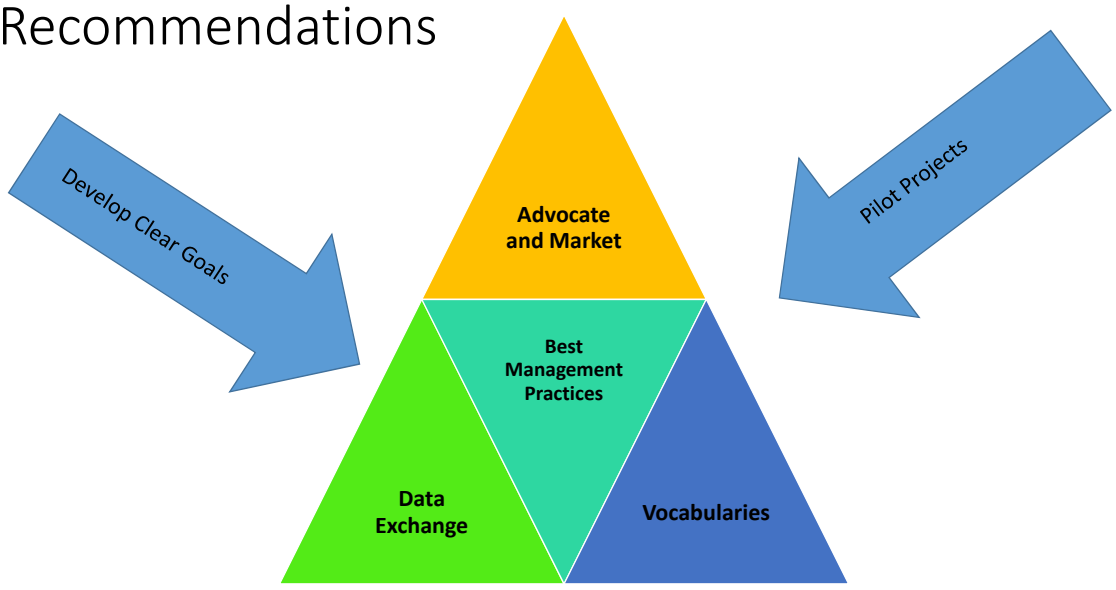
- Recommendations
  - Build a supportive culture
  - Provide training
  - Ensuring long term preservation and stewardship
  - Facilitating Data Discovery, Reuse, and Reproducibility
  - Developing new approaches

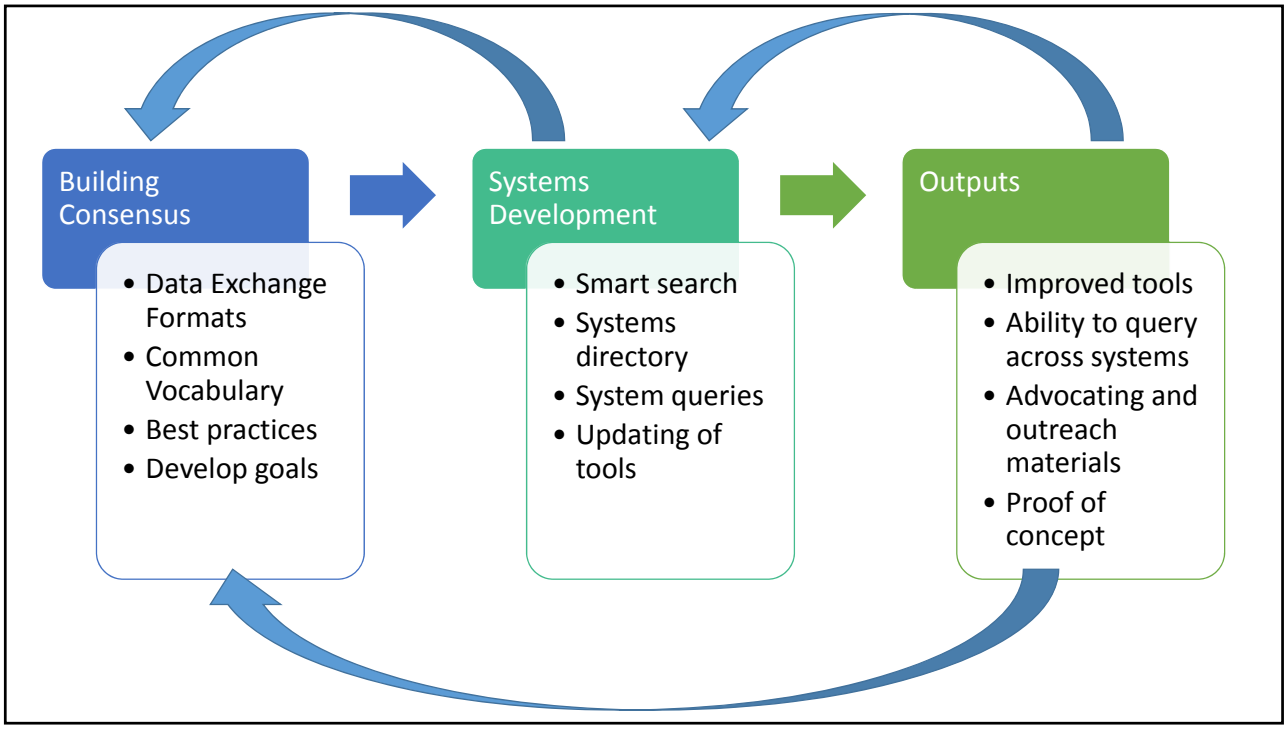
Working Group Outcomes

- Laid groundwork for current discussions
- Created useful outputs that can be expanded and improved



Recommendations





Possible workshop ideas

- Development of a data exchange framework
- Best management practices
- Vocabularies
- Search tools (fellowship)
- Marketing with SeaGrant

Path Forward

- Workshop report
  - With recommendations and actions
  - Other important materials available
- Core team regroup
  - Who will take responsibility of this going forward?
- Re-form the data funders group
- Presentations/examples of work already being done that relate to our recommendations
- Keep working and DO GOOD THINGS!!!

Stones Mooring

- Collaboration with Shell Oil
  - deepest oil and gas development project in the world
  - GRP will provide instrumentation that will provide real-time monitoring of ocean currents from 3,200 feet to the seafloor
  - Sea floor acoustic sensor
  - All data will be managed through GCOOS

8