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.”
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

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!!**
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.