2-2015

Environmental Disasters Data Management Workshop Report

Coastal Response Research Center

Follow this and additional works at: https://scholars.unh.edu/crrc

Recommended Citation

https://scholars.unh.edu/crrc/20

This Workshop Report is brought to you for free and open access by the Research Institutes, Centers and Programs at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Coastal Response Research Center by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.
Environmental Disasters Data Management Workshop Report

September 16 – 17, 2014

Coastal Response Research Center
University of New Hampshire

Publication Date: February 2015
Acknowledgements

The content for this workshop was developed in cooperation with NOAA Office of Response and Restoration (ORR) and the following Organizing Committee members:

Russ Beard, NOAA, National Coastal Data Development Center
Allen Dearry, National Institute of Environmental Health Sciences (NIEHS)
James Gibeaut, Harte Research Institute for Gulf of Mexico Studies, GRIIDC
Peter Giencke, Google, Crisis Response Team
Rick Greene, U.S. Environmental Protection Agency
Nancy Kinner, Coastal Response Research Center, University of New Hampshire
Christopher Krenz, OCEANA
Richard Kwok, National Institutes of Health (NIH)
CAPT Anthony Lloyd, U.S. Coast Guard (USCG)
Michael McCann, MBARI
Amy Merten, NOAA, Office of Response & Restoration
Aubrey Miller, National Institutes of Health
Mark Monaco, NOAA, National Centers for Coastal Ocean Science
Gary Petrae, Bureau of Safety and Environmental Enforcement (BSEE)
Pasquale Roscigno, Bureau of Ocean Energy Management (BOEM)
Stephanie Sneyd, Chevron
Evonne Tang, National Academy of Science (NAS)
Jeffrey Wickliffe, Tulane University, School of Public Health & Tropical Medicine

This workshop was facilitated by Dr. Nancy Kinner, Coastal Response Research Center (CRRC) at the University of New Hampshire (UNH). CRRC focuses on issues related to hydrocarbon spills. The Center is known for its independence and excellence in the areas of environmental engineering, marine science, and ocean engineering as they relate to spills. CRRC has conducted numerous workshops bringing together researchers, practitioners, and scientists of diverse backgrounds (including from government, academia, industry, and NGOs) to address issues in spill response, restoration and recovery.

We wish to thank all presenters for their participation in the workshop:
Charles Henry, NOAA Gulf of Mexico Disaster Response Center
Robert Haddad, NOAA Office of Response & Restoration, ARD
Jonathan Henderson, Gulf Restoration Network
Tracy Collier, Puget Sound Partnership
Aubrey Miller, National Institute of Environmental Health Sciences
Stephen Del Greco, NOAA National Climatic Data Center
Russ Beard, NOAA, National Coastal Data Development Center
Benjamin Shorr, NOAA, Office of Response & Restoration, Spatial Data Branch/ARD
Michael McCann, MBARI
Felimon Gayanilo, Harte Research Institute for Gulf of Mexico Studies
Steven Ramsey, Social & Scientific Systems, NIEH GuLF STUDY

We would also like to thank the breakout group leaders:
Carol Rice, University of Cincinnati, Department of Environmental Health
Henry Norris, Florida Fish and Wildlife Research Institute
Kim Jenkins, NOAA, National Ocean Service, ACIO
Mark Miller, NOAA, ERD, Technical Services Branch
1.0 Introduction

In the wake of the Deepwater Horizon (DWH) oil spill, a flood of information and new research has highlighted the need for improved coordination of data management for environmental applications (Figure 1). It is common for multiple entities (NGOs, academic institutions, responsible parties, federal and state agencies) to collect data that vary significantly in quality, collection methods, access, and other factors that affect use by others. These differences result in limitations for use of the data including comparing results or making inferences.

Figure 1. Courtesy of Russ Beard, NOAA, National Coastal Data Development Center

The Environmental Disasters Data Management (EDDM) project seeks to foster communication between collectors, managers, and users of data within the scientific research community, industry, NGOs, and government agencies, with a goal to identify and establish best practices for orderly collection, storage, and retrieval. The Coastal Response Research Center (CRRC) is assisting NOAA’s Office of Response and Restoration (ORR) with this effort.

The objectives of the EDDM project are to:
- Engage the community of data users, data managers, and data collectors to foster a culture of applying consistent terms and concepts, data flow, and quality assurance and control;
- Provide oversight in the establishment and integration of foundational, baseline data collected prior to an environmental event, based on user requirements;
- Provide best-practice guidance for data and metadata management;
- Suggest infrastructure design elements to facilitate quick and efficient search, discovery, and retrieval of data;
- Define the characteristics of a “gold standard” data management plan for appropriate data sampling, formatting, reliability, and retrievability; and
- Deliver workshop conclusions to end users in order to promote the use of the protocols, practices, or recommendations identified by participants.

An EDDM workshop was held on September 16-17, 2014 at the U.S. Fish and Wildlife National Conservation Training Center in Shepherdstown, WV. Participants at the workshop included individuals representing industry, government, NGOs, and academia on regional, national, and international levels who have a variety of experience related to data management during disasters (Participants in Appendix A). For the purposes of this workshop, environmental disasters are defined as floods, earthquakes, hurricanes, tornados, and discrete pollution events (e.g., oil spills).

The workshop consisted of plenary presentations and group breakout discussions (Agenda in Appendix B). It commenced with initial introductions and presentations on (1) how data are used for environmental disasters and (2) types of data management systems for these disasters.

The participants were split into breakout groups based on their expertise:
- Breakout Group A: Field Sample Collection (Data Collection/Sampling Protocols),
- Breakout Group B: Data Formatting/Entry (Data Consistency and Comparability),
- Breakout Group C: Data Reliability/Tracking (Accurate Transmission to Databases and QA/QC, Data Validation), and
- Breakout Group D: Discovery and Accessibility (Data to Users).

During the breakout sessions on Day 1, each group addressed questions that had been developed by the Organizing Committee (Breakout Group Questions in Appendix C). The discussions/answers from each breakout group were summarized and presented to all participants during a subsequent plenary session.

In the breakout session on Day 2, the groups discussed EDDM-related issues and challenges, the difficulty and priority to address them, and potential steps for a path forward. Each group presented the main points of its discussion in a final plenary session, which was followed by all participants discussing synthesis and next steps. Participants were given the opportunity to serve on the Organizing Committee as EDDM efforts move forward, or one of several topic-specific working groups that will be convened as a result of the discussions.

The following definitions are useful for the subsequent sections:

**Discovery:** User knowing that the data exists and then being able to find the specific data desired.

**Accessibility:** User accessing the data (by browser, mobile app, or other) and the level of access available (completely public or with restrictions).

**Data Model:** Rubric that documents and organizes data, defines how it is stored and accessed, and establishes the relationships among different types of structured and non-structured data.

**2.0 Plenary Sessions**
A summary of each presentation from the workshop is provided below. Slides for the presentations are available in Appendix E.

2.1 Use of data for environmental disasters

2.1.a Response – Charles Henry, NOAA, Office of Response & Restoration, Gulf of Mexico Disaster Response Center

Charles Henry provided an overview of how data are used for spill response during environmental disasters and discussed related data needs. He outlined the five key questions to be answered during a disaster: (1) What was spilled? (2) Where is it going? (3) What is at risk? (4) How will it hurt? and (5) What can be done to mitigate the hurt? Data are needed during environmental disasters to provide situational awareness and answer each of the five questions. To characterize the situation quickly, it is helpful if available data fit into a Common Operational Picture (COP) (a common available, easily displayed/used environment, such as the Environmental Response Management Application (ERMA®) used by NOAA). Both the quality of data, as well as how those data are processed and used, can critically affect decisions made in response to a disaster. Trajectories of spills are one critical component for monitoring and planning response efforts. To accurately predict trajectories, many types of data must be combined quickly. If the data are not accurate, or are processed or interpreted incorrectly, a poor trajectory can result. This could be disastrous because assets may be deployed inefficiently. Knowing the confidence in the available data is also important. Often at the beginning of a disaster, available information may be incorrect or sparse, but response decisions must still be made. It is important to update and correct this information/data as new information becomes available. The nature of disasters can add additional challenges to response (e.g., when levees broke in New Orleans after Hurricane Katrina, a road map was useless for planning response because many of the roads were flooded). Another challenge was the limited resources available, and resource use needs to be as effective as possible.

2.1.b Assessment – Robert Haddad, NOAA, Office of Response & Restoration, Assessment and Restoration Division (ARD)

The primary objectives of a Natural Resource Damage Assessment (NRDA) are to: (1) determine the extent and magnitude of injuries to the natural resources as a result of the release/spill and any injuries caused by the response activities, and (2) develop and implement appropriate restoration. The ability to integrate considerable amounts of different types of high quality data (and access related QA/QC information and metadata associated with them) and then see the results is critical to identifying and quantifying injury successfully. Assessment considers not only information derived during and after the spill, but also historical baseline data and material from various agencies. From a NRDA perspective, the term “data” includes: field and laboratory data, in situ measurements, climactic/meteorological data, photos, remote sensing, field observations and determinations, telemetry, model results, and metadata.

NRDA is a scientific and legal process – these both drive how data management is performed. With the potential for litigation, the data collected may be subject to the highest level of scrutiny. Each side in the case will search for inconsistencies that might preclude use of the data in court. The methods of data collection, analysis, and interpretation must be explained and defensible. If data management is not done properly, the data can be rendered useless and significant resources spent on data collection and management wasted.
Data management at NOAA ARD has evolved over the years. ERMA®, which is currently used by NOAA, enhanced the ability to see many types of data (including live feeds) and rapidly share them with stakeholders. Data Integration Visualization Exporting and Reporting (DIVER) is a collection of tools and processes that represent the most current evolution of data management at ARD. It standardizes and makes available to principal investigators/scientists the range of available data. DIVER enables data mining across diverse data types and spatially-explicit queries.

With more advanced instruments, the amount of information and data collected today far exceeds the amount that was collected historically. Much of the data collected 5 years ago or longer cannot be used, as those data need to be validated and managed so that they can be compared to current data. Funding is always limited, and each piece of data collected can be incredibly expensive, so the amount of data collected has to be balanced with the amount of available funding. The DWH case is an outlier because of its size (~200 million gallons of oil spilled) and scope. Education and communication among groups are needed to ensure data managers and users are not segregated. Ideally, everyone involved knows all the data and understands the analytical quality and all the steps that have occurred from collection to final interpretation.

2.1.c NGOs and the Public – Jonathan Henderson, Gulf Restoration Network, Coastal Resiliency Organizer

Jonathan Henderson provided an overview of the Gulf Restoration Network and the Gulf Monitoring Consortium (GMC). The Gulf Restoration Network (www.healthygulf.org), based in New Orleans, is a 20 year old member supported nonprofit environmental conservation organization. Its mission is to unite and empower people to protect and restore the natural resources of the Gulf region. The GMC is a rapid response alliance of various member organizations dedicated to monitoring and reporting pollution across the Gulf of Mexico. GMC uses satellite images and analysis of pollution detection trends to identify targets for monitoring. Airplane flights detect and verify pollution events using photos and GPS data. GMC has volunteers on land and water collecting samples and documenting impacts.

The GMC reports incidents to the National Response Center (NRC), and findings are publicly available. Websites such as the NRC should be able to withstand cyber-attacks. The biggest issue currently with the NRC system is transparency when a report is filed. Unless a Freedom of Information Act request is filed, the entity filing the report does not receive subsequent information about what happens after the report is filed except which agencies were notified of the spill/event. The EPA has better transparency than NRC. Because of the current lack of transparency with the NRC system, GMC cannot keep stakeholders informed about events. Often communities do not trust the agencies responding to disasters, and a clear and direct line of communication between scientists, government, NGOs, and industry is important to engendering trust. Data sharing among all parties also is important.

There is a critical need to respond and prepare the tools necessary for efficient data management. Members of the workshop highlighted two important points resulting from his talk: (1) More individuals need information in a disaster - how can they get it? (2) How do we use data generated by other sources (e.g., NGOs) to help inform additional research or other actions?

2.1.d Research: Ecological Health – Tracy Collier, Puget Sound Partnership

There are five types of data useful for determining ecological effects of oil spills: (1) water chemistry, (2) air chemistry, (3) chemicals in biota, (4) biological measures in individuals, and (5) population metrics.
The last three types of data are the hardest to get, but may be the most important. Pressing needs exist in the following areas: seafood safety, human health, dispersant use, and threatened and endangered species. These are interconnected, so “cross-walking” can occur between them in developing response strategies and sharing data. There are some 30,000 chemicals used in commerce, with only 4% routinely analyzed, and 75% unstudied. Many are designed to be toxic (pesticides) and 400 are estimated to be persistent. Some have unanticipated side effect (e.g. flame retardants). Petroleum contains thousands of unstudied chemicals.

Baseline data are critical information to have in the region of concern. Data must be quickly identified and captured. For Hurricane Katrina, there were no baseline data to compare with post storm conditions. Puget Sound has a long-standing monitoring program, but there is a lack of archived data in other areas. There are episodic attempts to establish this in some places, but it is not systematic or continuous sampling.

2.1.e  Research: Human Health – Aubrey Miller, National Institute of Environmental Health Sciences (NIEHS)

Environmental disasters come in all shapes and sizes, and human health is a component of most of them. Typically, health research in response to disasters has been quite limited and suffers from a number of problems including:

- Ad-hoc, convenience-based sampling,
- Non-systematic collection of health information,
- Late Data: Missing baseline & longitudinal health data,
- Exposures not measured,
- High risk groups not included: pregnant, elderly, pre-existing conditions,
- Lack of toxicity / health effect information for exposures, and
- Need for increased community engagement.

It is important to recognize that there are important human health questions associated with disasters that need to be addressed in order to prevent injuries and illnesses and promote recovery and future preparedness. Such questions include:

- What are the acute and long-term health implications (including mental health) of the exposures and stressors, especially among those most vulnerable?
- Are the impacted areas safe for people to live and work there?
- What must be known to help protect the public, address community concerns, and prepare for the future?

In order to address these questions we need to develop tools and processes to enable us to collect useful and timely information. Also, data and their management systems should be developed accordingly to support disaster response and research efforts.

With respect to the Gulf Oil Spill, the NIEHS and the Centers for Disease Control and Prevention (CDC) came together quickly to help coordinate and facilitate an assessment of data gaps and research needs related to spills and exposures. Subsequently, an Institute of Medicine (IOM) workshop held in New Orleans in June 2010 assessed the research needs related to the human health effects of the DWH spill. There are limited human health studies that have been performed for oil spills. Of 38 supertanker oil spills in the past 50 years, only eight have been studied for health effects, and all but one of those studies were short term. Only one study had estimates of exposure (using surrogate measures e.g.,
distance from spill). Exposures of concern during oil spills include: components of the crude oil, dispersants, mixtures of crude and dispersants, and chemicals resulting from burning.

Based on these and other considerations the IOM made the following recommendations:

- Longitudinal human health research is clearly indicated,
- Health studies should begin as soon as possible,
- Mental health & psychosocial impacts must be considered,
- Sensitive populations must be monitored,
- External stakeholders must be part of the process, and
- Data and data systems should be developed to support wider research efforts.

Subsequently, the NIEHS developed a number of intramural and extramural research efforts to respond to the IOM recommendations. The NIEHS GuLF STUDY (Gulf Long-term Follow-up Study) is an intramural health study of 32,762 oil spill clean-up volunteers and workers. The study follows participants for 10+ years and includes some combination of telephone interviews, in-home clinical assessments and biospecimen collection, comprehensive clinical exams, mental health and resiliency assessments, and a linkage to vital records and cancer registries. NIEHS also leads a NIH funded extramural DWH Research Consortia between four academic centers and community organizations focusing on research issues of concern to the coastal communities. The studies being performed by these groups will be looking at distinct populations (women, children, pregnant women, cultural/ethnic minorities) and will also cover seafood safety and community resiliency.

Additional lessons learned from oil spill research include the importance of rapid and ongoing communication with stakeholders, and the need for better capabilities to rapidly evaluate exposures and the resulting toxicity. Also, it is important to characterize the spill exposures to workers and the community to help understand any associated health effects. Such characterization and investigations include:

- Identify chemical profiles of different crude oils,
- Characterize changes in exposure impact due to oil weathering and degradation,
- Conduct research on chemical mixtures, and
- Document background ambient exposures as a baseline to evaluate impacts of future spills.

As part of the Gulf Oil Spill response, as well as responses to other disasters, a number of challenges for performing timely health research in response to disaster situations have been identified including: lack of baseline data (health and environmental), timeliness of funding awards and initiation of studies, study development (including getting approvals from Institutional Review Board (IRB), Office of Management and Budget (OMB), and obtaining Certificates of Confidentiality), identifying and enrolling study populations, and exposure reconstruction.

In response, the National Institutes of Health (NIH) have started a new Disaster Research Response (DR2) Project. This pilot project has been developed to help galvanize and accelerate the necessary infrastructure to mobilize quickly to perform needed health research in response to disasters. The DR2 will improve researchers access to data collection tools and create new platforms and networks to help facilitate engagement by federal, state, local, and community organizations in health data collection efforts. Objectives include the following:

- Development of a central repository for data collection tools and research protocols,
- Development of Rapid Data Collection Capability: baseline, clinical, and biospecimens; and new processes to hasten IRB and OMB approvals and address ethical issues,
• Timely collection of environmental data to accompany health data (including exploring roles of new technologies, social media, and “citizen science” in research),
• Training of intra/extramural disaster researchers,
• Development of Environmental Health Research Response Networks, and
• Development of a public website: “Disaster Research Responder”.

Next steps for the DR2 Project include efforts to facilitate the collection of exposure and environmental data by other agencies in support of the human health research studies and to increase our capabilities to perform toxicology research to further our understanding of various exposures of concern.

2.2 Existing data management systems, potential overlaps, shortfalls, opportunities for improvements, evolution of systems going forward

2.2.a Atmospheric Data – Stephen Del Greco, NOAA, National Climatic Data Center (NCDC)

NOAA’s National Climatic Data Center (NCDC) is the world’s largest archive of climate and weather data. NCDC is responsible for preserving, monitoring, assessing, and providing public access to the Nation’s climate and historical weather data and information. There is a rising demand for climate information, and the amount of climate data has increased tremendously in recent years. NCDC offers numerous climate products and services to a large variety of users on the local, regional, and national/global level, on weekly to decadal timescales. The Products and Services Guide available on the NCDC website (www.ncdc.noaa.gov) provides an overview of the offering. Services are delivered online, or via CD-ROM, DVD, computer tabulations, maps, and/or print. Data are accessed from disk (Storage Area Network) and tape (robotics system). NCDC does not store data in all formats, but instead data are formatted on demand to suit a specific need/format. Google Analytics is used to provide usage statistics and patterns. Drupal Content Management System provides the content infrastructure.

There are three data access portals: the Climate.gov portal, the Drought Portal, and the Model Portal. Many partners are involved in the portals, across NOAA, other agencies, and at the regional and state levels. The Climate.gov portal is designed to reach a wide segment of users – scientists, businesses, decision/policy-makers, news media, public, etc. The Drought Portal is geared toward providing critical information to decision-makers. The Model Portal provides access to reanalyses and numerical model output. NCDC also provides access to model data via the Climate Forecast System Reanalyses which is available online via NOMADS. NCDC also hosts international data - the Global Observing Systems Information Center (GOSIC) and the World Data Centers for Meteorology and Paleoclimatology. The GOSIC Portal provides one-stop access to data and information identified by the Global Climate Observing System, the Global Ocean Observing System, the Global Terrestrial Observing System, and their partner programs. The World Data Centers are a component of a global network of sub-centers that acquire, catalog, archive, and facilitate international exchange of scientific data without restriction.

The Climate Data Online (CDO) system and GIS Map Services provide centralized access to numerous US and global datasets and products. Data users are provided access to the data and metadata and allowed direct machine-to-machine access. Data visualization tools (e.g., Multigraph) provide graphical display of various parameters. For CDO, a “Batch” process allows users to submit orders for data and receive a link via email to the data. The underlying structure of CDO includes Oracle databases with tiered server infrastructure. These services continue to be built out to accommodate additional datasets and products. NCDC also has a weather and climate toolkit, which is based on community developed tools and standards. It is a desktop application providing simple visualization and data export to various formats. It supports 22 data formats (Model, Satellite and Radar), and provides interoperability with
diverse user communities. It is interoperable with Google Earth - exporting 3D radar sweeps and isosurfaces for Google Earth visualization. The Comprehensive Large Array-Data Stewardship System (CLASS) website (www.class.noaa.gov) provides users with access to CLASS information holdings and receives the users’ requests for information. CLASS manages data user’s logins, contact information, preferences, shopping cart, etc.

2.2.b Oceanographic Data – Russ Beard, NOAA, National Coastal Data Development Center (NCDDC)

The National Oceanographic Data Center (NODC) at NOAA manages the world’s largest collection of publicly available in situ and remotely sensed physical, chemical, and biological oceanographic data. It includes data taken from sources such as ships, CTD/Niskin casts, buoys, plankton tows, laboratory experiments, models, satellites, gliders, ocean currents, instrumented animals, and Expendable Bathythermograph (XBT). The NODC website (Nodc.noaa.gov) provides a list of all the available products. NODC’s data are being used for aquaculture, policy, ocean sciences, hazards response, national defense, industry, and climate-related work. Data management should be judged by its usefulness to current and future users.

The National Coastal Data Development Center (NCDDC), a Division of NODC, provides comprehensive end-to-end data management for the coastal environment. It has a regional approach, with a wide constituent base and liaison officers for customer service and user outreach. It provides metadata development (semantic search and ontologies), data discovery, mining, access, transport, archive, entry tools, collaborative web tools, data integration and fusion, geospatial enablement and visualization (e.g., ARC GIS and Google map), and biological data considerations.

NODC hosts global data sets of satellite and in situ data. The NODC Advanced Very High Resolution Radiometer (AVHRR) Pathfinder Version 5.2 sea surface temperature (SST) Climate Data Record provides the longest (1982 – 2012), most accurate, and highest resolution, consistently-reprocessed SST climate data record from the AVHRR sensor series. The World Ocean Database (WOD) and World Ocean Atlas (WOA) provide quality controlled comprehensive data collection and global in situ climatologies of temperature, salinity, oxygen, and nutrient measurements. The WOA is created from the WOD, and is a set of objectively analyzed climatological fields and associated statistical fields of observed oceanographic profile data interpolated to standard depth levels.

The NOAA Gulf of Mexico Data Atlas (gulfatlas.noaa.gov) provides digital discovery and access to Gulf data. Based on the traditional atlas format, it allows a wide range of users to browse a growing collection of datasets seen as map plates. The goal of the Atlas is to provide access to datasets that characterize baseline conditions of Gulf of Mexico ecosystems in order to assist long-term research, monitoring, and restoration programs. It includes metadata, web mapping services, and data download and access links, as well as access to Representational State Transfer (REST) services. The Atlas benefits from over 30 federal, state, non-governmental, and academic partnerships.

NCDDC’s OceanNOMADS (National Operational Model Archive and Distribution System for Oceans) is a web portal providing access to output from data-assimilating ocean-models from NOAA and Navy. It supports NOAA research on marine ecosystems and can be a backup (note: not primary) data source during events. Data from operational, data-assimilating ocean models provide 4-D ocean state estimates, and web tools simplify the task of accessing model data in useful formats. OceanNOMADS staff have worked with NOAA and academic scientists on oceanographic input for whole-ecosystem models as well as marine habitat, larval transport, and marine mammal ecology studies. OceanNOMADS is a data source for OR&R’s GNOME Online Oceanographic Data Server (GOODS), however
OceanNOMADS is operated primarily as an aid to retrospective analysis, and so does not guarantee reliable real-time data delivery during an event.

The National Centers of Coastal Ocean Science (NCCOS) provides coastal managers the information and tools they need to balance society’s environmental, social, and economic goals. NCDDC is working with NCCOS to create a geoportal-based application to enhance easy discovery of and access to the NCCOS data inventory.

A common data model should be platform and format independent. It should stretch across different users, with a consistent vocabulary and glossary. Multiple formats can be used and integrated, as opposed to needing a standard format. If everyone can agree on the metadata (suggests the nine parameters of metadata), then anyone can search for, locate, and discover the data. DIVER is an example of a model that contains different types of data and uses best practices to provide transparency, discoverability, and accessibility.

2.2.c Chemistry Data – Benjamin Shorr, NOAA, Office of Response & Restoration (ORR), Spatial Data Branch/ARD

A data warehouse integrates and makes information and data available from one location. Standard tools are generally used to collect and manage the information. The recommended approach is flexible and scalable. A data management effort in the midst of an emergency will default to existing tools and processes. The sooner field collected and lab processed data streams are integrated, the better the connections and management of the data. Ideally, data are combined beyond high level metadata. One of the ultimate goals is to provide environmental intelligence (using an online query to make an informed decision) and make information available in a useful format. Often in disasters, data have to be managed with an agile development approach (i.e., not all necessary information is known in the moment, but data management must move forward regardless and evolve along the way to meet ever changing needs). This agile approach was implemented during the DWH damage assessment, and frequent brief video conferences enhanced accountability, minimized silos, and helped to facilitate communication and create a team approach.

Common data model(s) (which refers to schemas or structures of data organization) should be flexible and scalable, with the ability to query across types of information. Data delivery and query requirements should drive how the information is managed. Data should be collected digitally if possible, and contain structured information (records with a field such as analytical data) and unstructured information (no records or columns to query such as reports or scanned field sheets). Data are connected by core fields at a high level across data sets/models. The first step in a common data model is to collate source data. The next step is extraction, transformation, and loading (ETL). ETL extracts data from homogeneous or heterogeneous data sources. Steps include defining the common model, accommodating additional data, and standardizing it. Source data and queries should be audited. Data are then brought into the data warehouse and integrated. Then data can be explored, visualized, and reported. Information collected can include: chemical and biological samples; oceanographic data; observations of shoreline, marsh, and species; animal telemetry; photography; and restoration data (potential and implemented, budget and activities). There are data specific information (e.g., results, methodology, units) and related information (field information, source data packages, reports, graphs). Existing standards and nomenclature can be used, and expanded and standardized, when necessary. Metadata is an important component. Existing contaminant chemistry source databases include: Historical Contaminant Chemistry (Query Manager), DWH Response collected (EPA ETL → NOAA QA/QC), and British Petroleum (BP) Natural Resource Damage Assessment (NRDA).
DIVER is an explorer data management and query tool developed and used by NOAA. It has a flexible query providing guided or custom searches, which can be saved for later use. DIVER provides export of data packages (including from NRDA and external datasets) which can then be used for analysis, visualization, and processing. Data tables showing query results are integrated with a mapping function. Information can be displayed as points, lines, and/or polygons and exported into GIS formats. Charts provide a summary of query results and are interactive, showing filtered data when clicked. Information is linked to source data files, and related data and information (e.g., documents, photographs, study notes). Metadata is a critical component, containing information such as query details (e.g., fields and data chosen), data details (e.g., when datasets were updated), data caveats (notes about the data), and field definitions. Metadata meets Federal Geographic Data Committee (FGDC) compliant (Extensible Markup Language (XML) and Hypertext Markup Language (HTML)) specifications; moving to International Organization for Standardization (ISO) geospatial metadata standards. DIVER is interoperable with ERMA® - query results can be shown in the ERMA® application. DIVER staff are currently working on enhanced data search functionality, and more widely available DIVER tools for the Gulf of Mexico, the Great Lakes, and nationally. NOAA is creating a flexible and scalable national approach with the goal of using DIVER as part of NOAA’s approach to data collection and management for the next environmental disaster. NOAA is also trying to address Internet data security needs and concerns of federal organizations, while also broadening the community accessibility and usability.

2.2.d Sensors – Mike McCann, Monterey Bay Aquarium Research Institute (MBARI)

Oceanographic research involves using a wide variety of surface and subsurface observation and sampling platforms (e.g., gliders, drifters, moorings, shipboard systems). For example, an autonomous underwater vehicle (AUV) is a mobile platform that measures properties (e.g., dissolved oxygen, nitrate, genetics, fluorescence, chlorophyll) while moving through the water. Data can be received from the AUV in real-time or delayed mode. A long-range AUV can be at sea for two weeks with continuous communication to shore. Examples of instruments placed on these platforms include the Seabird CTD, Wetlabs ECO Puck, ISUS Nitrate analyzer, Oxygen optode, and the Environmental Sample Processor.

Mike McCann discussed managing, visualizing, and understanding in situ oceanographic measurement data using the Spatial Temporal Oceanographic Query System (STOQS). STOQS is an open source geospatial database package that provides efficient access to these kind of data. Data ingest depends on using CF-NetCDF 1.6 discrete sampling geometry format for archiving information from the instruments. After loading into STOQS all of the data and metadata are viewable in a web-based user interface, which enables interactive exploration and analysis of large collections of data. The STOQS user interface provides these specific features:

- Spatial and temporal overview of all the data,
- Selection of data by platform, parameter, time, depth, and data value,
- Plotting of selected measured parameter along time-depth sections,
- Plotting of selected measured parameter on the map,
- Plotting any parameter against any other parameter, e.g. T-S plots,
- Visualizing the data in 3D, and
- Export to other formats, e.g.: CSV, JSON, KML.

The STOQS software is under continual development at MBARI. Current efforts include incorporating more laboratory analyses from physical samples and developing machine-learning algorithms to aid in decision making.
2.2.e Biological Data – Felimon Gayanilo, Harte Research Institute for Gulf of Mexico Studies

Biological data are commonly stored or archived in: (1) desktop computer or stand-alone system not accessible or shared with others, (2) databases developed by short-term funded projects that in many cases becomes inaccessible after the project funds are exhausted, (3) institution-wide information systems with institutional support and long-term initiatives, (4) federal, regional, and state programs that are generally accessible to the public, and (5) information systems from multi-national programs and efforts.

The type and structure of biological data are very much dependent on the objective of the study. Although the data management life cycle (which includes planning, collecting/generating, processing/analyzing, archiving, and discovering/re-using) is fairly standard, there are no community-wide encoding standards or vocabulary for biological data. These are just some of major issues that inhibit the re-use/repurposing of biological datasets from data centers. Instances of there being insufficient information to establish data provenance (metadata), absence of data review process (quality control), limited temporal and spatial coverages, and insufficient efforts to allow the interoperability of disparate information systems are the other issues with the management of biological datasets.

2.2.f Human Health Data – Steven Ramsey, Social & Scientific Systems Inc.; NIEHS GuLF STUDY

Objectives of disaster epidemiology include:

- Prevent or reduce the number of deaths, illnesses, and injuries caused by disasters,
- Provide timely and accurate health information for decision-makers, and
- Improve prevention and mitigation strategies for future disasters by collecting information for future response preparation.

Related surveillance work includes assessment of mortality (deaths) and morbidity (disease). A wide variety of resources, data/information, and data collection tools are used to assess these early in disaster situations and some examples were provided. Understanding the short and long-term health effects of disasters requires research that should be another component of the response to disasters. It is being done, but it takes too long to get into the field. Working with human subjects presents unique challenges and complications that are not associated with the study of animals and ecosystems. Human research protections require “Rules of engagement” for interacting with human subjects and strict study protocol must be followed, requiring much time and coordination. In addition, it can be difficult to get people to respond to and participate in research over long periods of time. Certain approaches work better than others depending on the population of interest. A workshop participant mentioned the idea of involving community organizations as one method that resulted in improved response. More work is needed to integrate data from sources such as weather satellites, monitors, sensors, and models with human specimens and questionnaire data to better understand exposures and related sequel. The nature of disasters can also present challenges to the logistical feasibility of conducting research, such as lack of power for refrigeration of samples, and closure of shipping as a means to send samples for analysis. Several examples of research study data management systems were discussed and some pros
and cons of each were presented.
3.0 Breakout Sessions

Based on their expertise, each workshop participant was assigned to one of the breakout groups:
- Field Sample Collection (Data Collection/Sampling Protocols),
- Data Formatting/Entry (Data Consistency and Comparability),
- Data Reliability/Tracking (Accurate Transmission to Databases and QA/QC, Data Validation), or
- Discovery and Accessibility (Data to Users).

The following is a summary of the discussions and conclusions for each of the breakout groups.

3.1 Breakout Group – Field Sample Collection (Data Collection/Sampling Protocols)

The Field Sample Collection Group answered the following questions during the workshop.

Is there a common data model that can be shared across entities?
No. A good place to start would be to create a performance-based conceptual model that unifies data types and variables.

What are the essential core parameters to be collected and recorded for any field collection (e.g., sample ID, date/time, lat/long)?
Essential core parameters should include media being sampled, as well as spatial and temporal components. At the detailed level, there is a long list of parameters, which can become a challenge between different groups. The goal should be to collect parameters that allow an evaluation of data quality and determination of utility with other data resources in order to evaluate exposure and effects.

What are the essential core parameters to be included in the metadata record?
Essential core parameters should be in compliance with Open Data Policy and standard-specific metadata guidelines. Additionally, information regarding how and why data were generated in a particular way (e.g., protocols, SOPs, strategies) and data use and access documentation should be included. A unique identifier and data contact/custodian should also be included. Mandatory and mandatory if applicable fields and their corresponding fields in a variety of metadata standards are available from the Open Data Project website at https://project-open-data.cio.gov/metadata-resources/.

What are the standard data types and protocols for emergency response?
There are numerous protocols for sampling particular agents in a particular matrix. Tiered protocols as needed for emergencies should follow a performance-based approach. This needs to be developed before the emergency because it can take too long once the disaster occurs.

What are best practices for reducing transcription errors?
Electronic field data entry reduces copying and transcription errors. An investment in this technology and the training to use it can substantially reduce data entry costs and errors and provide more rapid access to the results.

What are the roadblocks for getting data from field collection into an electronic format?
Electronic field data entry is preferred for reducing copying and transcription errors and eliminates later transfer to an electronic format. However, electricity (for charging) and Internet (for transmitting) are not always available on-site. Planning is needed to assure adequate storage capacity on-site, until data can be transmitted at a later time.
How is field collection designed to maintain Personally Identifiable Information (PII) (personal identification, human health etc.)?
Personal information should be maintained on a separate computer, with a linking identifier to the files of field data.

How is field collection designed to ensure accuracy of data?
Different collection plans have different criteria to ensure accuracy. Protocols can also depend on who is collecting the data. Ensuring accuracy of the data should be performance-based.

How is field collection designed to maintain data security?
This varies between agencies.

What are requirements for field data collection in order to ensure good data?
- Use of standard sampling protocol,
- Trained data collectors, particularly related to emergency response (protocol for preparedness),
- Coordination of sampling efforts,
- Performance-based,
- Standard Operating Procedures,
- Accurate and thorough metadata documentation, and
- Pre-plan for anticipated emergency response scenario needs, and incorporate into Sampling and Analysis Plan.

What are the types of media that should be sampled for an environmental disaster with respect to human and ecological health?
Both human and ecological health:
- Air,
- Soil/sediment,
- Water,
- Biological samples (e.g., urine, blood, fish bile),
- Characterization of toxicity of hazard (e.g., what chemicals present? e.g., oil, dispersant), and
- Archive a variety of samples that can be analyzed with high sensitivity later (for other analytes that are not known at time of incident). This can be done for background conditions too, prior to incidents. However, that can be expensive. If background sampling is cost limited, an alternative is to collect these samples outside of the disaster area during the event.

Note: Leveraging existing reference sites, as well as existing citizen science and NGO networks, should be considered to increase the data resource.

Human health specific (in addition to above):
- Dermal,
- Time, location, and activity (changes by day),
- Biological sampling (urine, blood, other human health information), and
- Mold, mildew.

Note: Focus initially on characterizing the exposure of the public and emergency responders.

The Field Sample Collection Breakout Group also developed the following table regarding issues and challenges, and a path forward. The group felt that all of these items were high priority.

<table>
<thead>
<tr>
<th>Issues and Challenges</th>
<th>Difficulty</th>
<th>Path Forward</th>
</tr>
</thead>
</table>

Coastal Response Research Center
### Breakout Group – Data Formatting/Entry (For Consistency and Comparability)

The Data Formatting/Entry Group answered the following questions during the workshop.
Is there a common data model that can be shared across entities?
No, there is not a common data model across all disciplines. What is considered the “best” data model depends on why data are being collected. Best practices and models exist, but there is nothing universal. However, there are many commonalities across disciplines. Many of the data models needed for disaster data management have a spatial component. Census Data, ISO191, and GIS are popular encompassing ones. Data models can have a similar structure, but within the models, there needs to be a glossary/index/dictionary that defines similar terms (e.g., variables, units) and clarifies them for comparison between models. An overarching model is not necessary, as long as there are standards. Crosswalk methods can allow existing data models to connect to each other. Adaptive management can be used as models are adopted and linkages are established.

What are the essential core parameters to be collected and recorded for any data collection (e.g., sample ID, date/time, lat/long)?
- Unique identifier,
- 4-D locations (time, X, Y, Z),
- Parameter measured or observed,
- Actual values,
- Units, and
- Metadata.

What are the essential core parameters to be included in the metadata record?
It is difficult to draw a clear line between data and metadata – they go “hand-in-hand”. Metadata is an essential part of data. Some of the core parameters listed for data collection apply to metadata (e.g., unique identifier). Other information for metadata includes:
- Information on what the dataset is, who collected it, what its purpose was?,
- Spatial reference (coordinate system and datum),
- Collection methodology,
- Instruments used,
- Limits of detection by methodology,
- Review status and what type of quality control was done,
- User restrictions, and
- Shareability (How can this be used or shared? Federal data?, Proprietary?, Contains PII?).

What are best practices for reducing transcription errors?
- Electronic data capture, when possible/practical,
- Transcription verification/dual data entry,
- Multiple people review, if possible/practical, and
- Safeguards in the system (unable to enter unrealistic data (e.g., that a person is 16ft tall)).

What are the rate limiting steps for getting data from field collection into an electronic format?
- Time,
- Office of Management and Budget (OMB) requirements (any federal data collection needs their clearance and it is a slow process),
- Data sharing and ownership issues, data sharing agreements,
- Difficulty reading handwriting issues, data sharing agreements,
- Non-standardized data (e.g., personal notes or a small sketch may end up in text fields),
- Platform dependency (Android vs. iOS, PC vs. Mac),
- No access to Internet,
• Running out of battery with electronic devices, and
• Procedural differences among agencies. No clear protocol or process established for data transfer. Adjusting the data into different digital formats for multiple stakeholders.

How are data formatting/entry designed to maintain PII (e.g., personal identification, human health, SSN, birth date)?
The focus should be on how much information is needed to identify a specific individual from a pool – this is different at each scale. Data are needed to make sure that the same person is not surveyed twice and to make sure people with the same name get surveyed individually. Only collect components of PII that are needed. Do not collect PII that is not needed. Perhaps PII may not be needed at all. If PII is available already, do not collect it again. Only use PII that has been collected when it is needed. PII does not have to be put into the electronic record – it can be kept archived. Encrypt the data.

How are data formatting/entry designed to maintain data security?
There needs to be safety and protection from collection to archiving. Data should have a “sharing status” providing information about who it can be shared with and how. For example, approval may be needed before data are shared, and/or there may be a part of the data that cannot be shared prior to public release. Once data are shared, they still have to be protected.

The Data Formatting/Entry Breakout Group also developed the following table regarding issues and challenges, and a path forward.
<table>
<thead>
<tr>
<th>Issues and Challenges</th>
<th>Difficulty</th>
<th>Priority</th>
<th>Path Forward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common data model(s)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common language (controlled vocabulary)</td>
<td>high</td>
<td>high</td>
<td>Each discipline develops its common data model. Have workshops among groups to develop common data model. If individual models are interoperable that may be sufficient.</td>
</tr>
<tr>
<td>Data structure</td>
<td>medium</td>
<td>high</td>
<td>Create pre-defined forms (e.g., have key tracking terms like keys, ID). Constraint lists (drop down menu - must choose).</td>
</tr>
<tr>
<td>Extensibility &amp; useability</td>
<td>high</td>
<td>high</td>
<td>Engage data and field practitioners in data model development and end user verification/testing. Run drills. Integrate organizations to keep everyone regularly informed of how data is being used. At conferences, each organization talk about their data. Frequent virtual meetings to check progress and discuss. Charter for each working group says what they do, frequency of meetings, etc. Have a representative held accountable and hold working groups accountable.</td>
</tr>
<tr>
<td>Data sharing &amp; ownership</td>
<td>highest</td>
<td>high</td>
<td>Draft memoranda of data sharing agreements so they can be executed at time of disaster. (Group agrees important item, uncertain of best solutions)</td>
</tr>
<tr>
<td>Core parameters recorded during data collection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique identifier quality: not unique, lengthy, complex</td>
<td>easy</td>
<td>high</td>
<td>Use barcodes to replace long IDs. Use meaningful/logical/sequential IDs so know if something went wrong (alphanumeric order).</td>
</tr>
<tr>
<td>4-D locations quality</td>
<td>easy</td>
<td>high</td>
<td>Agreement on time zone/reference time and encoding of time. Standardization and training on coordinate system, precision &amp; accuracy, significant figures, calibration, and crossing time zones. Standard operating procedures. Report inconsistencies immediately.</td>
</tr>
<tr>
<td>Parameter measured or observed quality</td>
<td>easy</td>
<td>high</td>
<td>Document the method used.</td>
</tr>
<tr>
<td>Actual values</td>
<td>easy</td>
<td>high</td>
<td>Calibrating equipment, agreement on flag values, significant figures. Checks to make sure the data 'make sense' in the big picture.</td>
</tr>
<tr>
<td>Units</td>
<td>easy</td>
<td>high</td>
<td>Standardize and be explicit.</td>
</tr>
<tr>
<td>Core parameters for metadata</td>
<td>Metadata</td>
<td>medium</td>
<td>high</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------</td>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>Confusion regarding definition of metadata</td>
<td>medium</td>
<td>medium</td>
<td>Transformation tools from machine generated nonstandard metadata to standard metadata. One-page clear guidance on what standards are. Make sure metadata gets filled out completely and it is provided by the person collecting the data.</td>
</tr>
<tr>
<td>Reducing transcription errors during data formatting/entry</td>
<td>Missing data</td>
<td>easy</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td>Invalid data</td>
<td>easy</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td>Illogical data (e.g., a male can't be pregnant)</td>
<td>medium</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td>Typos or inversions</td>
<td>easy</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td>Illegible data</td>
<td>easy</td>
<td>high</td>
</tr>
<tr>
<td>Version control</td>
<td>medium</td>
<td>medium</td>
<td>Gold standard with rules and goals; standard methodologies, routines, and checklists. Training and regular communications (pre-departure meetings, morning assemblies, etc.). People confirm version using and turn old versions in. Project Lead takes ownership.</td>
</tr>
<tr>
<td>Getting field data into electronic formats</td>
<td>Resources limitations: equipment &amp; people (analysis takes time)</td>
<td>medium</td>
<td>low</td>
</tr>
<tr>
<td></td>
<td>Time delay between collection and processing, and then loss of information that is needed for a complete record</td>
<td>easy</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td>Inconsistency in questionnaires, unable to compare groups</td>
<td>easy</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td>Operating equipment in hazardous areas</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td>Untrained teams that have different focuses</td>
<td>easy</td>
<td>high</td>
</tr>
</tbody>
</table>
### 3.3 Breakout Group – Data Reliability/Tracking (accurate transmission to database & QA/QC, data validation)

The Data Reliability/Tracking Group answered the following questions during the workshop.

**Is there a common data model that can be shared across entities?**
A metadata standard is needed. We can generate flexible and extensible usage of existing standards (models). QA/QC and metadata come in different levels. There need to be agreements in place between stakeholders, and active relationships, for data management before incidents occur.

**What are the essential core parameters needed for tracking the reliability of data?**
A set of core parameters should be developed and used. There should be a process that is known and followed by all; as part of the incident planning process. There should be transcription verification and subject matter expert validation. Having an “authoritative source” and verifying this is a big challenge.

**What are the system requirements for data reliability and tracking?**
There needs to be flexibility across platforms. Users should be accessing data through loosely coupled web services. IT issues will include security (need data backup), and archiving and maintaining the original. There needs to be security of data while in transit, and security of data at rest. There will always be a hybrid data system using both paper and electronic (need to track both) - the issue is the dynamic of the system.

**How are data reliability/tracking designed to maintain data security?**
Checksums can be used to detect errors that may have occurred during data transmission or storage. When applied, a checksum function or algorithm calculates a number based on the data. If the checksums calculated before and after storage or transmission are the same, it is a good indicator that the data has not been corrupted or altered. Data should be encrypted in transit and at rest. Version control can be employed regarding version information for devices that are collecting and processing data.

**What are the QA/QC processes used and are they community and/or scientifically accepted standards?**
Peer review is not practical at the incident. Third party validation of data should be considered.

**What is important for data reliability, QA/QC and validation when moving data from field collection into an electronic format?**
The physical object and electronic object should be tracked together along with their characteristics (i.e., disposal, location, sample id, sample expiration date, other information to allow the sample to be identified). A robust, flexible system and processes is needed to move data from the field into electronic form. Inconsistencies in nomenclature can present a challenge to proper interpretation. A common vocabulary must be established and consistently used.

What is the process for informing data generators/users about the status of data from collection to archives?
If this is not done well, the system may be viewed as not being transparent. There can be a notification process to inform people that their data has been received and for what it is being used. A reverse Chain of Custody communication should be implemented.

The Data Reliability/Tracking Breakout Group also developed the following table regarding issues and challenges, and a path forward. The group felt that all of these items were high priority.

<table>
<thead>
<tr>
<th>Common data model(s) &amp; core parameters for tracking reliability of data (combined)</th>
<th>Issues and Challenges</th>
<th>Difficulty</th>
<th>Path Forward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defining metadata standards</td>
<td>high</td>
<td>Engage NOAA and metadata experts to establish a training plan/path forward.</td>
<td></td>
</tr>
<tr>
<td>Adopting metadata standards</td>
<td>high</td>
<td>See above</td>
<td></td>
</tr>
<tr>
<td>Implementing metadata standards</td>
<td>high</td>
<td>See above</td>
<td></td>
</tr>
<tr>
<td>Building comprehensive QC plan (validation levels, useability, methodologies, versioning, links to publications, historical and baseline data, links to source, study plan, QAP)</td>
<td>high</td>
<td>Scan, analyze, adapt/adopt; review existing large-scale plans</td>
<td></td>
</tr>
<tr>
<td>Need agreement and active relationships for data before incidents</td>
<td>medium</td>
<td>See above</td>
<td></td>
</tr>
<tr>
<td>Easy translation and communication to public - common language/public outreach on understanding data quality and importance of metadata</td>
<td>medium</td>
<td>Make this a priority and work with incident command structure; forms, job aids, info inserts for incident management handbook &amp; workflows</td>
<td></td>
</tr>
</tbody>
</table>

**Maintaining data security**

| Defining data security - what is necessary (checksums, digital signatures, chain of custody) | medium | Policy recommendation; creating a plan; having a panel of experts from different domains to verify/validate protocols to validate authenticity |
| Defining who should have access, levels of access (system level, local admin rights, not requiring an IT person in the field) | high | See above |

**Developing community and scientifically accepted standard QA/QC processes**

<p>| Need for a coalition of government, public, scientific, academia, stakeholders | high | Identify, organize, and deal with the low-hanging fruit; implement plans noted above |</p>
<table>
<thead>
<tr>
<th><strong>Data reliability, QA/QC, and validation when moving data from field to electronic format</strong></th>
<th>Scanning original source data to store alongside electronic data file; transcription verification and validation</th>
<th>easy</th>
<th>Develop best practices for capturing and submitting data types; supply tools and training to enable field personnel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical and electronic objects to be tracked together along with their characteristics (e.g., disposal, location, ID, expiration date, sample identifying information)</td>
<td>medium</td>
<td>Determine importance of sample to set time to be kept; identify potential for legal ramifications</td>
<td></td>
</tr>
<tr>
<td>Robust, flexible, system and processes to move data from field to electronic form</td>
<td>high</td>
<td>Very important for QA/QC, see group A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Informing data generators/users about the status of their data &amp; tracking disparate data sets as they are processed</strong></th>
<th>Designing and implementing flexible infrastructure to provide multiple types of access. Clearly defined roles and responsibilities. Should have point of contact for feedback from data providers.</th>
<th>high</th>
<th>Have provisional pathway built in to data flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status on “push-pull” basis</td>
<td>Need subject matter expert</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A system to keep generators and users engaged/informed on where the data is in the process.</td>
<td>Information at a granular level to be able to communicate where things are in the process; and be able to track it</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Require a data source and a contact mechanism; whoever receives the data is now an “informer”</td>
<td>To provide data, must provide contact – chain of custody</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.4 Breakout Group – Discovery and Accessibility (getting data to the users)

The Discovery and Accessibility Group answered the following questions during the workshop.

Is there a common data model that can be shared across entities?

- No there is not a common data model, and there may never be one. However, the ability for multiple ones to work together (interoperability) is critical.
- Data sharing agreements need to be developed before events happen. The agreements would establish things like a common ontology, a standard file format for data exchange (including standardized metadata), and requirements that everything is platform independent (works with everything else).

What are the essential core parameters needed for discovery and accessibility?

- Essential core parameters are: spatial, temporal, and keywords.
- Ontologies are important for searching the data. Ontology is a classification, while vocabulary is a definition. Ontology can be used to show links between concepts (e.g., shrimp to chemistry).

What are the system requirements for discovery and accessibility?

- Robust infrastructure to host during emergency situation (lots of bandwidth)
- Online access
- Publicly accessible
- Platform independent
- Accessibility controls
- Vocabulary/ontology built into the system by software (i.e., user-centered design).
- Every sample accompanied by certain necessary parameters. Need to use common vocabulary.
- Metadata automatically generated as data is collected
- Valid links to metadata, data, contacts
- System has to be dynamic modified for access

The Office of Science and Technology Policy (OSTP) has an Open Data Policy that could be a good model/example. It provides guidelines on discoverability and access. Any data generator with federal funding will be required to follow it.

What are the best practices for data visualization, discovery, and accessibility?

- Consider what questions the end user is trying to answer when deciding how to structure information gathered. It should be a user-centered design.
- Develop an inventory of existing best practices that can be shared (there are a lot of them).
- Do not conflict with existing statutes, regulations, and guidance.
- Have a quality statement go along with the data, to tell how it can be used.
- Have good metadata, and provide good metadata training.

What are the best practices for maintaining PII (e.g., personal identification, human health, SSN, birth date) and Chain of Custody in discovery and accessibility?

- Follow guidance of Open Data Policy – there is a section on PII and controlled access.
- Use best practices of metadata (e.g., instead of name, use a position title).
- The National Coastal Data Development Center (NCDDC) has documented best practice for chain of custody during the Deepwater Horizon spill.
- Share best practices widely.
How is access to data granted to users given that PII data are available and need to be protected? The group expanded on this question and included any controlled data (e.g., preliminary data during a response, marine archeology, budgeting data).

- Security is an important consideration in maintaining data quality, as well as data accessibility.
- See Open Data Policy guidance. Training is needed.
- Make a list of restricted data types that could be shared and put in metadata records.
- Data can still be discoverable, even if it is not accessible, for transparency. If the user does not have the required credentials, they will see the data exists, but will not be able to access it.

The Discovery and Accessibility Breakout Group also developed the following table regarding issues and challenges, and a path forward.

<table>
<thead>
<tr>
<th>Common data model(s)</th>
<th>Issues and Challenges</th>
<th>Difficulty</th>
<th>Priority</th>
<th>Path Forward</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Many</td>
<td>medium</td>
<td>high</td>
<td>Ensure interoperability between the models through training, awareness, consistency of the existing systems, and core elements. Required by Open Data Policy for federal entities to move in this direction.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Core parameters for discovery and accessibility</th>
<th>Issues and Challenges</th>
<th>Difficulty</th>
<th>Priority</th>
<th>Path Forward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited awareness of core parameters/elements</td>
<td>easy</td>
<td>high</td>
<td></td>
<td>Nine core elements plus nine if-applicable elements from Open Data Policy. See “Common Core” elements. Make this information more commonly known through evangelizing, training, publications.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System requirements for discovery and accessibility</th>
<th>Issues and Challenges</th>
<th>Difficulty</th>
<th>Priority</th>
<th>Path Forward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infrastructure (hardware) exists for sharing data across entities.</td>
<td>easy - technical, industry, internally medium - process high - security</td>
<td>high</td>
<td></td>
<td>Develop data sharing agreements and have discussions before incidents.</td>
</tr>
<tr>
<td>Storage and archiving the data long term so it can be accessible</td>
<td>easy - storage medium - archive</td>
<td>high</td>
<td></td>
<td>Data centers already exist for archiving issues, but there are issues that go beyond that. Recognize that data centers are underfunded. Register data with, and make it known to, use Data.gov and HAZUS.gov.</td>
</tr>
<tr>
<td>Sharing process and policy information</td>
<td>easy</td>
<td>high</td>
<td></td>
<td>Develop a two pager from federal perspective to list/explain all policies affecting data access; share broadly.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Developing best practices for data visualization, discovery, and accessibility</th>
<th>Issues and Challenges</th>
<th>Difficulty</th>
<th>Priority</th>
<th>Path Forward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information officer when incident occurs to coordinate data accessibility</td>
<td>high</td>
<td>high</td>
<td></td>
<td>Adjust incident management handbooks to include this, which is a high level decision.</td>
</tr>
<tr>
<td>Need metadata training</td>
<td>easy</td>
<td>high</td>
<td></td>
<td>Online metadata training is currently available. Different levels of metadata training for different roles. Determine which entities need to take it.</td>
</tr>
<tr>
<td>Topic</td>
<td>Implementation of keywords and ontologies used by data generators</td>
<td>Developing best practices for maintaining Chain of Custody during discovery &amp; accessibility</td>
<td>Granting users access to data while maintaining PII and controlled access data</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>medium</td>
<td>medium</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lack of accountability and ownership</td>
<td>medium</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple processes for chain of custody, per collector</td>
<td>medium</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transparency of users knowing the data exists even if they cannot get access to the actual data</td>
<td>medium</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>When request comes for multiple data sets, uploader does not always have enough information about data and if it contains sensitive information.</td>
<td>medium</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>low during incident as everything is sensitive, high long term</td>
<td>Raise awareness of the Open Data Policy, which gives policy guidance on this issue. Make users aware of why data is being restricted.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Raise awareness of the Open Data Policy, which gives policy guidance on this issue. Responsibility falls upon authoritative source, who should know laws and policy. Flagged in the metadata.</td>
<td></td>
</tr>
</tbody>
</table>

It was noted during the question period that data management should be budgeted at the beginning of a project (15-25%). When it is not done until later it becomes more expensive. Every time budgets are renewed (for O&M etc.), the data management cost should be included.
The Discovery and Accessibility Breakout Group developed Figure 2 as a conceptual model.

**Figure 2.**
4.0 **Recommendations for the Path Forward**

The presentations and subsequent discussions resulted in a number of conclusions and “next steps” which should be part of the path forward and continued dialogue regarding data management during environmental disasters.

- **Use Existing Resources.** Mine existing resources (e.g., information, policies, data management plans) to ensure EDDM’s efforts do not overlap or contradict existing guidelines, and that established best practices are used where appropriate to avoid “reinventing the wheel”. Check the Open Data Policy, Ocean Exploration Research best practices, and others.

- **Review Open Data Policy.** Form a small working group (WG) to examine the Open Data Policy to determine if it can be the guiding principle for EDDM’s efforts. Include a representative from each type of organization (e.g., Federal, State, industry, NGO) on the WG.

- **Employ Existing Tools.** Enable the reuse of existing tools for new processes. Employ existing tools at all levels, rather than developing new tools/processes. Inventory existing tools at each step of the data process. Start at the field collection level – identify what information is collected in the field and how. List any existing tools currently used. See the Open Data Policy as a starting list. Identify gaps in tools.

- **Compile Background Data.** Develop, manage, and maintain a disaster data package for background data that refers to historical baseline data in specific regions, in order to understand changes post-disaster. This data package would mine existing baseline data and/or data currently being collected across all disciplines and identify any data gaps. This work must be done before disaster events occur. It is easier to do this before an emergency. It provides a dry run in preparation for an emergency. This effort could be the focus of a working group and would help drive the interconnectivity goal of EDDM.

- **Work Toward a Common Data Model and Interoperability.** Create a WG to document what specific common data models people are using across different disciplines and compile details regarding each one. Crosswalk existing common data models (i.e., translate between data models) to see if there are similar elements (perhaps under different names). Incorporate data dictionaries. At all levels (field collection, synthesis, analysis) inventory/identify existing ways to be interoperable. Find and build connections to create something that is more extensive and broad. Unify models that exist. Create a virtual infrastructure connecting the nodes. Demonstrate interoperability of the databases.

Figure 3 provides a conceptual model that incorporates these recommendations.
Figure 3.

- **Identify and Answer Fundamental Questions - User Centered Design.** Work with smaller working groups or society meetings to identify the mission and fundamental questions that need to be answered during a disaster response by domain/discipline. Questions would include: (1) What is their recommendation for a common data model? (2) What are their data requirements? (3) What data are collected? (4) What quality is required? These questions determine how/what data should be collected, which can feed into the model(s). A common data model and the related procedure/approach need to be flexible and adaptable.

- **Identify Data Dictionaries.** Identify data dictionaries, common language across disciplines, and have a clearinghouse of terminology. This can be included in data sharing agreements to help ensure consistent terminology.

- **Include NGO and Academic Data.** Consider how to incorporate NGO and academic data that feeds into decision making during a disaster. The data may have been collected with different objectives and timeframes, but the information is still important. Determine how the data can be incorporated into a common data model and decision making?

- **Incorporate Data Management Plans.** Data Management Plans must be incorporated into the Concept of Operations (CONOPS).

- **Include Data Managers in Response.** Data managers should be incorporated into the incident response plan and Unified Command.

- **Address Planning and Training.** Planning and training are essential, and there is a large need for
them. Create a WG to address what planning and training needs to be done. One thing missing currently is cross-training and collaboration across different sectors. Provide specific recommendations on cross-training (e.g., citizen science, human health). Make training available to producers and users of data, perhaps online. The National Response Team (NRT) might be a venue to move forward with this work.

- **Work Across Disciplines.** Pair different disciplines within working groups (e.g., pair environmental toxicology people with meteorological people to share experience with natural hazards). Weather and climate data are critical components, but data managers may lack experience with this kind of data.

- **Prepare Outreach Materials.** Prepare a one-page document (and slides) for all target audience organizations, with a consistent message regarding what EDDM is doing and why.

- **Perform Outreach.** Have an “inside champion” for each discipline, who is a member of appropriate organizations, to lead the outreach effort (e.g., a chemist within the EDDM group to take the message to the American Chemical Society). Consider sending someone to the organization’s meeting and/or plan a roundtable for the meeting. Pair people from different disciplines to go to these meetings as a team (i.e., one within society and one outside society) to share experiences. There is value in obtaining the perspective of various stakeholders. Mini working groups held at society meetings could gather their core data requirements.
5.0 Appendixes

Appendix A: Agenda
Appendix B: Breakout Group Questions
Appendix C: Breakout Group Members
Appendix D: Participants
Appendix E: Presentations
Appendix F: Group A: Field Sample Collection Breakout Groups Notes
Appendix G: Group B: Data Formatting/Entry Breakout Groups Notes
Appendix H: Group C: Data Reliability/Tracking Breakout Groups Notes
Appendix I: Group D: Discovery/Accessibility Breakout Groups Notes
Appendix A
AGENDA

Day 1: Tuesday, September 16

8:00 am  Registration

8:15 am  Welcome & Overview of Meeting
         Nancy E. Kinner, UNH Coastal Response Research Center
         Amy Merten, NOAA Office of Response & Restoration, Spatial Data Branch/ARD
         Russ Beard, NOAA National Coastal Data Development Center

8:45 am  Participant Introductions

Presentations: How are data used for environmental disasters?

9:10 am  ➢  Response: Charlie Henry, NOAA Gulf of Mexico Disaster Response Center

9:35 am  ➢  Assessment: Robert Haddad, NOAA Office of Response & Restoration, ARD

10:00 am BREAK

10:15 am ➢  NGOs and the Public: Jonathan Henderson, Gulf Restoration Network

10:40 am ➢  Research – Ecological Health: Tracy Collier, Puget Sound Partnership

11:05 am ➢  Research – Human Health: Aubrey Miller, National Institutes of Health

Presentations: Data Management Systems

Existing data management systems, potential overlaps, shortfalls, opportunities for improvements, evolution of systems going forward

11:30 am ➢  Atmospheric data: Stephen Del Greco, NOAA National Climatic Data Center

11:45 am ➢  Oceanographic data: Russ Beard, NOAA, National Coastal Data Development Center

12:00 pm LUNCH

1:00 pm  ➢  Chemistry data: Benjamin Shorr, NOAA, Office of Response & Restoration, Spatial Data Branch/ARD

1:15 pm  ➢  Sensors (e.g., ROV, AUV): Mike McCann, MBARI

1:30 pm  ➢  Biological Data: Felimon Gayanilo, Harte Research Institute for Gulf of Mexico Studies

*For the purposes of this workshop environmental disasters is defined as floods, earthquakes, hurricanes, tornados, and discrete pollution events.*
1:45 pm  ➢  Human Health Data: Steven Ramsey, Social & Scientific Systems; NIEH GuLF STUDY

2:00 pm  Instructions for Breakout Group: Nancy Kinner

2:15 pm  BREAK

2:30 pm  Breakout Groups - Session I

   Breakout Group A: Field Sample Collection (Data Collection/Sampling Protocols)
   Breakout Group B: Data Formatting/Entry (for consistency and comparability)
   Breakout Group C: Data Reliability/Tracking (accurate transmission to database & QA/QC, data validation)
   Breakout Group D: Discovery and Accessibility (getting data to the users)

4:00 pm  Plenary Report Out

5:00 pm  ADJOURN

6:15 pm  Dinner & Social Hour

Day 2: Wednesday, September 17

8:30 am  Recap and Recalibrate

8:45 am  Breakout Groups: Session II
   How do you overcome the challenges and move forward?
   Breakout Group A: Field Sample Collection (Data Collection/Sampling Protocols)
   Breakout Group B: Data Formatting/Entry (for consistency and comparability)
   Breakout Group C: Data Reliability/Tracking (accurate transmission to database & QA/QC, data validation)
   Breakout Group D: Discovery and Accessibility (getting data to the users)

12:15 pm  LUNCH

1:30 pm  Plenary Report Outs

2:30 pm  BREAK

2:45 pm  Plenary Discussion of Path Forward

3:45 pm  Closing Remarks

4:00 pm  ADJOURN
BREAKOUT GROUP QUESTIONS

Breakout Group A: Field Sample Collection (Data Collection/Sampling Protocols)

- Is there a common data model that can be shared across entities?
- What are the essential core parameters to be collected and recorded for any field collection (e.g., sample ID, date/time, lat/long, etc.)?
- What are the essential core parameters to be included in metadata record?
- What are the standard data types and protocols for emergency response?
  - Shoreline and/or soils
  - Watercolumn
  - Air
    - Human Health
  - Other
- What are best practices for reducing transcription errors?
- What are the rate limiting steps for getting data from field collection into an electronic format?
- How is field collection designed to maintain PII (personal identification, human health etc.)?
- How is field collection designed to ensure accuracy of data?
- How is field collection designed to maintain Chain of Custody?
- How is field collection designed to maintain data security?

Breakout Group B: Data Formatting/Entry (for consistency and comparability)

- Is there a common data model that can be shared across entities?
- What are the essential core parameters to be collected and recorded for any data collection (i.e., sample ID, date/time, lat/long, etc.)?
- What are the essential core parameters to be included in metadata record?
- What are the standard data types and protocols for emergency response?
  - Shoreline and/or soils
  - Water column
  - Air
    - Human Health
  - Other
- What are best practices for reducing transcription errors?
- What are the rate limiting steps for getting data from field collection into an electronic format?
- How are data formatting/entry designed to maintain PII (personal identification, human health, SSN, birth date, etc.)?
- How are data formatting/entry designed to maintain Chain of Custody?
- How are data formatting/entry designed to maintain data security?
Breakout Group C: Data Reliability/Tracking (accurate transmission to database & QA/QC, data validation)

- Is there a common data model that can be shared across entities?
- What are the essential core parameters needed for tracking the reliability of data?
- What are the system requirements for data reliability and tracking?
- How are data reliability/tracking designed to maintain data security?
- What is the process for informing data generators/users about the status of data from collection to archives?
  - What are the software and techniques for tracking disparate data sets for structured and unstructured data; where are they in process (at what lab, have they been analyzed? Have they been validated?)?
- Optional: What are the standard data types and protocols for emergency response?
  - Shoreline and/or soils
  - Water column
  - Air
  - Human Health
  - Other

Breakout Group D: Discovery and Accessibility (getting data to the users)

- Is there a common data model that can be shared across entities?
- What are the essential core parameters needed for discovery and accessibility?
- What are the system requirements for discovery and accessibility?
- What are the best practices for data visualization, discovery, and accessibility?
- What are the best practices for maintaining PII (personal identification, human health, SSN, birth date, etc.) and Chain of Custody in discovery and accessibility? Human subjects data protections?
- How is access to data granted to users given that PII data are available and need to be protected?
Appendix C
EDDM WORKSHOP BREAKOUT GROUPS

**Breakout Group A**: Field Sample Collection (Data Collection/Sampling Protocols)
- Lead: Carol Rice
- Recorder: Laura Belden
  - Kim Anderson
  - Courtney Arthur
  - Tracy Collier
  - Shawn Fisher
  - Jim Gibeaut
  - Jonathan Henderson
  - Sairah Malkin
  - Amy Merten
  - David Mica
  - Aubrey Miller
  - Geoff Scott
  - Patricia Stewart
  - Kent Thomas

**Breakout Group B**: Data Formatting/Entry (for consistency and comparability)
- Lead: Henry Norris
- Recorder: Ian Gaudreau
  - Matthew Foster
  - Amna Greaves
  - Kevin Hobbie
  - Matt Howard
  - Dan Hudgens
  - Stephane Leblanc
  - Lewis Leinenweber
  - Zach Nixon
  - John Parker
  - Steven Ramsey
  - Kari Sheets
  - Stephanie Sneyd
  - Laura Weems

**Breakout Group C**: Data Reliability/Tracking (accurate transmission to database & QA/QC, data validation)
- Lead: Kim Jenkins
- Recorder: Stefanie Tetreault
  - Steve Delgreco
  - Chander Ganesan
  - Felimon Gayanilo
  - Charlie Henry
  - Ann Jones
  - Anthony Lloyd
  - Wendy McDowell
  - Greg Minnery
  - Ben Shorr
  - Jason Weick
  - Kyle Wilcox

**Breakout Group D**: Discovery and Accessibility (getting data to the users)
- Lead: Mark Miller
- Recorder: Angela Sallis
  - Russ Beard
  - Dennis Beckmann
  - Brandon Brewer
  - Derek Eggert
  - Bob Haddad
  - Michele Jacobi
  - Hugh Johnson
  - Mike McCann
  - Jaci Mize
  - Peter Murphy
  - Mark Stenzel
  - Evonne Tang
Appendix D
Kim Anderson  
Oregon State University  
kim.anderson@oregonstate.edu

Courtney Arthur  
NOAA Marine Debris  
courtney.arthur@noaa.gov

Russ Beard  
NOAA National Coastal Data Development Center  
russ.beard@noaa.gov

Dennis Beckmann  
BP, Science and NRDA Data Assurance Manager  
dennis.beckmann@bp.com

Brandon Brewer  
Shawn Douglas, The Crisis Communicator  
brandon@shawn-douglas.com

Tracy Collier  
Puget Sound Partnership  
tracy.collier@psp.wa.gov

Steve Delgreco  
NOAA National Climatic Data Center  
stephen.a.delgreco@noaa.gov

Derek Eggert  
Chevron Corporate Emergency Response Advisor  
DHYG@chevron.com

Shawn Fisher  
U.S. Geological Survey  
scfisher@usgs.gov

Matthew Foster  
National Fish & Wildlife Foundation  
matthew.foster@nfwf.org

Chander Ganesan  
Open Technology Group, Inc  
chander@otg-nc.com

Felimon Gayanilo  
HRI/GOMRI  
felimon.gayanilo@tamucc.edu

Jim Gibeaut  
HRI/GRIIDC  
james.gibeaut@tamucc.edu

Amna Greaves  
MIT Lincoln Laboratory  
amna@ll.mit.edu

Robert Haddad  
NOAA ORR ARD  
robert.haddad@noaa.gov

Jonathan Henderson  
Gulf Restoration Network  
jonathan@healthygulf.org

Charlie Henry  
NOAA Gulf of Mexico Disaster Response Center  
charlie.henry@noaa.gov

Kevin Hbbie  
Oregon State University  
Kevin.Hbbie@oregonstate.edu

Matt Howard  
GCOOS/GOMRI/GRIIDC  
mkhoward@tamu.edu

Dan Hudgens  
Industrial Economics, Incorporated (IEc)  
dhudgens@indecon.com

Michele Jacobi  
NOAA ORR ARD  
michele.jacobi@noaa.gov

Kim Jenkins  
NOAA National Ocean Service, ACIO  
kim.jenkins@noaa.gov

Hugh Johnson  
NOAA Information Management Division  
hugh.johnson@noaa.gov

Ann Jones  
Industrial Economics, Incorporated (IEc)  
ajones@indecon.com

Stephane Leblanc  
Environment Canada  
stephane.leblanc@ec.gc.ca

Lewis Leinenweber  
Open Geospatial Consortium (OGC)  
leinenweber@opengeospatial.org

CPT Anthony Lloyd  
U.S. Coast Guard  
Anthony.S.Lloyd@uscg.mil

Sairah Malkin  
University of Georgia/Ecogic  
smalkin@uga.edu

Mike McCann  
Monterey Bay Aquarium Research Institute (MBARI)  
mccann@mbari.org

Wendy McDowell  
McDowell Safety and Health Services  
wmc dowell@mcdowellssafety.com

Amy Merten  
NOAA ORR ARD, Spatial Data Branch  
amy.merten@noaa.gov
Appendix E
WELCOME

Environmental Disasters Data Management Workshop

September 16 - 17, 2014

EDDM

September 16 & 17, 2014

Nancy E. Kinner
Coastal Response Research Center
University of New Hampshire
Logistics

• Fire Exits
• Restrooms at each end of hallway
• Coffee, tea, water available all day
• Dining: breakfasts (onsite in dining hall), lunches & snacks (all), welcome to bring food in meeting rooms
  • Breakfast 6:30 - 9:00
  • Morning snack (in hallway): 9:30 - 10:00
  • Lunch: As scheduled on agenda
  • Afternoon snack (in hallway): 2 - 3:30
  • Dinner: Tonight as scheduled; Tomorrow: 5:30 - 7:30

Logistics

• Dining: Tuesday night dinner (all)
  • Cash bar is located lower level in social area
  • 5:30 - 11:00
  • Welcome to bring drinks upstairs to dining area
• All meals - please sit in designated area behind dividers reserved for EDDM
• Logistical questions see Kathy Mandsager or me
Thank You

- Thank you to National Oceanic and Atmospheric Administration (NOAA)
  - Amy Merten - Office of Response & Restoration, Spatial Data Branch
  - Russ Beard - National Coastal Data Development Center

THANK YOU
Participants!
Coastal Response Research Center (CRRC)

• Partnership between NOAA’s Office of Response and Restoration and the University of New Hampshire
• Since 2004
  • UNH Co-Director - Nancy Kinner
  • NOAA Co-Director - Amy Merten

Overall CRRC Mission

• Conduct and oversee basic and applied research and outreach on spill response and restoration
• Transform research results into practice
• Serve as hub for oil spill R&D
• Facilitate workshops bringing together ALL STAKEHOLDERS to discuss spill issues and concerns
Workshop Objectives

- Engage community to apply consistent terms and concepts, data flow, and QA/QC.
- Provide oversight for foundational, baseline data collected prior to environmental event, based on user requirements.
- Provide best-practice guidance for data and metadata management.

Workshop Objectives

- Suggest infrastructure design elements to facilitate quick and efficient search, discovery, and retrieval of data.
- Define characteristics of “gold standard” data management plan for appropriate data sampling, formatting, reliability, and retrievability.
- Promote use of workshop protocols, practices, and recommendations.
After Workshop

- Website:
  - Presentations
- Report distributed
- Working Group meetings
- EDDM “Evangelism”

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
- Stop me if I am not doing this!
Participant Pledge

- Be Engaged
  - Turn off cell phones and computers, except at breaks
- Listen to others
- Contribute
- Speak clearly: Use microphones in plenary
- Learn from others
- Avoid side conversations
- Avoid using acronyms

Participant Introductions

- Name
- Affiliation
- Community/organization representation
- What is your interest for this workshop?
Environmental Disasters Data Management Workshop

Response Data Needs

Charlie Henry
Director, NOAA's GOM Disaster Response Center
Office of Response and Restoration

16 Sept. 2014

Disclaimer:
The information presented reflects only the views of the presenter, and does not necessarily reflect the official positions or policies of NOAA or the Department of Commerce.
NOAA Estimation of Floodwater Depth

The Five Response Questions?

- What was spilled?
- Where is it going?
- What’s at risk?
- How will it hurt?
- What can be done to mitigate the hurt?
What was spilled?

Search
Find response information for thousands of hazardous materials, including fire and explosion hazards, health hazards, firefighting techniques, cleanup procedures, protective clothing, and chemical properties.

MyChemicals
Build a list of chemicals. For example, substances involved in an incident response (such as a train derailment) or chemicals stored in your community.

Reactivity
See what hazards might occur if chemicals in your MyChemicals collection are mixed together.

Get started by finding a substance of interest with a search.
Learn more by checking the help for background information, a glossary of terms, and guidance on using this database.

About | Privacy Policy | Contact Us | User Survey | Mobile Site
CAMEO Chemicals version 2.4.1 rev 3.
Where is it going?

What's at risk?
“Information (data) is very critical to providing situational awareness.”

How will it hurt? Toxicity Data
What can be done to mitigate?

“Knowing the confidence in the information can be just as critical.”

Uncertainty – Spill Responders are OK with uncertainty, if they know the information is uncertain.

**Significant Figures – Implied Accuracy**

- 10,000,000
- 10,000,063
  - 63
  - 60
  - 100
  - 10.345 (ppm)
Types of Emergency Responses - All-Hazards

Types of Data – All Kinds

To “paint a good picture” quickly, it is very helpful if the data fits into what is called a Common Operation Prospective.”

common, easily displayed environment

ERMA
Last comment…

- Much of the early information known during an emergency response is wrong, and response decisions must be made anyway.

  (It is important to update and correct bad information.)
THE NOAA GOM DISASTER RESPONSE CENTER
Mobile, Alabama

Core Business Function:
A Hub for NOAA Emergency Preparedness, Response, Recovery, and Resilience...
Data and Information Management Needs for Natural Resource Damage Assessments
(why this is important to the users and abusers!)

Environmental Disasters Data Management Workshop
Sept 16-17, 2014
National Conservation Training Center

Robert Haddad, PhD
Amy Merten, PhD
NOAA Office of Response & Restoration and
Damage Assessment, Remediation, & Restoration Program

Why worry?

- The assessment of Natural Resource injuries following an oil spill or hazardous substance release is mandated under a variety of authorities, including OPA and CERCLA.
- Currently, NRD authority exists under:
  - CERCLA – U.S.C. 42 § 9601 et seq./ CFR 43 § 11 et seq.
    - See also CFR 40 § 300 et seq.
  - OPA – U.S.C. 33 § 2701 et seq./ CFR 15 § 990 et seq.
The primary objectives of the Natural Resource Damage Assessment (NRDA) are:

- To determine the extent and magnitude of injuries to the natural resources as a result of the release/spill and any injuries caused by the response activities, and
- To develop and implement appropriate restoration to make the public whole.

However, at its heart, the NRDA is both a scientific and legal process.

Some Objectives of this Workshop

- Common data models?
- Best practices for reducing transcription errors?
- Issues with getting data from field collection into an electronic format?
- Essential key fields needed to tie data types together?
- Infrastructure needs?
- Data visualization, discovery, and delivery best practices?
- Security best practices?
My Drivers... Science

- Science performed in any system, but especially within the complexity of natural ecosystems requires strong hypothesis testing.
  - In turn, our ability to test hypotheses and reduce or minimize natural variability inherent in these systems requires a considerable amount of high quality data.
  - Access to these data, to the underlying QA/QC information and other metadata is critical for any scientific investigation.
  - The ability to integrate all of the different data types and then visualize the results of our analyses are also critical to our success in identifying and quantifying injury.

My Drivers... Legal Framework

- Every NRDA is conducted within a legal framework. It is always our desire to settle, but we have to be prepared for litigation.
- Within this legal framework, all of the data we collect and use to develop our injury analyses are subject to the highest level of scrutiny.
  Thus, we must be able to explain and defend the appropriateness of how these data were collected, analyzed, and interpreted in the adversarial arena of the courts.
Really??

As an example, let’s discuss a NRDA case from the 1990s. To set the stage, this site was a waste site, where data collection had been ongoing for the past 5-7 years.

In preparing for NRDA, we began to review the data to assess what information was available and what information needed to be obtained.

The outcome of the review was that the QA/QC analysis showed the data to be significantly flawed or indefensible. As a result, we demonstrated that nearly $250K dollars worth of data were essentially useless.

This changed the opposition’s expectations very quickly and we were able to achieve a favorable settlement.

What does this have to do with DM?

The amount of information collected today far exceeds that collected historically – in almost every case:
- More laboratory analytical samples
- More in situ instrument-derived results
- More telemetry information
- More digital photos
- More modeling results

And for assessment purposes, we need to capture data and information not only derived from the assessment:
- We need to be able to access and integrate historical baseline data/information
- We need to be able to access and integrate much of the data/information developed during the response by the response agencies (think SCAT, dispersant usage, oil trajectory information [models, photos, observations, & remote sensing], etc.)
- All of these data become pertinent to the NRDA.
The need to quickly review and integrate disparate information is critical to our scientific understanding of complex systems.

The need to ensure all PIs are obtaining correct and similar information.

The ability to connect every piece of data/information back to its source and understand its complete provenance is, more and more, a key area of attack/need in litigation.

Finally, the need to organize, summarize, visualize, and explain these large amounts of data has increased seemingly exponentially.

There are others in the room who will talk about what types of data are important – and Amy touched on this point, earlier.

However, from the NRDA perspective, we see data as an encompassing term – A few broad examples of what we use:

- Field and laboratory collected analytical data & methodologies (e.g., analytical lab derived chemistry, in situ measurements of DO, etc.),
- Climatic/meteorological data,
- Photos and data derived from remote sensing
- Field observations (e.g., SCAT observations, species identification),
- Field determinations (e.g., how many critters in a quadrat?),
- Telemetry output,
- Laboratory observations,
- Mathematical model inputs & outputs,
- QA/QC data,
- All associated meta data, etc.
In our Division (ARD), we have been evolving along multiple pathways.

Early efforts where focused on what was needed for a specific case and generally involved a number of excel spreadsheets shared between staff.

Later evolution (e.g., Query Manager) resulted from a need to better integrate analytical data from multiple sites to draw more universal conclusions and develop widely applicable models.

QM was a foundational component of NOAA’s Watershed Database and Mapping Projects which provided a rapid method to create maps that displayed analyzed, sorted, and summarized data on a watershed-wide basis.
ERMA – the Environmental Response Management Application enhanced our needs for and ability to visualize many different types of data (including live feeds) and rapidly share these views with stakeholders.

DIVER – Data Integration, Visualization, Exploration, and Reporting – represents our current evolution of data management. This is a collection of tools and processes to standardize and make available to the principle investigators/scientists the vast range of data we have already discussed. This includes the ability for data mining across diverse data types with the ability to ask spatially explicit questions.
What this means to me

As a scientist, I need to have a high degree of confidence in the data and other information upon which I will base my conclusions.

I also need to have access to the widest base of knowledge available. Many answers will not be simple, and instead will be identified as a probability in a weight of evidence analysis.

What this means to me

As an NRDA practitioner, I need to know that the data and information I use is scientifically valid and that the interpretations I draw from those data will be scientifically and legally defensible.

As head of NOAA’s Damage Assessment group, I need to know that at a programmatic level, our Damage Assessment Claims are scientifically and legally defensible.
Data Management Plans – why?


"Now, as to my actual data management plan, here is how I plan to deal with research data in the future.

I will store all data on at least one, and possibly up to 50, hard drives in my lab. The directory structure will be custom, not self-explanatory, and in no way documented or described. Students working with the data will be encouraged to make their own copies and modify them as they please, in order to ensure that no one can ever figure out what the actual real raw data is.

Backups will rarely, if ever, be done.

When required to make the data available by my program manager, my collaborators, and ultimately by law, I will grudgingly do so by placing the raw data on an FTP site, named with UUIDs like 4e283d36-61c4-11df-9a26-edddf420622d. I will under no circumstances make any attempt to provide analysis source code, documentation for formats, or any metadata with the raw data. When requested (and ONLY when requested), I will provide an Excel spreadsheet linking the names to data sets with published results. This spreadsheet will likely be wrong -- but since no one will be able to analyze the data, that won't matter."
What We Do

• Rapid Response Alliance
• Space: Satellite images and analysis of pollution detection trends identify targets for monitoring
• Air: Overflights detect and verify pollution events with photos and GPS data
• Earth & Sea: Volunteers on land and in the water collect samples, document impacts
• Report incidents to National Response Center
• Publish our findings to the Public

GMC Flyover Flight Path—March 24, 2013
SKYTRUTH ALERT: NRC REPORT: OIL: CRUDE IN GULF 2013-07-08

**Report Details**
- **Report ID:** 555555
- **Incident Date:** 2013-07-08 07:00:00
- **Nearest City:** LA
- **Location:** GULF OF MEXICO LAT: 38° 20° 28’ LONG: 89° 16° 45’
- **Incident Type:** PLATFORM
- **Material:** OIL
- **Discharge:** WATER

**Skytruth Analysis**
- **Latitude:** 38.344444
- **Longitude:** -89.279167
- **Discharged ÖlVolume:** 43 gallons
- **Reported Spill Volume:** 43 gallons

**Report Description**
- **Incident Description:** COLLEGE REPORTING A LEAKAGE OF CRUDE OIL FROM A WELL DUE TO AN EQUIPMENT FAILURE WHILE PLUGGING THE WELL.

---

**Map**

---

**Location:** 38.575555, -89.275556

**Tags:** NRC, release, oil, LABO, ShaleGasAlerts
Incident Report # 1042025
McDuffie Coal Terminal, Mobile River
March 24th, 2013 10:00am.
GMC Website:
http://www.gulfmonitor.org/
Assessing ecological effects following environmental disasters, and data needs

Tracy Collier
Science Director (until Thursday)
Puget Sound Partnership
Tacoma, WA USA

Environmental Disasters Data Management Workshop

National Conservation Training Center
Shepherdstown, WV

September 16, 2014
Prior to working for PSP, I worked for 30+ years at NOAA’s Northwest Fisheries Science Center, where I managed assessments of several environmental ‘disasters’:
Prior to working for PSP, I worked for 30+ years at NOAA’s Northwest Fisheries Science Center, where I managed assessments of several environmental ‘disasters’:

Prior to working for PSP, I worked for 30+ years at NOAA’s Northwest Fisheries Science Center, where I managed assessments of several environmental ‘disasters’:

- **EXXON Valdez oil spill** (1989-1992) Ecosystem and human health
- **North Cape oil spill** (1996-1998) Ecosystem health
Prior to working for PSP, I worked for 30+ years at NOAA’s Northwest Fisheries Science Center, where I managed assessments of several environmental ‘disasters’:

- **EXXON Valdez oil spill** (1989-1992) Ecosystem and human health
- New Carissa oil spill (1999) Ecosystem health
Prior to working for PSP, I worked for 30+ years at NOAA’s Northwest Fisheries Science Center, where I managed assessments of several environmental ‘disasters’:

- New Carissa oil spill (1999) Ecosystem health
- Prestige oil spill (2002) Ecosystem and human health
Prior to working for PSP, I worked for 30+ years at NOAA’s Northwest Fisheries Science Center, where I managed assessments of several environmental ‘disasters’:

<table>
<thead>
<tr>
<th>Event</th>
<th>Dates</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Cape oil spill</td>
<td>(1996-1998)</td>
<td>Ecosystem health</td>
</tr>
<tr>
<td>New Carissa oil spill</td>
<td>(1999)</td>
<td>Ecosystem health</td>
</tr>
<tr>
<td>Prestige oil spill</td>
<td>(2002)</td>
<td>Ecosystem and human health</td>
</tr>
<tr>
<td>Hurricanes Katrina and Rita</td>
<td>(2005)</td>
<td>Human health</td>
</tr>
</tbody>
</table>
Prior to working for PSP, I worked for 30+ years at NOAA’s Northwest Fisheries Science Center, where I managed assessments of several environmental ‘disasters’:

- New Carissa oil spill (1999) Ecosystem health
- Prestige oil spill (2002) Ecosystem and human health
- Hurricanes Katrina and Rita (2005) Human health
- Cosco Busan oil spill (2007-2009) Ecosystem health
Prior to working for PSP, I worked for 30+ years at NOAA’s Northwest Fisheries Science Center, where I managed assessments of several environmental ‘disasters’:

- **EXXON Valdez oil spill** (1989-1992)  Ecosystem and human health
- **North Cape oil spill** (1996-1998)  Ecosystem health
- **New Carissa oil spill** (1999)  Ecosystem health
- **Prestige oil spill** (2002)  Ecosystem and human health
- **Hurricanes Katrina and Rita** (2005)  Human health
- **Cosco Busan oil spill** (2007-2009)  Ecosystem health

Since ‘retiring’ from NWFSC in 2010, I’ve been a science advisor to NOAA’s Oceans and Human Health Initiative, a technical advisor to NOAA’s marine mammal and sea turtle TWGs for the DWH NRDA, and have been informally liaising with NIEHS on their DWH long-term human health study (GuLF).
Types of data useful for determining ecological effects of oil spills, focused on fish and higher vertebrates
Types of data useful for determining ecological effects of oil spills, focused on fish and higher vertebrates

Water chemistry
Types of data useful for determining ecological effects of oil spills, focused on fish and higher vertebrates

Water chemistry

Air chemistry
Types of data useful for determining ecological effects of oil spills, focused on fish and higher vertebrates

Water chemistry

Air chemistry

Chemicals in biota
Types of data useful for determining ecological effects of oil spills, focused on fish and higher vertebrates

Water chemistry

Air chemistry

Chemicals in biota

Biological measures in individuals
Types of data useful for determining ecological effects of oil spills, focused on fish and higher vertebrates

- Water chemistry
- Air chemistry
- Chemicals in biota
- Biological measures in individuals
- Population metrics
New research on petroleum in the water column is raising concerns for eggs and larvae

Oil spills and fish health: exposing the heart of the matter

JOHN P. INCARDONA, TRACY K. COLLIER AND NATHANIEL L. SCHOLZ

Environmental Conservation Division, Northwest Fisheries Science Center, National Oceanic and Atmospheric Administration, Seattle, Washington, USA

Oceans and Human Health Initiative, National Oceanic and Atmospheric Administration, Silver Spring, Maryland, USA

Address all correspondence to john.incardona@noaa.gov

Journal of Exposure Science and Environmental Epidemiology published online 10 November 2010;
doi:10.1089/jes.2010.51

The chemical complexity of crude oil and its fuel products poses many important challenges for exposure science in marine ecosystems that support productive fisheries throughout the world. Meeting these challenges will enable better decisions on approaches to protecting and restoring these ecosystems.

BACKGROUND

Major oil spills typically trigger heightened public concern for highly visible species such as birds and marine mammals. However, because these events do not occur every day and are difficult to study, we know much less about the unseen and more subtle effects of oil exposure in marine ecosystems. Consequently, academic
Cardiac dysfunction from exposure of fish embryos to very low levels of weathered crude oil

**control 39 hpf**
HR @ 56 hpf = 189 ± 8

**weathered ANS crude 39 hpf**
HR @ 56 hpf = 143 ± 12, p < 10^{-8}
Cardiac dysfunction from exposure of fish embryos to very low levels of weathered crude oil

control 39 hpf
HR @ 56 hpf = 189 ± 8

weathered ANS crude 39 hpf
HR @ 56 hpf = 143 ± 12, p < 10^{-8}
Air chemistry data is a major need for assessing the effects of oil spills on human health;
After the EXXON Valdez oil spill, Pacific herring were heavily impacted. Air chemistry data is a major need for assessing the effects of oil spills on human health;
The fate of chemicals in biota after oil spills

**PAHs**

**Vertebrates:**
- fast, efficient metabolism
- Uptake
- Metabolites
- Reactive intermediates
- Released in bile

**Invertebrates:**
- slow, inefficient metabolism
- Uptake
- Accumulation in tissues
- Little or no PAHs in issues
- PAHs in tissues

Little or no PAHs in issues

Released in bile
The fate of chemicals in biota after oil spills

PAHs

**Uptake**

**Vertebrates:**
- fast, efficient metabolism

**Invertebrates:**
- slow, inefficient metabolism

**Metabolites**
**Reactive intermediates**
**Accumulation in tissues**

**Released in bile**
**Little or no PAHs in issues**
**PAHs in tissues**

Accumulation in tissues

Uptake
Sampling bile from herring using a syringe
Biological measures in individuals, taken during dolphin health evaluations
Biological measures in individuals, taken during dolphin health evaluations

- Physical exam & ultrasound
- Blood, urine samples
  - CBC, serum chemistry
  - endocrinology, immunology, serology
  - urinalysis
  - chemical analysis
- Blubber biopsy
- Satellite/VHF tagging
Biological measures in individuals, taken during dolphin health evaluations

Health of Common Bottlenose Dolphins (*Tursiops truncatus*) in Barataria Bay, Louisiana, Following the *Deepwater Horizon* Oil Spill


†National Centers for Coastal Ocean Science, National Oceanic and Atmospheric Administration, 331 Fort Johnson Road, Charleston, South Carolina 29412, United States
‡National Marine Mammal Foundation, 2240 Shelter Island Drive, Suite 200, San Diego, California 92106, United States
§Bayside Hospital for Animals, 251 Racetrack Road NE, Fort Walton Beach, Florida 32547, United States
‖Chicago Zoological Society, c/o Mote Marine Laboratory, 1600 Ken Thompson Parkway, Sarasota, Florida 34236, United States
¶Joint Office for Science Support, University Corporation for Atmospheric Research, 3300 Mitchell Lane, Boulder, Colorado 80301, United States
Focus Sites for *Tursiops* Nearshore Assessments
Longitudinal Vessel-Based Surveys of Dolphin Populations

• Remote biopsy tissue sampling
• Photo-identification for mark-recapture
  – Robust Design
  – Estimate abundance for each primary session
  – Estimate survival rates across primary sessions
  – Document calving events

Poster 32:6 – L. Schwacke et al.
Dolphin Photo-ID, Biopsy, & Capture-Release

Photo: NOAA, B. Rone
Types of data useful for determining ecological effects of oil spills, focused on fish and higher vertebrates

- Water chemistry
- Air chemistry
- Chemicals in biota
- Biological measures in individuals
- Population metrics
Types of data useful for determining ecological effects of oil spills, focused on fish and higher vertebrates

- Water chemistry
- Air chemistry
- Chemicals in biota
- Biological measures in individuals
- Population metrics

"Other"
Based on my involvement in and observations of the Deepwater Horizon oil release, and previous experience with EXXON Valdez, North Cape, New Carissa, Prestige, Hurricane Katrina, and Cosco Busan, in 2012 I told a group of European response specialists (PREMIAM—POLLUTION RESPONSE IN EMERGENCIES: MARINE IMPACT ASSESSMENT AND MONITORING) that there are pressing needs in the following areas, regarding pollution emergencies:
NOAA Fisheries’ Seafood Safety Response to the Gulf Oil Spill

Aquaculture America 2012
Las Vegas, NV

Calvin C. Walker and Cheryl L. Lassitter
National Seafood Inspection Laboratory
Office of Sustainable Fisheries
Pascagoula, MS
The use of dispersants

After the EXXON Valdez oil spill, Pacific herring were heavily impacted. The use of dispersants...
The effects of oil exposure on human health
After the EXXON Valdez oil spill, Pacific herring were heavily impacted.

Assessing the effects of oil spills on threatened and endangered species
After the EXXON Valdez oil spill, Pacific herring were heavily impacted.

Seafood safety

Human health

The toll free number is 1-855-NIH GULF.
That's 1-855-644-4853.

Dispersant use

T&E species
After the EXXON Valdez oil spill, Pacific herring were heavily impacted. Seafood safety, human health, and dispersant use were significant concerns. T&E species were also affected.
After the EXXON Valdez oil spill, Pacific herring were heavily impacted.

Seafood safety

Human health

The toll free number is 1-855-NIH GULF. That's 1-855-644-4853.

Dispersant use

T&E species
After the EXXON Valdez oil spill, Pacific herring were heavily impacted. Seafood safety, human health, and dispersant use are critical considerations for T&E species.
After the EXXON Valdez oil spill, Pacific herring were heavily impacted. Seafood safety, human health, dispersant use, and T&E species are crucial considerations.
After the EXXON Valdez oil spill, Pacific herring were heavily impacted. Seafood safety, human health, dispersant use, and T&E species are critical factors to consider.
After the EXXON Valdez oil spill, Pacific herring were heavily impacted. Seafood safety, human health, dispersant use, and T&E species are among the issues considered.
After the EXXON Valdez oil spill, Pacific herring were heavily impacted. Seafood safety, human health, dispersant use, and T&E species are all important considerations.
While each of these areas requires their own advance planning, response, and monitoring capabilities, interconnections mean that we need to crosswalk between them in developing strategies for handling pollution emergencies, and especially for sharing data.
Hazardous chemical releases associated with Hurricanes Katrina and Rita, and assessment of seafood safety.
30,000 chemicals in commerce

- 400 estimated to be persistent
- 4% routinely analyzed
- 75% unstudied
- Many are designed to kill (pesticides)
- Unanticipated (side) effects (e.g. flame retardants)
- Pharmaceuticals in sewage treatment discharge
- Petroleum = thousands of unstudied chemicals
After Katrina: tracking the toxic flood

BATON ROUGE

Three weeks after Hurricane Katrina ravaged the coasts of Mississippi and Louisiana, marines and engineers are starting to assess the safety of fish and shellfish exposed to toxic flood waters in the Gulf of Mexico.

The flood waters are teeming with *E. coli* bacteria and a wide range of chemicals (see Nature 437, 101; 2005). And engineers are appealing the need for the city to pass the Gulf coast. With its shrimp, oyster, crab and other fish stocks raised at around $3.1 billion, the coast is one of the most productive fishing grounds in the United States.

Shaler Cummings of the National Oceanic and Atmospheric Administration led a food safety team into the Gulf last week for a three-day expedition on board the research vessel Nancy Foster. Until then the box was being used to check the safety of the region's major ports, doing readings to check for obstruction under the water, among other tasks.

Cummings had just two days to organise the expedition — a huge challenge under normal circumstances and more so as personnel were in short supply. "You're trying to get a research team from a distant area who are hungry, stressed and looking for their family like everybody else," he says. So he recruited a team from the Northwest Fisheries Science Center in Seattle, Washington.

His team of some 15 scientists collected shrimp, oysters and Atlantic croakers — a common ground fish — and sent them to a Seattle laboratory to be tested for bacterial contamination and pollutants. Aware of the risk of a second public health disaster in the wake of the hurricane, the researchers worked around the clock. "Nobody sleeps," Cummings told Nature after three rough days aboard the Nancy Foster. "We're doing this to make sure the food supply is safe."

On their tour through the Gulf, the crew took samples from the muddy placenta that had spread from the Louisiana coast over hundreds of square kilometers. Katrina mobilized the water by washing sediments from the Mississippi into the sea, explains Cummings. And the fresh water is sitting on top of the salty water, spreading "like tea on a table top."

So far, the team has seen no evidence of the algal bloom that might result from the freshwater influx, and the circulation pattern of the water seems normal. But "we don't know what's in the water," says Cummings.

As well as testing water and sediment for pathogens and chemicals, the researchers collected and labelled fish samples for storage until they could be tested back in Seattle. It was close, feverish and stressful work. The liver and brain from all the fish were tested for fish-based contaminants such as polyethers, hydrocarbons and metals. The gut and gill will be tested for pathogens, and the muscle will be tested for mercury and other contaminants.

"It's an unprecedented situation," says Tracy Collier, head of the ship's wet lab and director of Environmental Conservation at the Northwest Fisheries Science Center. "We're trying to sample as broadly as we can," he explains, to detect anything that might harm people.

"The results will not be ready until the end of September but it is unlikely anyone will get ill in the meantime," says Do Scheninger of the Louisiana Department of Wildlife and Fisheries, because no fishing is likely to take place for many weeks. Katrina has brought the fishing community in the region to a standstill. Boats and pens have been destroyed, ice houses flushed, and fish processing centres damaged. US authorities estimate that about 4,000 fishermen in the area are now out of work.

Although serious attention is being paid to food safety, there is also concern that Katrina has damaged fish and shellfish stocks. Flood water has covered oyster beds along a stretch of inland bays with up to a metre of mud, and the system has collapsed.

Some wildlife seems to be bouncing back, however. Biologist Michael Baron was on board the Nancy Foster to watch for marine mammals in distress. On her third 12-hour shift, she said she had seen turtles and dolphins behaving normally.

Adrienne Kobl

"We're making sure the food supply is safe."

for mercury and other contaminants.

"It's an unprecedented situation," says Tracy Collier, head of the ship's wet lab and director of Environmental Conservation at the Northwest Fisheries Science Center. "We're trying to sample as broadly as we can," he explains, to detect anything that might harm people.
Chemicals in Our Waters

- fossil fuels
- metals
- pesticides
- other commercial chemicals
NIEHS Disaster Research Response: Recent Lessons & Future Steps

Environmental Disasters Data Management Workshop

September 16, 2014

Importance of Research
World Trade Center September 11, 2001

Widespread Contamination
Complex Mixed Exposures
>50,000 Healthy Workers

**Study 9 yrs later** *
- Asthma 28%
- Sinusitis 42%
- Lung Tests 42%
- PTSD 9%
- Panic 8%
- Depression 28%


Events come in all shapes and sizes
Environmental Health a part of most!

9/11 and Anthrax
Katrina, Rita, Wilma
Ike, Gustav
H1N1 Pandemic
Deepwater Horizon
11/2011 Tornadoes
Japan Earthquake Nuclear Event

Common Themes Across Disasters

- What are the health, including mental health, implications of the exposures & stressors, not just acute but long term?
  - Especially among those most vulnerable.

- Are the impacted areas safe for communities to live and work?

- What do we need to know to: help protect the public, address community concerns, and prepare for the future?

Deepwater Horizon Oil Spill, April 2010

- Interagency Research Work Group (May 2010)
  - NIEHS, NTP, NIOSH, ATSDR, SAMHSA, HHS/ASPR
  - NIEHS & CDC: Coordinated & facilitated assessment of data gaps and research needs related to spill & exposures

- Seed Funding by NIH Director
  - Dr. Collins: $10M support for research June 2010
  - Focus on health effects among those involved in various clean-up activities
Key Points
- Longitudinal human health research is clearly indicated
- Health studies should begin as soon as possible
- Mental health & psychosocial impacts must not be overlooked
- Sensitive populations need to be monitored
- External stakeholders must be part of the process
- Data and data systems should be developed to support wider research efforts

Limited Health Studies on Oil Spills

- 38 supertanker oil spills in past 50 years
- Only 8 studied for health effects, all but one cross-sectional or very short term

<table>
<thead>
<tr>
<th>Year</th>
<th>Name</th>
<th>Barrels of Oil</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989</td>
<td>Exxon Valdez, USA</td>
<td>270,000</td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>MV Braer, UK</td>
<td>620,000</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>Sea Empress, UK</td>
<td>525,000</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>Nakhodka, Japan</td>
<td>&gt;44,000</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>Erika, France</td>
<td>146,000</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>Prestige, Spain</td>
<td>460,000</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>Tasman Spirit, Pakistan</td>
<td>270,000</td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>Hebei Spirit, South Korea</td>
<td>73,000</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>Deepwater Horizon, USA</td>
<td>4,900,000</td>
<td>Dispersant Use &gt; 1.8 M gallons</td>
</tr>
</tbody>
</table>

Exposure Assessment:
- only 1 study had estimates of exposure (used surrogate measures e.g. distance from spill)
### Health Concerns from Previous Studies of Oil Spills

- **Acute**
  - Dermal, Eye, Respiratory Irritation

- **Longer-term health effects**
  - Pulmonary Symptoms and Abnormalities
  - Genotoxicity
  - Generalized Anxiety, Post Traumatic Stress Disorder, Depressive symptoms

---

### GOS health findings though August 27th (NIOSH Report August 13, 2010)

- **Injuries and Illness through July 27th**
  - N=2130 (1136 injuries (53%) 994 illnesses (47%))

- **For illnesses about 75% Onshore vs Offshore**
  - 192 Heat stress
  - 171 Multiple Symptoms (more than one organ system with no specific underlying cause)
  - 127 Headache / Dizziness
  - 122 Gastrointestinal
  - 78 Dermatologic
  - 42 General Symptoms (malaise, fatigue, non-specified allergic reactions)
  - 28 Cardiovascular

---

### Exposures of Concern

- **Crude Oil**
  - Polycyclic Aromatic Hydrocarbons (PAHs)
  - Volatile Organic Compounds (VOCs) (benzene, napthalene, toluene, xylene)
  - Heavy Metals (cadmium, nickel, lead, zinc)

- **Dispersants**
  - Detergents (sulfonic acid salts)
  - Solvents (2-butoxyethanol, propylene glycol)
  - Petroleum Distillates (paraffins, PAHs)

- **Burning**
  - PAHs, respirable particulate, hydrogen sulfide, sulfur dioxide

- **Other: Heat Stress, Physical Hazards, Mental Health**
**GuLF STUDY (Gulf Long-term Follow-up Study)**

- Prospective study of 32,762 adults involved in oil spill clean-up or support
  - Enrolled Mar 2011 to Mar 2013
  - Baseline telephone interview on clean-up jobs, symptoms, health
  - In-home clinical assessment and biospecimen collection – 11,210 from Gulf states
  - Followed 10 or more years
    - Telephone interview every 2-3 years
    - Subgroup with repeated mental health and resiliency assessments
    - Linkage to vital records and cancer registries
  - Comprehensive clinical exam (~4,000 from AL, LA) started 8/14
Deepwater Horizon Consortium

- 5-year $25.2 M program
- Four university/community partnerships
  - Tulane
  - LSU
  - Univ. of Florida
  - Univ. of Texas Medical Branch
- Steering group leadership
  - Includes GuLF STUDY
  - Input from NTP
- Distinct populations & foci
  - Women and children
  - Pregnant women
  - Cultural/ethnic minorities
  - Seafood safety
- Shared approaches
  - Seafood
  - Resiliency
  - Population studies
  - Community outreach & dissemination

Funding: NIEHS, NCI, NHLBI, NIMH, NIMHD, NINR, NCATS, OBSSR

Characterizing Spill Exposures to Understand Health

- Identify chemical profiles of different crude oils
- Better characterize changes in exposure impact due to weathering and degradation
- Conduct research on chemical mixtures
- Characterize background ambient exposures as a baseline to evaluate impact of future spills
Oil Spill Research Challenges

- **Study Populations: Workers and Volunteers**
  - Use of NIOSH roster & combining multiple lists (BP, national guard)

- **Study Development Process**
  - IRB, OMB, & Certificates of Confidentiality

- **Baseline Data for Comparison**
  - Available only for small fraction of cohort (e.g., Coast Guard)
    - Health information, biospecimens, relevant tests ?
    - Environmental baselines & monitoring (seafood, water, air, etc.)

- **Exposure Reconstruction**
  - Methods, sensitivity/specificity, time/location, area vs. personal samples, etc.
  - Multiple databases that need to be integrated
  - Available data difficult to use to reconstruct exposures

- **Timeliness of Extramural Awards & Initiation of Studies**

---

Oil Spill Research: Lessons Learned

- **Atypical workers involved in disaster responses:**
  - fisherman & others who lost jobs, unemployed from other areas

- **Rapid and ongoing communication with stakeholders**

- **Need better capabilities to rapidly understand exposures & evaluate toxicity of exposures**
Elk River WV Chemical Spill, Jan. 2014

Is it really safe?

- ~10,000 gallons of 4 methylcyclohexane methanol (MCHM) + polyglycol ethers (PPH) leaked into Elk River
- No water for over 300,000 residents, affecting some for more than a week
- About 500 patients seen in response
- Limited toxicology and health data available
- Missed opportunity to assess exposures and health impacts
- **Currently**: CDC looking at surveillance opportunities and NIEHS/NTP developing toxicology studies

Disaster Environmental Health Research Issues

- Ad-hoc, convenience based sampling
- Non-systematic collection of health information
- Late Data: Missing baseline & longitudinal health data
- Exposure data not measured to understand effects
- High risk groups: pregnancy, elderly, pre-existing illness
- Lack of toxicity / health data for exposures
- Minimal community engagement
Disaster Research Response (DR2) Project Genesis

- National Biodefense Safety Board Recommendations (Apr, 2011)

- Sep. 2012 NIH/ASPR Federal Partners Meeting: Identified Areas of Concern
  - Funding, IRB/OMB, Data collection tools, trained research workforce, infrastructure support, & community engagement

- Deployment of research too slow & Data is perishable!
  - H1N1 Response- treatment research, IRB issues
  - DWH Oil Spill- 9 months to start GuLF Study
  - Hurricane Sandy- 11 months to fund extramural efforts

“Timely research is critical to prevent injury & illness and support recovery”

Lurie, Manolio, Patterson, Collins, Frieden.  NEJM Mar 2013:

NIH Disaster Research Response (DR2) Project

Pilot project to help galvanize and accelerate needed infrastructure as part of a larger HHS Effort

Objectives

1. Central repository data collection tools & research protocols
2. NLM public website: “Disaster Research Responder”
3. Rapid Data Collection Capability: baseline epi., clinical, & biospecimens
4. Environmental Health Research Response Network (EHS Network)
5. Training intra/extramural disaster researchers
6. Share & Integrate: HHS/federal response & recovery frameworks
Research Responder Training & Education

• Training & Education
  1. National response plans & HHS mechanisms
  2. Training on DR2 Project & EHS emerging issues
  3. Health & Safety issues relevant to the disaster/situation

• Training Exercises on identified scenarios & issues

4/7 Port of Los Angeles Training Exercise
- USGS Tsunami Scenario
- 140 involved: fed, state, academia & community
- Evaluate DR2 Project concepts & support
- Discussion: integration, issues of concern

Tour of Area & Tabletop Training Exercise
Day 15

- A storage tank at a local oil refinery has caught fire resulting in a large plume of smoke and leakage into the coastal waters

- Local hospital ED’s and poison centers are experiencing increased complaints of respiratory, gastrointestinal, and neurologic symptoms

- Refinery workers and clean-up workers experience similar symptoms
NIH ASPR CDC IOM

Disaster Research Response Workshop: June, 2014*

Enabling Public Health Research During Disasters

- Build a broader network
- Frame a national research agenda & action items
- Integrate research into existing response structures
- Identify critical research needs & priorities
- Identify obstacles & barriers to research
- Discuss structures & strategies needed for deployment
- Share ideas, innovations, technology to support research
- Explore data collection tools & sharing mechanisms

*IOM Report available by November 2014

NIEHS Disaster Research Response Looking Forward

1. Build on DR2 Repository & NLM Website, Training, & Integration
2. RAPIDD Protocol for health data collection (IRB & OMB approvals / issues)
3. Expand “EHS Network” & collaborations with federal, state, academia, & communities
4. Exercises to further test research response strategies, protocols, field-implementation, and training.

5. Rapid collection of environmental data to go with health data!
   - Explore role of new technologies, social media, & “citizen science” in research
THANK YOU!
QUESTIONS?
For more information contact: CAPT Aubrey Miller, MD, MPH
miller.aubrey@nih.gov

Or email the DR2 Staff at:
dr2@niehs.nih.gov

Project Webpage
NOAA’s National Climatic Data Center

*World’s Largest Archive of Climate and Weather Data*

Presented to: Coastal Environmental Disasters Data Management Workshop
September 16, 2014

Stephen Del Greco
Deputy Chief, Climate Services and Monitoring Branch

---

NCDC Strategic Vision

**MISSION**

*Steward the Nation’s Climate Information*

NCDC is responsible for preserving, monitoring, assessing, and providing public access to the Nation’s treasure of climate and historical weather data and information.

**VISION**

*Be the Nation’s Trusted Authority on Climate and Historical Weather Information*

NCDC will be the most comprehensive, accessible, and trusted source of state-of-the-art climate and historical weather data, information, and climate monitoring.
NCDC Geographic Locations

- 160 Federal Employees
  - Alaska, Colorado, Hawaii, Maryland, Missouri, New York, North Carolina, Texas, Washington, Wisconsin
- 153 NCDC Headquarter Contractors
- 6 Regional Climate Centers
- 2 Cooperative Institutes

Rising Demand for Climate Information

- Sustainability of Marine Ecosystems
- Coasts and Climate Resilience
- Climate Impacts on Water Resources
- Changes in Extremes of Weather and Climate
- Agriculture
- Energy
- Health
- Transportation
Climate Products and Services

- Local
  - Snowfall Impact Index
  - Heating & Cooling Degree Days
  - Precipitation Outlooks
  - Billion $ Disasters, Climate Extremes Index
  - FEMA, disaster response
  - Energy Sector
  - Agriculture
  - Insurance

- Regional
  - Hurricane Tracks
  - Heat Wave Prediction
  - Drought Outlook
  - Construction, Infrastructure, Agriculture
  - Emergency Planners
  - Public Health Officials
  - Agriculture

- National & Global
  - Drought Monitor
  - Monthly State of the Climate Reports
  - Annual State of the Climate Reports
  - National Climate Assessment
  - Agriculture
  - Decision Makers

Storage Volume and Ingest Rate

Total Archive Volume: 13.4 Petabytes

Volume (Petabytes)

- 2000: 0
- 2001: 3000
- 2002: 6000
- 2003: 9000
- 2004: 12000
- 2005: 15000

Fiscal Year
NCDC Access to Data Received from Many Sources

- Voluntary U.S. Observers
- Global Weather Reports
- NCEP Weather Charts & Models
- Ship, Buoy Reports
- Rocketsonde
- Weather Balloons
- Storm Data
- Doppler Radar
- GOES, POES, NPOESS, and many other Satellites
- Aircraft Observations
- Wind Profiler
- Airport Weather Reports (ASOS)
- U.S. Climate Reference Network
- Climate Models
- Paleoclimate Data

NCDC Homepage/Website

- [www.ncdc.noaa.gov](http://www.ncdc.noaa.gov)
- New site design implemented in 2012 with continued enhancements since then
- Provides access to NCDC datasets, products, and services
- NCDC’s regional partners are also featured
- Data are accessed from disk (Storage Area Network) and tape (robotics system)
- Google Analytics used to provide usage statistics and patterns
- Drupal Content Management System provides the website content infrastructure
- Contact information is provided so that customers can call or email as needed
Monitor and Describe the Climate

- [http://www.ncdc.noaa.gov/climate-information](http://www.ncdc.noaa.gov/climate-information)
- Numerous monitoring products at the US and global levels
- Products related to extreme events, hurricanes, tornadoes, etc
- Temperature and precipitation data monitored in detail, regarding climate variability and change

Data Access Portals

**NOAA Climate.gov Portal, Drought Portal, Model Portal**

- Ongoing development and integrated to provide one-stop access to widely distributed datasets, products, services
- Drought Portal geared toward providing critical information to decision-makers
- Climate.gov Portal designed to reach a very wide segment of users — scientists, businesses, decision/policy-makers, news media, public, etc
- Model Portal provides access to reanalyses and numerical model output
- Many partners involved across NOAA, other agencies, and at the regional/state level
Data Access – International Partnerships
Global Observing Systems Information Center, World Data Centers for Meteorology and Paleoclimatology

- http://gosic.org/
- GOSIC Portal and the World Data Centers for Meteorology and Paleoclimatology are hosted by NCDC
- GOSIC Portal provides one-stop access to data and information identified by the Global Climate Observing System (GCOS), the Global Ocean Observing System (GOOS) and the Global Terrestrial Observing System (GTOS) and their partner programs
- The World Data Centers are a component of a global network of sub-centers that acquire, catalog, archive, and facilitate international exchange of scientific data without restriction

Data Access
Access to Model Data
Climate Forecast System Reanalyses (CFSR)

- 350 TBs of CFSR data now accessible online via NOMADS, along with other model data
- 60-80 TBs per month of model data downloaded
- Data volumes continue to grow as data are ingested from NCEP
- Access provided via OPeNDAP/HTTP/FTP/GridFTP
- Accommodates user’s most requested data by sub-setting long time series
- Fosters research within geo-science communities (ocean, weather, and climate)
Data Access
Climate Data Online (CDO) System

- Centralized access to numerous US and global datasets and products
- Web Services allow users direct machine-to-machine access for use in applications
- “Batch” process allows users to submit orders for data (e.g., by station, state, country, etc), then receive email with link to the data
- Underlying structure includes Oracle databases with tiered server infrastructure
- Services continue to be built-out for additional datasets and products

Data Access
Map/Web Services and Data Visualization

- GIS Map Services provide centralized access to numerous US and global datasets and products
- Web Services such as WMS, WFS, KML/KMZ (for Google Earth, etc)
- Access to the data and metadata, including machine to machine access
- Data visualization via tools such as Multigraph provide graphical display of various parameters
- Services continue to be built-out for additional datasets and products
CLASS Data Access

The CLASS Website

- CLASS Website
  - The CLASS Web Interface provides users with access to CLASS information holdings.
  - Displays the CLASS welcome page and help pages
  - Manages user login
  - Maintains each user’s contact information and preferences for searching and ordering
  - Receives users’ requests for information
  - Obtains requested information from other functions (e.g. browse images from Generate Associated Descriptions, result sets from the Data Management, metrics reports from Administration)
  - Maintains result sets and shopping carts for users
  - Forwards order specifications to the Process Orders function
  - HTML pages generate requested information, and return responses to users
  - Software components: Cocoon, Tomcat, HTTPD

www.class.noaa.gov
**Data Access**

**Weather and Climate Toolkit**

- Desktop application providing simple visualization and data export of weather and climate data
- Supports 22 data formats (Model, Satellite and Radar)
- Based on community developed tools and standards
- Data interoperability with diverse user communities
- Export data to GIS, KMZ, NetCDF and text formats

---

**Data Access**

**Interoperability with Google Earth**

- Weather and Climate Toolkit export of 3D Radar sweeps and isosurfaces for Google Earth visualization
NCDC offers a wide range of products and services. Our users range from large engineering firms designing the latest in safe energy efficient structures to the attorney documenting a weather event to the individual planning for a retirement move to universities and government agencies engaged in climate research. Services offered include data resource consultations, publications, copies of original records, certifications, and a wide range of online datasets, products, and reports. Services are delivered on a variety of media including online access, CD-ROM, DVD, computer tabulations, maps, and publications.

The NCDC 2014 Products and Services Guide provides a good overview of everything we have to offer. A free PDF copy of the guide is available on the NCDC website. Free hard copies of the guide are also available for order via the Online Store.
OCEANOGRAPHIC DATA MANAGEMENT

Presented to:
Environmental Disaster Data Management Meeting

Russ Beard
Director, National Coastal Data Development Center
Interim Science Coordinator, Gulf Restoration Council
September 16, 2014

NODC Stewards a Variety of Data

NODC manages the world’s largest collection of publicly available in situ and remotely sensed physical, chemical, and biological oceanographic data.
How is NODC’s Data being used?

Success of data management is judged by its usefulness to current and future users.

Key focus of the National Coastal Data Development Center

Comprehensive end-to-end data management for the coastal environment

- Gulf of Mexico Data Atlas
- Metadata development (semantic search and ontologies)
- Data discovery, mining, access, transport, archive, entry tools, and collaborative web tools
- Liaison Officers / Regional Approach
- Wide constituent base, customer service and user outreach
- Biological data considerations
- Geospatial enablement and visualization, e.g., ARC GIS and Google map
- Data integration, fusion and partnerships
Global Data Sets: satellite and in situ data

NODC AVHRR Pathfinder Version 5.2 SST Climate Data Record
Provides the longest (1981-2012), most accurate, and highest resolution consistently-reprocessed SST climate data record from the AVHRR sensor series.

World Ocean Database and World Ocean Atlas
Quality controlled comprehensive data collection and global in situ climatologies of temperature, salinity, dissolved oxygen, AOU, nutrients.

World Ocean Atlas 2013
World Ocean Atlas (WOA) is a set of objectively analyzed (1 degree grid) climatological fields at standard depth levels of in situ:

- Temperature
- Salinity
- Dissolved Oxygen
- Apparent Oxygen Utilization
- Percent Oxygen Saturation
- Phosphate
- Silicate
- Nitrate

It also includes associated statistical fields of observed oceanographic profile data interpolated to standard depth levels on 5°, 1°, and 0.25° grids.

The World Ocean Atlas 2013 (WOA) was created from the quality controlled data of the World Ocean Database 2013 (WOD)

www.nodc.noaa.gov/OCS/woa13/
The National Centers for Coastal Ocean Science (NCCOS) provides coastal managers the information and tools they need to balance society’s environmental, social, and economic goals. This geoportal provides an interface in which to discover and access the NCCOS data inventory.

**NOAA Gulf of Mexico Data Atlas**

*Digital Discovery & Access to Gulf Data*  
gulfatlas.noaa.gov

Over 235 map plates in 70 subject areas, from over 30 federal, state, non-governmental and academic partnerships.

Based on the traditional atlas format, the Gulf of Mexico Data Atlas is a data discovery and data access tool that allows a wide range of users to browse through a growing collection of datasets visualized as map plates. The goal of the Atlas is to provide access to datasets that characterize baseline conditions of Gulf of Mexico ecosystems in order to assist long-term research, monitoring, and restoration programs.

- Metadata, Web Mapping Services and Data Download and Access Links
- Access Representational State Transfer (REST) Services

- Additional Federal and State fisheries-independent species
- Federal fisheries-dependent catch for shrimp
- Hypoxia 10-year frequency of occurrence with animated annual-mean contour maps
NODC’s OceanNOMADS node supports NOAA research on marine ecosystems

- Data from operational, data-assimilating ocean models provides 4-D ocean state estimates
- Web tools simplify task of accessing model data in useful formats
- NODC staff working with NOAA and academic scientists on (FY 12-13):
  - Oceanographic input for whole-ecosystem models
  - Marine habitat models
  - Larval transport
  - Marine mammal ecology

QUESTIONS?
DIVER Data Warehouse and Query Tools: Samples & Contaminant Chemistry Focus

Environmental Disasters Data Workshop

September 16-17

Ben Shorr
NOAA’s Office of Response and Restoration

Overview

Discuss “Data Warehouse” approach
Actual framework and processes; flexible and scalable

Common Data Models
Overview of data models and standards; focus on samples and chemistry and related information/data

Data Query and Delivery
Requirements that drive development of data discovery, query, reporting and export tools
Data Warehouse and Business Intelligence

Data Warehousing concept and reality

- **Default** to existing tools and processes; databases and data sources with faults and inefficiencies
- The earlier field collected and lab processed data streams are integrated, the better connections and management. **UP FRONT EFFORT** pays big dividends
- **Data Warehouse and Data Vaulting** concepts
  - Ideally combine *data* beyond high level metadata
- **Business (Environmental) Intelligence**

Data Warehouse and Business Intelligence

- Use Industry standard tools
  - Collect and manage structured and unstructured information
- DWH Damage Assessment managing data with an Agile development approach
  - Evolve to meet data and development needs
  - Frequent brief video conference enhances accountability; minimizes silos; creates “team”
Data Warehouse and Business Intelligence

- Common Data Models – flexible and scalable
  - Core fields across datasets
  - Collect all digital possible (structured and unstructured) with key connections and hierarchy
  - ETL (Extract, Transform, Load)

Our Approach: Promoting Common Data Models
Common Data Model Examples (schemas)

- **Samples:** chemistry, biological+
- **Oceanographic:** cruise-collected sensor data
- **Observations:** shoreline, marsh, birds and mammals
- **Telemetry:** whales, dolphins, turtles, tuna
- **Photography:** keywords
- **Restoration data:** potential and implemented; budget and activities

---

Common Data Model Examples (schemas)

- **Core Fields**
  Higher level across data models e.g. Analysis Type, Data Source, Status, Spatial

- **Data Specific**
  Results, Methodology, Units

- **Related Information**
  raw data, field information, source data packages, unstructured documents (reports, graphs, charts etc...)
**Common Data Model: Contaminant Chemistry**

**Samples**
- Used existing data standard and data processing (Query Manager) for contaminant chemistry
- Electronic Data Deliverables (lab templates)
- Work with data providers (owners)
- Use existing standards and nomenclature; expand and standardize when necessary
- Metadata, metadata, metadata

**Common Data Model: Contaminant Chemistry**

- Ingest into Data Warehouse
  - Contaminant chemistry source databases include:
    - **Historical Contaminant Chemistry** (Query Manager)
    - **DWH Response** collected (EPA ETL → NOAA QA/QC)
    - **BP NRDA** provided
  - Audit source data and queries
  - Integrate with other data streams (e.g. additional field collection information, related field and lab documents and raw packages, “value added” analysis like oil source fingerprinting)
Common Data Model: Samples
(Includes Contaminant Chemistry)

Conceptual View

Location
State
Site
Lat / Long

Data/Time
Sample ID
Sample Date
Sample Time

Sample Attributes
Trip ID
Sample ID
Notes

Status
Analysis Status
Staging Status
Validation Status

Sample Results
Lab Name
Lab Results
Result Units

Species
Family
Genus
Species

Analysis
Analysis Category
Analysis Type
Analysis

External Links
Field Files
Lab Packages
QM/QC

Workgroup
Workgroup
Workplan
Study

...Just a few favorite fields
DIVER (Data Integration Visualization Exporting and Reporting): Explorer Tool

• Objectives & Requirements
  – Flexible query and export of all data including NRDA collected and external datasets
  – Documented lineage and connections to data holdings
  – Metadata, Metadata, Metadata
  – Export for analysis, visualization and processing

• Queries: Guided, Custom & Saved
• Download Data Packages
• Map & Legend
• Charts
• Data Tables
• Photos
• Metadata
• Study Notes
• Export
Guided Queries

Choose from Field List

Add Filters

Specify Date (or Depth Range)

Choose value(s)

Selected Value(s)
Mapping Requirements:

- Display geometry: points, lines, polygons
- Select on map
- Spatial selection/subset
- Symbolize by different aspects
- Interoperability (Show In ERMA*, save Shapefile/KML)

*NOAA's Environmental Response Management Application
Query Results (Table)

Requirements and Functionality:

- Present tabular results
- Integrated with map
  - Selected row highlighted in map
  - Select in map creates filtered table
- Link to source data files, related data and information (e.g. documents, photographs)
Charts

Requirements and Functionality:
• Provide overview summary of query results
• Interactive - click on charts to show filtered data
• Flexible - built to handle new information
Export

Requirements and Functionality:
• Spreadsheet and GIS formats (CSV, Shapefile, KML)
• Include metadata and related study notes (contaminant chemistry)
• Export results
• Export related data (additional fields and collection forms)
Metadata

Requirements and Functionality:
- “Lite” version with key information (HTML)
- FGDC compliant metadata (XML and HTML)
  – moving to ISO 19115
- Query Details: fields and values chosen
- Data Details: when were datasets updated?
- Data Caveats: notes about data
- Field Definitions
Validated public NRDA data available at these websites:
Next Steps & Challenges

- More widely available DIVER tools
  - Gulf of Mexico and Great Lakes
- Enhanced data search functionality
- Create flexible & scalable national approach
  - Ability to ingest digital field collected data and unstructured information

Internet Security

Discussion? Happy Hour?

Acknowledgements

DIVER Data Management Team: Mike Jackson, Dan Hudgens, Jim Anderton, Ann Jones, Amy Merten, Ben Shorr, Kevin Kirsch
NOAA: Amy Merten, Kevin Kirsch, Jay Coady
IEC: Kate Doiron, Ann Jones, Jess Fydenkevez, Lena Flannery, Neal Etre, Amy Anderton
Sirius Solutions: Vincent Luzzo, Nicole Williams, Brian Thompson
Ocean Sensor Data

Managing, Visualizing, and Understanding data using STOQS

16-17 September 2014
Environmental Disasters Data Management Workshop

Mike McCann
Monterey Bay Aquarium Research Institute
Oceanographic Observations
Long Range AUV

- Mobile platform measures properties while moving through the water
- Seabird CTD, Wetlabs ECO, ISUS, Optode, ESP, …
- T, S, optical backscatter, chlorophyll, fluorescence, DO, nitrate, genetics, …
- realtime and delayed mode
Workflow

1. Install STOQS from stoqs.googlecode.com
2. Conduct missions that collect data
3. Create CF-NetCDF 1.6 files of the data
4. Construct STOQS load script
5. Create PostgreSQL database and run script
6. Explore, visualize, and understand data
STOQS Architecture

All free and open source components

**Client**
- HTML5
- jQuery & AJAX
- Flot
- OpenLayers
- Twitter Bootstrap
- X3DOM

**Server**
- Python with pydap, numpy, ...
- Minnesota Mapserver
- GeoDjango ORM & web framework
- PostgreSQL + PostGIS
Relational Database
See video at stoqs.googlecode.com
Live Demo
Extra Slides
Tables to support machine learning
Labeled Data to support machine learning
DMS: Biological Data

Environmental Disasters Data Management Workshop (EDDM)
National Conservation Training Center (NCTC) USFWS, Shepherdstown, WV
16-17 September 2014

Felimon Gayanilo
Systems Architect
Harte Research Institute, Texas A&M University Corpus Christi
(email: fgayanilo@tamu.edu)
Overview

“…to quickly reach a science-based consensus about the defining characteristics and regulating processes of an ecosystem to address environmental disaster due to natural or anthropogenic causes using best available technology and data.”
Data Life Cycle

- Plan
- Collect/Generate
- Process/Analyze
- Archive
- Discovery/Re-use
Data Life Cycle

Data Aggregation/Collection/Archive Centers: Biological Data

**Type 1: Desktop/Stand-alone**
- Data is stored on off-line services

**Type 2: Short-term funded projects**
- Objective-based; data encoded offline and sometime served over the Internet

**Type 3: Institutional and Long-term initiatives**
- BCO-DMO; GRIIDC; DataONE; eBIRD
- Academic/Research Institutes; LTER; IEDA; ACADIS; fishbase.org; ecosystemresearch.org

**Type 4: Federal, regional and state programs**
- NOAA’s NODC/NGDC/NCDC/Fisheries/ESI/DIVERS;
- USDA; NIH NCBI; IOOS; GCOOS TOAST; Dept. of Ecology, WA; [http://www.ecy.wa.gov/database.html](http://www.ecy.wa.gov/database.html)

**Type 5: Multi-national programs**
- GBIF/Pangaea/CGIAR/UN Data/ World Bank
- Environmental Data
Processes: Most Common

<table>
<thead>
<tr>
<th>Planning</th>
<th>Collection/Generation</th>
<th>Processing</th>
<th>Archiving/Storing</th>
<th>Distribution/Discovery</th>
</tr>
</thead>
</table>
| - CDL
  - DMPTools
  - Online || Templates | - Objective-based (Forms) | - Objective-based | - Dataset |
|          |                       |            |                   | - Desktop              |
|          |                       |            |                   | - Institutional (DataONE) |
|          |                       |            |                   | - Programs (BCO-DMO, GRIIDC) |
|          |                       |            |                   | - Federal (NOAA, USDA, EPA, etc.) |
|          |                       |            |                   | - Multi-national (fishabse.org; GBIF; WB; FAO) |

CDL
DMPTools
Online || Templates
Objective-based (Forms)
Objective-based
Raw data
Processed data
Metadata (ISO/EML/CSDGM)
Publications
Common DMS Elements

User account: Data Provider
- Data Submission
- Data Review
- Metadata Review
- Data Posting
- Website (Search Engine)
- User account: Download

LDAP/CAS
- InCommon
- OpenID

{URL}
- HTTP
- FTP
- SFTP
- GridFTP

DOI
- CSW
- WAF
- ERDDAP

Feedback
- ISO/EML/CSDGM
• Insufficient information to establish data provenance (metadata) and data review (quality control)
• Failed to establish a common standard (collection, vocabulary, ontology and structure) throughout inhibiting the re-use/re-purposing of the datasets
• Insufficient interoperability and network capabilities
• Temporal and spatial limited (highly heterogeneous; goal setting)
DMS: Biological Data

Thank you!

Felimon Gayanilo
Systems Architect
Harte Research Institute, Texas A&M University Corpus Christi
(email: fgayanilo@tamu.edu)
Processes: Data Loss

ISSUES
- Interoperability
- Archival practice and accountability
- Insufficient feedback
Disaster Health Research Data Systems

Steve Ramsey, MPH

Overview

• Disaster epidemiology
• Types of human health data
• Data systems
Disaster Epidemiology

Objectives

• prevent or reduce the number of deaths, illnesses, and injuries caused by disasters
• provide timely and accurate health information for decision-makers
• improve prevention and mitigation strategies for future disasters by collecting information for future response preparation

Surveillance

• Mortality
  – Vital Records
  – CDC Disaster-related mortality surveillance form
• Morbidity
  – Laboratory
  – Sentinel sites
  – Syndromic surveillance
  – Absenteeism
  – Insurance
  – Pharmacy
  – Shelter
• Response
  – CASPER
  – OEMS Systems
Disaster Epidemiology

Research Data

• Registry/Cohort

• Short and long-term data
  – Medical history, occupational, recreational, residential exposures, mental health, social and behavioral factors

• Anthropometric and physiological measures
  – HT / WT, HC/ WC, HR/ BP, pulse ox, lung function

• Biospecimens
  – Blood, urine, toenails, hair, saliva for DNA

• Environmental measurements
  – Household dust and GPS coordinates

Cohort/Registry

Data Sources

• Training rosters
• Contractor lists
• Shelter manifests
• Evacuee manifests
• FEMA
• Social services
• Public datasets
• Local jurisdictions
Data Collection Tools
Research Management Systems

Basic Registry Information

- Contact Information
- Demographic and Sociological Factors
- General Health
- Deployment Information
- Exposure Information
- Medical Records
Biospecimen Collection Considerations

Disaster → Resource Constrained? → Time & space for specimen collection? → Self-collected Specimens
AND
Transport available to move specimens to CPL?

Resources Available? → Collect Convenience Specimens

Possible Convenience Specimens

Primary Specimens
- Serum and clot
- Plasma and PCV
- Whole blood or Lymphs
- Trace metals
- RNA, DNA studies
- Metabolic, endocrine, stress
- Metabolic, endocrine, stress, TM
- Metabolic, endocrine, stress
- Metabolic, endocrine, stress
- Endocrine, TM

Aliquots and possible assays
Disaster Epidemiology

Exposure Assessment

• Weather Data
• Monitors
• Sensors
• Models

Questions?
Instructions to Recorder: If conversation follows the questions and you can take notes below each question, please do so. If conversation does not closely follow questions you may take notes at the end, below all questions, under “General Discussion”. There is also a “Final Decisions/Conclusions/Summary” area at the end. Alternatively, if you have “final decision” items under each question you can highlight them in a certain color to designate. If appropriate, note who says things. There may not be group consensus – if so, please note the differences in opinion and give reasons/examples. At the end of the breakout session, please email me a copy of your notes (laura.belden@unh.edu). Even if they are very rough, it will just be a backup copy so you don’t lose your information in case of technology malfunction.

Recorder: Laura Belden

**Breakout Group A: Field Sample Collection (Data Collection/Sampling Protocols)**

- Is there a common data model that can be shared across entities?
  - No
  - Ideal data model characteristics:
    - Adaptive Management:
    - Possible models that can be adapted:
    - What data do we really want:
    - Creating a conceptual model that is unifying of data types etc would be good place to start

  Some common data models could be described conceptually (tomorrow).

- What are the essential core parameters to be collected and recorded for any field collection (e.g., sample ID, date/time, lat/long, etc.)?
  - Media, spatial, temporal
  - Time, date, sampling person, result, method, where to find more information,
  - Things that get at exposure and effects
  - Establishing nomenclature first is important

- What are the essential core parameters to be included in metadata record?
  - Unique identifier, data contact/custodian,
  - Protocol, SOP, strategy – why and how data were generated this way
• What are the standard data types and protocols for emergency response?
  Consider info that can be taken faster – proxy, representative, indicators
  Include in sampling design long term monitoring sites (certain percentage) then can get error bars etc.
  Sampling plan and protocol may differ from long term sampling one.
  We have lots of protocols for sampling particular things in particular matrix – need tiered one for emergency – need to develop before emergency because otherwise can take too long.
  Need to identify areas where baseline data doesn’t exist.
  
  o Shoreline and/or soils
    Notes
  
  o Watercolumn
    Notes
  
  o Air
    Notes
  
  o Human Health
    Notes
  
  o Other
    Notes

• What are best practices for reducing transcription errors?
  No 18 hours days for staff
  Ways can value check data
  Making sure it gets implemented can be a challenge
  Electronic field data entry reduces a lot of copying and transcription errors (small investment in this makes big difference). This makes it easier for next level reviewer to review too.
  Electronic tool will generate chain of custody (COC)
  Scribe system. Time stamping. Date stamping.
  GPS, camera, all included
  Automatic geocoding...

• What are the roadblocks for getting data from field collection into an electronic format?
  Electronic field data entry reduces a lot of copying and transcription errors (small investment in this makes big difference). This makes it easier for next level reviewer to review too.
  Electronic tool will generate chain of custody (COC)
  Scribe system. Time stamping. Date stamping.
  GPS, camera, all included
  Automatic geocoding...

  Challenge: internet for transmitting not always available (can transmit later).
  Electronic can generate labels (waterproof) (helps with “room” to write on label)
• How is field collection designed to maintain PII (personal identification, human health etc.)? Keep personal info on separate computer from actual data. Or don’t put PII on computer (in case lose computer).

• How is field collection designed to ensure accuracy of data?

• How is field collection designed to maintain Chain of Custody?

• How is field collection designed to maintain data security?

Notes

General Discussion:

If data is already being taken, how can we optimize it? (to be able to use environmental data for human health etc).

How do we get agencies to buy into this? Get them to collect baseline data etc. (to be discussed tomorrow)

Found a lot of errors in past disasters – transcription errors, missing AM/PM, missing data, etc. Should have been double checked upon entry.

Need to have SOP. Process for how train folks on SOP and documenting this. (But one challenge w/DWH was running out of enough trained people)

Plan to do it right for all disasters – big and small – do it right all the time, not just when White House watching.

Way to harmonize methods is to make it performance based (what is performance of your analytical method etc)

Would be helpful more trainings for NGOs and citizen science – training helpful if free/come to them because of limited resources

How do we start creating platforms that NGOs and citizen science can put their info into and others can start looking at it?

Phytoplankton Monitoring Network (used by NOAA) is model citizen science group.

Citizen Science could be front-line boots on ground where applicable.

Debris program uses a lot of citizen input.
Kent: items not sure addressed with questions here: quality control in emerg field sampling is often nonexistent or poor, developing improved sensor systems (b/c measurements can be cumbersome to collect data), interagency differences in how collect share use data (is National Response Team Science and Technology Committee the right place to bring these things to start address? Aubrey says conversations there more operational so not best place. Amy thinks this could work if we generate synthesis of workshop)

In between major events, work on developing technology, sensors, so that during next emergency can be that much more efficient and effective.

Sensors – have value/advantages, but less accurate than analytical method, define needed precision/accuracy – gets back to performance based. Sensor advantages include greater spatial/temporal data – this can be very valuable. Development of new sensors needed – work with industry. SMIR (small business innovative research programs) one place haven’t seen many budget cuts – look at this how can it better serve this All Hazards response – this is how you expedite technology.

We don’t have good way of bringing NGO and citizen data into govt system, assimilating it, and putting it back out.

Need data management systems flexible enough to incorporate new technologies.

Focus on what question want to answer, and how do I get that data (as opposed to shotgun approach of data collection – lots of data collected from everywhere). Smart sampling – predictive watershed models (Geoff, one effort doing currently). Ground truth things with local knowledge.

“Final Decisions”/Conclusions/Summary:
BREAKOUT GROUP DAY 2, RECORDER NOTES

Recorder: Laura Belden

**Breakout Group A: Field Sample Collection (Data Collection/Sampling Protocols)**

1) What are requirements for field data collection in order to assure good data?
   - Sampling Protocol
   - Trained data collectors, particularly related to emergency response (protocol for preparedness)
   - Coordination of sampling efforts
   - Performance based metrics
   - SOPs
   - Accurate and thorough metadata collection
   - Sampling and analysis plan specific to emergency response, anticipated scenarios

2) What are the types of media that should be sampled for an environmental disaster with respect to:

   **Human Health**
   Focus on exposure initially...
   Public, responders
   - Dermal
   - Time, location, and activity (changes by day)
   - Biological sampling (urine, blood, other human health information)
   - Mold, mildew

   **Ecological Health**

   **Both Human and Ecological**
   - Air
   - Soil/sediment
- Water
- Characterize Toxicity of hazard (what chemicals present, oil, dispersant, etc)
- Archival variety of samples that can analyzed with high sensitivity later (for other analytes aren’t known at time of incident) (done during background conditions too – can be expensive, if cost limited, plan to take these outside of disaster area during event)
- Biological sampling (urine, blood, other human health information, fish bile, )
- Leverage existing reference sites, (NEERS, NEON)
- Leverage existing citizen science and NGO networks

Note: USGS going to collect background data on east coast. Archive sediment samples. In event of hurricane etc, can go back and analyze select ones.
Appendix G
Breakout Group B: Data Formatting/Entry (for consistency and comparability)

- Is there a common data model that can be shared across entities?
  - Depends on who you are.

  Data Model: Field lists, definitions, values, and definition on these values + all key information from all data sets that will be used

  No, there isn’t anything across all disciplines. But there are many commonalities that different disciplines use. Best practices and models exist, but nothing universal. If we combine a bunch of things, we might be able to get a universal

  All data models are spatial. ISO191, Census Data, GIS are popular encompassing ones

  Data models can have similar structure, but within the models, there needs to be a glossary/index/dictionary that defines similar terms, variables, units, etc. and clarifies them for when you compare between models

  There doesn’t have to be an overarching model as long as their time and place standards.

  Weather: METAR,
  Contaminant chemistry: QUERY Manager Database
  Field observational contaminant chemistry: DIVER & photos
  ESRI GIS Models – Petroleum Engineering

  Adaptive Management
• What are the essential core parameters to be collected and recorded for any data collection (i.e., sample ID, date/time, lat/long, etc.)?

  **Summary:** Unique identifier, 4-D locations (time, X, Y, Z), parameter measured or observed, actual values, units + the metadata that goes with it

• What are the essential core parameters to be included in metadata record?
User restrictions. Information that is required for metadata standard. Core standards that go across all metadata record.

  Limits of detection by methodology

  Spatial reference/coordinate system (=coordinate system + datum)

  Also, how you got the data (Instruments have different levels of detection and different parameters), collection methodology

  Unique identifier also applies here

  What is this dataset, who collected it, review status (what type of quality control was done),

  essential core parameters (what was its purpose, spatial reference, who contributed to the data set, what **instruments** did you use?, shareability (proprietary, personal information [business identifiable], is it federal data? How can/cant this be used or shared?),

• What are the standard data types and protocols for emergency response?

  o Shoreline and/or soils
  Notes

  o Water column
  Notes

  o Air
  Notes

  o Human Health
  Notes

  o Other
  Notes
• What are best practices for reducing transcription errors?
  Naming conventions, terminology, domain definitions, safeguards in your system (so you can’t say a person is 16ft tall), use electronic data capture (see challenge at bottom), the original data collectors need to sign off saying that anyone else that edits the data has been approved, transcription verification/dual entry, how are you archiving the data capture?,

• What are the rate limiting steps for getting data from field collection into an electronic format?
  Time, inavailability of experts to integrate information, does the system have an offline mode, you can’t get trained workers (not hazardous waste trained), no access to internet, not able to read handwriting from paper documents and having to find the original data recorder to clarify, Difficulty to transfer from paper to digital, paper documentation requirements, Office & Management of Budget (any federal collection needs clearance by this – slow process), platform dependency (android vs. iOS, PC vs. Mac), running out of battery with electronic devices, lack of clarity for data transfer (no clear protocol or process established – different stakeholders and trying to adjust the data into a digital format for multiple stakeholders), data sharing issues/ownership issues, shifting culture that might only prefer paper documents into using electronic device, different versions of same software program or of documents “version control”. Non-standardized data (in the context of personal notes or a small sketch)

• How are data formatting/entry designed to maintain PII (personal identification, human health, SSN, birth date, etc.)?
  How much information is needed to identify a specific individual from a pool – different at each scale


  If it doesn’t need to be collected, don’t collect it. Only use it as needed. Use IDs

  You need enough data to make sure that you don’t survey the same person twice or make sure people with the same name get surveyed individually

  We don’t have to put PII information into the electronic record – we can keep it archived

  Encrypt data, enter the data twice to make sure it matches (computer will register the match to make sure there is no spelling error)

• How are data formatting/entry designed to maintain Chain of Custody?
  Referring to the transfer of documents from one person to another, what they did with the documents, how long they had it, etc.

  Electronic System Auditing
  Automatic system clearing to delete personal information, but backs it up in a secure location

• How are data formatting/entry designed to maintain data security?
Sharing status – information is appropriate for sharing with whom. – is there a part of that data that is not allowed to be shared yet before it is released to the public? Needs to be approved before it is shared. But, once it is shared, it still has to be protected. Safety and protection from collection to archiving

General Discussion:
Collecting Raw Data – Information – Knowledge (we are operating in the data, records, data set, collection of data points, etc. that we know a lot about). Even a raw data collection is entered in Excel and is data, but is before QAQC process. Our ideas are a bit more focused quality information. Essentially, the atomic bits that create our information is within these parameters.

We are looking at data that is collected but has not been put into a database and QAed, but is quite far from being discoverable in a database.

Perspective on data/information is the determiner. Data might be considered information to other groups, etc. Instruments that give data can still give QAQC data.

CHALLENGE: is equipment intrinsically safe and usable? Damaged in transport, can’t use in cold weather, need for backup systems

“Final Answers”/Conclusions/Summary:
<table>
<thead>
<tr>
<th>Issues and Challenges</th>
<th>Group B: In regards to Data Formatting/Entry:</th>
<th>Path Forward?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty in addressing issues</td>
<td>Importance (High/Medium/Low)</td>
<td>Each subgroup [like oceanography] needs to come up with its own common data model. Have workshops with those groups and have them develop a common data model.</td>
</tr>
<tr>
<td>Common Data Model(s)</td>
<td></td>
<td>Create pre-defined forms (like a drop-down menu and you have to choose something from the menu)</td>
</tr>
<tr>
<td>Common Language (Controlled vocabulary)</td>
<td>Red</td>
<td>Making sure you have data and field practitioners in the creation of the data model and end user verification/testing. Create a hypothetical natural disaster and use data from a bunch of different organization to see where the inconsistencies any full-scale drills to see where our resources are needed and how successful the data is used. “The White House is looking for X data in a few hours, see if you can make it work.” Drills will be a good time to make sure all software is up-to-date and people know how to use their equipment/forms. Have a calendar/check ins to see how progress is being made. Integrate different organizations to keep everyone informed of how data is being used by other organizations (a continued effort between pulses of drills or spikes in data use). As Conferences, add another day for each organization to talk about and explain their data + frequent virtual meetings to regularly check in and talk openly. Have a champion for every group that says what they do, the frequency of how they meet, etc. and have a representative from each group that can be held accountable for not following through with these ideas and letting people know the progress. Everyone agrees it is important but is not sure how to fix this.</td>
</tr>
<tr>
<td>Data structure</td>
<td>Yellow</td>
<td></td>
</tr>
<tr>
<td>Extensibility &amp; Usability</td>
<td>Red</td>
<td></td>
</tr>
<tr>
<td>Data Sharing &amp; Ownership</td>
<td>Red</td>
<td></td>
</tr>
<tr>
<td>Confusion between what is metadata? Is it international standards or just a scientist opinion of what metadata is</td>
<td>Yellow</td>
<td>What is reasonable? [we don’t have a perfect answer for this,] transformation tools from machine generated and standard metadata to standard metadata. One-page clear guidance [explain to general audience what standards ‘we’ are accepting]. Make sure metadata gets filled out completely whenever you ask to provide, make sure it is provided by the person collecting the data.</td>
</tr>
<tr>
<td>Core Parameters recorded during data collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>unique identifier quality: not unique, length, complexity</td>
<td>Green</td>
<td>Tracking replace lengthy (IDs, Meaning)/logical/sequential (IDs so you can tell if something has gone wrong [alphanumeric order]). Forms don’t have all proper IDs based on them. Link campaign to photo or other data so you can easily distinguish what data it is a part of.</td>
</tr>
<tr>
<td>4-D locations quality</td>
<td>Green</td>
<td>Agreement on time zone and reference time. Training with electronic devices as they change across time zones (knowing how to deal with that). Agreement on encoding of time. Standardization and training on coordinate system, precision &amp; accuracy, significant figures. Standard operating procedure (take a point with GPS and take a picture as double check). Calibrate equipment to ensure accuracy. Avoid redundant copying. Report inconsistencies immediately in order to clear confusion and find the correct answer.</td>
</tr>
<tr>
<td>actual location units</td>
<td>Green</td>
<td>What was the method?</td>
</tr>
<tr>
<td>metadata</td>
<td>Yellow</td>
<td>Calibrating equipment, agreement of what a flag value is (for example, if you have missing data, it is 0, blank, -1, etc.). Significant figures. Sanity checks to make sure the data makes sense” in the big picture</td>
</tr>
<tr>
<td>Core Parameters for metadata</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confusion between what is metadata? Is it international standards or just a scientist opinion of what metadata is</td>
<td>Yellow</td>
<td>What is reasonable? [we don’t have a perfect answer for this,] transformation tools from machine generated and standard metadata to standard metadata. One-page clear guidance [explain to general audience what standards ‘we’ are accepting]. Make sure metadata gets filled out completely whenever you ask to provide, make sure it is provided by the person collecting the data.</td>
</tr>
<tr>
<td>Reducing Transcription Errors during data formatting/entry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing data</td>
<td>Green</td>
<td>Input Validation (controlling the fields that they can enter data into in both paper and electronic forms), elimination of free text</td>
</tr>
<tr>
<td>Invalud Data</td>
<td>Green</td>
<td>Different inputting techniques (null vs. 0)</td>
</tr>
<tr>
<td>Missing Data (parameters that don’t support each other - If you are a make you can’t be pregnant)</td>
<td>Yellow</td>
<td>Logs on the forms to track consistency between fields.</td>
</tr>
<tr>
<td>Types of Inversions</td>
<td>Green</td>
<td>have two people enter same data and cross validate, transcription verify, have someone check, collect in electronic format, Field Lead or Originator verifies data after entry.</td>
</tr>
<tr>
<td>Illegible Data</td>
<td>Green</td>
<td>Have selectable drop down boxes</td>
</tr>
<tr>
<td>Unreadable Data</td>
<td>Green</td>
<td>Communication, training [understand the forms], make sure your gold standard has rules and defined goals. Incorporate pre-flight checks and pre-departure meetings to make sure everyone is up-to-date on data, refer to meetings. Governance or a Project Lead will help in order to achieve this. Standardize methodologies that are communicated to the team. Design a routine to eliminate errors from doing things out of order (“a standard habit”). Have a “morning assembly” to make sure people know what is going on and how to use forms.</td>
</tr>
<tr>
<td>Getting field data into electronic formats</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource limitations: Equipment &amp; People (analysis taken time)</td>
<td>Yellow</td>
<td>Get more equipment [money] or higher other labs that have more people or equipment (but going to other people can create inconsistencies)</td>
</tr>
<tr>
<td>Time delay between collection and processing, and then loss of information needed for a complete record</td>
<td>Green</td>
<td>Reduce time delay - collect and process data electronically compatible paper. If you are going to create a gold standard, have a time expectation for when you expect the data to be returned (build timelines into the plan). Delete the paper step from the plan so you are only using electronics. Look at systems that upload information to a cloud instantly. Consider data security</td>
</tr>
<tr>
<td>Inconsistency in questionnaires/surveys so you can compare groups</td>
<td>Green</td>
<td>Have a list of available questionnaires pertaining to certain topics</td>
</tr>
<tr>
<td>Inoperability in certain environments (constrained on use of certain technologies in certain environmental conditions) (i.e. a hazardous area like explosive, rain, flood, dry, etc. you can’t always use certain technologies)</td>
<td>Red</td>
<td>Ensuring that these limitations are considered when choosing solutions</td>
</tr>
<tr>
<td>Operating Equipment in certain environments (constrained on use of certain technologies in certain environmental conditions) (i.e. a hazardous area like explosive, rain, flood, dry, etc. you can’t always use certain technologies)</td>
<td>Green</td>
<td>Standardize data entry by calibrating team to make sure they understand the forms and variable being tested = do actual field exercises with them to practice - Designate a data manager in the team</td>
</tr>
<tr>
<td>Untrained team that have different focuses</td>
<td>Green</td>
<td></td>
</tr>
<tr>
<td>Maintaining FF</td>
<td>Refer all to Maintaining Data Security</td>
<td></td>
</tr>
<tr>
<td>Maintaining FF Chain of Custody</td>
<td>Refer all to Maintaining Data Security</td>
<td></td>
</tr>
<tr>
<td>Maintaining Data Security</td>
<td>Refer all to Maintaining Data Security</td>
<td></td>
</tr>
<tr>
<td>Functionality for user authentication on actual mobile devices</td>
<td>Red</td>
<td>Commercial industry way of addressing the technology challenge to meet that requirement mandated by federal government</td>
</tr>
<tr>
<td>Inoperability for application within the device (digital signatures)</td>
<td>Red</td>
<td>Commercial industry way of addressing the technology challenge to meet that requirement</td>
</tr>
<tr>
<td>something that happens for security adds friction in the field</td>
<td>Yellow</td>
<td>Have field practitioners involved in your decision making, adapt the minimally sufficient security requirements + reduce the footprint of security to as small as possible. Plan test explanation of what is required and how to meet that requirement.</td>
</tr>
</tbody>
</table>
Instructions to Recorder: If conversation follows the questions and you can take notes below each question, please do so. If conversation does not closely follow questions you may take notes at the end, below all questions, under “General Discussion”. There is also a “Final Decisions/Conclusions/Summary” area at the end. Alternatively, if you have “final decision” items under each question you can highlight them in a certain color to designate. If appropriate, note who says things. There may not be group consensus – if so, please note the differences in opinion and give reasons/examples. At the end of the breakout session, please email me a copy of your notes (laura.belden@unh.edu). Even if they are very rough, it will just be a backup copy so you don’t lose your information in case of technology malfunction.

Recorder: Stefanie Tetreault

Breakout Group C: Data Reliability/Tracking (accurate transmission to database & QA/QC, data validation)

- Is there a common data model that can be shared across entities?
  - Data is so different; could be a common format, metadata standard
    - Communities of practice
    - Generate flexible and extensible usage of existing standards (model(s) plural)
    - As data is processed and validated – multiple levels of QA/QC and validation
    - Day-to-day models vs incident models – overlap?
    - Common model yet to be built; hazard-based model
    - Need agreement and active relationships for data before incidents
    - Certain set of data points that need to be common – space, time, who collected it,
    - Metadata standard is somewhat of a data model
      - But most is collected/organized after-the-fact
    - QA/QC and metadata come in different levels – to different people at different points in the process
    - Metadata, Harmonized data, Specific data

- What are the essential core parameters needed for tracking the reliability of data?
  - Reviewed? By whom? – verify the source
  - Not necessary to go across disciplines – use generic parameters to capture enough for a specific discipline to use
    - A process that is known by all and followed; incident planning process
  - Reliability to access and validity of data
  - By discipline, by _, by source
Transcription verification; Subject matter expert validation;
Removal from raw data, transcription verification
“Authoritative source” is a big challenge – how do you verify
Peer review – initiated by the public, questioning
Time scale – incident scenario – not going to get peer review; provisional data

- What are the system requirements for data reliability and tracking?
  Metadata
  Across platform, flexibility
  Cannot be tied; central location
  Principles required?
  Accessing data through web services, loosely coupled – allows for federated databases
  IT issues, security – data backup, archive and maintaining original
  Network communication, stability of the comms
  Always going to have a hybrid data system – paper and electronic; issue is the dynamic of the system
  Validate that the data sent is the data received
  Not thinking too big, take a modular approach and scale up
  Being able to query
  Design methodology to bring all data together
  Each domain of knowledge has different rules for QA/QC
  Need a checkbox “has QA/QC been performed?”
  Data validation and reliability is the responsibility of the owner

- How are data reliability/tracking designed to maintain data security?
  Checksums
  Not necessarily concerned with the data but that it is not tampered with
  What is sent vs what is received
  Encryptions in transit and at rest
  Crowd-source data
    Let people develop a reputation of reliability
  Certify archived data – maintain integrity of original copy
  Version control; version information for devices that are collecting and processing data

- What are the QA/QC processes used and are they community and/or scientifically accepted standards?
  Peer review not practical at incident, third party validation
  QA/QC at high level are basic, at domain specific
  Work groups; producing derivatives of the data
  Record the source of the data
  Can’t go back after collecting original source data

- What is important for data reliability, QA/QC and validation when moving data from field collection into an electronic format?
Scanning original source data to store alongside electronic data file; transcription verification

Archiving material – holding times to consider

Physical sample to archive

Physical object and electronic object to be tracked together along with characteristics of them (including disposal, location, sample id, sample expiration date, info to allow sample to identified)

Data collection for some final informational product is product-specific

Fed/state/local response plans in place

Robust, flexible, system and processes to move data from field to electronic form

Nomenclature – a challenge to proper interpretation; need a common vocabulary

Overall standardization – vocabulary, units,

- What is the process for informing data generators/users about the status of data from collection to archives?
  Should have point of contact for feedback
  Posted production schedule
  Push-pull
  Define the user; how much access allowed
  Having provisional pathway built in to data flow –
    Status on push-pull basis
    Versions
  **If not done well, may be viewed as not transparent**
  Not including information because it might be misinterpreted
  Need subject matter expert –
  A system to keep generators and users engaged/informed on where the data is in the process
  “Generators and users” doesn’t necessarily capture everyone – information at a granular enough level to be able to communicate where things are in the process; and be able to track it
  Require a data source and a contact mechanism; whoever receives the data is now an “informer”
  To who is the data meaningful for?
  If you want to provide data, must provide contact – chain of custody

- What are the software and techniques for tracking disparate data sets for structured and unstructured data; where are they in process (at what lab, have they been analyzed? Have they been validated?)

Notes

- Optional: What are the standard data types and protocols for emergency response?

Notes

  - Shoreline and/or soils

Notes
General Discussion:

Crowd sourcing – sometimes it works

“Final Answers”/Conclusions/Summary:
<table>
<thead>
<tr>
<th>Issues and Challenges</th>
<th>Difficulty in addressing issues</th>
<th>Priority in Addressing</th>
<th>Path Forward?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Common Data Model(s) &amp; Core Parameters for tracking reliability of data (combined)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>defining metadata standards</td>
<td>Green</td>
<td>High</td>
<td>clarify the concept to enable a coalition to develop a project-based approach; leveraging existing systems and how they can be adapted; design an easy-reading training/internal outreach strategy</td>
</tr>
<tr>
<td>adopting metadata standards</td>
<td>Red</td>
<td>High</td>
<td>engage NOAA and metadata subject matter experts to establish a training plan/path forward</td>
</tr>
<tr>
<td>implement metadata standards</td>
<td>Red</td>
<td>High</td>
<td>see above</td>
</tr>
<tr>
<td>version control</td>
<td>Green</td>
<td>Med</td>
<td></td>
</tr>
<tr>
<td>cryptographic signature</td>
<td>Green</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>checksum</td>
<td>Green</td>
<td>Med</td>
<td></td>
</tr>
<tr>
<td>building comprehensive QC plan (validation levels, usability, methodologies, versioning, links to publications, historical and baseline data, links to source, study plan, QAP)</td>
<td>Red</td>
<td>High</td>
<td>scan, analyze, adapt/adopt; review existing large-scale plans</td>
</tr>
<tr>
<td>System Requirements</td>
<td>Support for implementing the common data model/higher level plan</td>
<td>yellow</td>
<td>high</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------</td>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>each domain of knowledge has different rules for QA/QC</td>
<td>across platform flexibility &amp; within-plan flexibility for emergency response</td>
<td>red</td>
<td>high</td>
</tr>
<tr>
<td>robust volunteer (non-federal, individual, organization) coordination plan</td>
<td></td>
<td>green</td>
<td>med</td>
</tr>
<tr>
<td>Interaction with Community (citizen watch, crowd sourcing, overarching social media)</td>
<td>Yellow</td>
<td>High</td>
<td>Tie into education and outreach; establish outreach plan; develop sub-template of requirements to ask of these people (informal 'metadata'); flexibility in platform (twitter, etc);</td>
</tr>
</tbody>
</table>

**Maintaining Data Security**

| Defining data security - what is necessary (checksums, GPG signatures, chain of custody) | Yellow | High | Policy recommendation; creating a plan; having a panel of experts from different domains to verify/validate protocols to validate authenticity; |

| At rest and in transit encryption (organizational agreements) | Green | Medium | See above |

| Adopting security plan | Red | High | See above |

| Defining who should have access, levels of access (system level, local admin rights, not requiring an IT person in the field) | Red | High | See above |

<p>| Defining organizational vs common levels of data security | Red | High | See above |</p>
<table>
<thead>
<tr>
<th>Developing community and scientifically accepted standard QA/QC processes</th>
<th>Need for a coalition of government, public, scientific, academia, stakeholders</th>
<th>red</th>
<th>high</th>
<th>Identify, organize, and deal with the low-hanging fruit; implement the plans noted above</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data reliability, QA/QC, and validation when moving data from field to electronic format</td>
<td>Scanning original source data to store alongside electronic data file; transcription verification and validation</td>
<td>green</td>
<td>high</td>
<td>Developing best practices for capturing and submitting data types; supply tools and training to enable field personnel</td>
</tr>
<tr>
<td>Archiving material, physical samples – holding times to consider</td>
<td>Physical objects and electronic object to be tracked together along with characteristics of them (including disposal, location, sample id, sample expiration date, info to allow sample to identified)</td>
<td>red</td>
<td>high</td>
<td>Consider early on, how long to be kept and reasonable limitations</td>
</tr>
<tr>
<td>Physical objects and electronic object to be tracked together along with characteristics of them (including disposal, location, sample id, sample expiration date, info to allow sample to identified)</td>
<td>Robust, flexible, system and processes to move data from field to electronic form</td>
<td>yellow</td>
<td>high</td>
<td>Determine importance of sample to set time to be kept; identify potential for legal ramifications</td>
</tr>
<tr>
<td>Robust, flexible, system and processes to move data from field to electronic form</td>
<td></td>
<td>red</td>
<td>high</td>
<td>Very important for QA/QC, see group A</td>
</tr>
<tr>
<td>Nomenclature – a challenge to proper interpretation; need a common vocabulary</td>
<td>yellow</td>
<td></td>
<td>identify appropriate standards; invite experts to develop vocabulary across domains</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Informing data generators/users about the status of their data &amp; Tracking Disparate Data Sets as they are processed</td>
<td>designing and implementing flexible infrastructure to provide multiple types of access; clearly defined roles and responsibilities; Should have point of contact for feedback from data providers</td>
<td>red</td>
<td>high</td>
<td></td>
</tr>
<tr>
<td>Having provisional pathway built into data flow –</td>
<td></td>
<td></td>
<td>The following points relate to informing generators/users status of the data:</td>
<td></td>
</tr>
<tr>
<td>Status on push-pull basis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Versions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If not done well, may be viewed as not transparent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not including information because it might be misinterpreted</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Need subject matter expert –</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A system to keep generators and users engaged/informed on where the data is in the process</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“Generators and users” doesn’t necessarily capture everyone – information at a granular enough level to be able to communicate where things are in the process; and be able to track it

Require a data source and a contact mechanism; whoever receives the data is now an “informer”

To who is the data meaningful for?

If you want to provide data, must provide contact – chain of custody

Preliminary Conversation:
short term vs long term priorities
the hardest problems yeild greatest gains

system of preparedness - science/academia information management; data and sample management plans
what makes data valid and useful - bringing this into the preparedness plan

QA/QC needs to come from the source of the information; and be
Appendix I
Instructions to Recorder: If conversation follows the questions and you can take notes below each question, please do so. If conversation does not closely follow questions you may take notes at the end, below all questions, under “General Discussion”. There is also a “Final Decisions/Conclusions/Summary” area at the end. Alternatively, if you have “final decision” items under each question you can highlight them in a certain color to designate. If appropriate, note who says things. There may not be group consensus – if so, please note the differences in opinion and give reasons/examples. At the end of the breakout session, please email me a copy of your notes (laura.belden@unh.edu). Even if they are very rough, it will just be a backup copy so you don’t lose your information in case of technology malfunction.

Leader: Mark Miller  
Recorder: Angela Sallis  
Reporter: Mike McCann

Introduction
Nobody is talking about a single place to store data. We have a giant continuum: still involves huge temporal time scale. Decision makers need data immediately, but others have long time frame. Let’s not confuse the time scales. Data management has to meet all the needs. Intersecting these requirements: focus on the right problems. Process, hardware, whole range. May be useful to put in the context of use case(s).

Breakout Group D: Discovery and Accessibility (getting data to the users)
- Is there a common data model that can be shared across entities?
  - Definition: a data model documents and organizes data, defines how it is stored and accessed, and establishes the relationships among different types of structured and non-structured data. Data modeling techniques and methodologies are used to manage data in a standard, consistent, predictable manner in order to manage data as a resource.
  - We do not have a common data model now.
  - Multiple parties that come with their own data model to the response and how do they come to have an understanding of, accepts and using the existing administration policies (open data policy) or common data model.
  - Agreement that whatever you bring to the table works with everything else (platform independent)
  - How do we have an ontology for all the different players?
  - Need a data sharing agreement that establishes things like an ontology.
- This includes using machine-readable and open formats, data standards, and common core and extensible metadata for all new information creation and collection efforts.

- **What are the essential core parameters needed for discovery and accessibility?**
  - Security is important consideration in maintaining the quality of the data as well as the accessibility. Can still be discoverable, even if it isn’t accessible, for transparency. See Open Data Policy.
  - How do we make sure that all archives are secure, like universities? This impacts response requirements of 24/7 accessibility. Open Data Policy has access level requirements.
  - Security has other meanings.
  - Important to define why data might be classified inaccessible.
  - How do systems access each other on a system to system basis?
  - How do I find the databases I need? How do I apply for access? How do I download data?
  - How do I search the data? Keywords, ontologies, classifications between multiple ontologies. Including common misspellings. Has to be defined ahead of time.
  - Difference between ontology and vocabulary. Ontology is classification, vocabulary is definition.
  - Ontology can be used to show links between concepts—shrimp to chemistry, DO, boats
  - Use Case: Multiple datasets collected by different groups that have to be linked together in order to answer a specific environmental question. DOI

- **What are the system requirements for discovery and accessibility?**
  - Vocabulary/ontology—should be built into system by software i.e., user-centered design.
  - There are certain necessary parameters that should go with every sample. Need to use vocabulary.
  - Online access
  - Public accessible
  - Platform independent
  - Accessibility controls
  - Metadata—hopefully automatically generated as data is collected
  - System has to be dynamic modified for access
  - See OER Data Management model-Rolling Deck to Repository (R2R) model
  - Need valid links to metadata, data, contacts
  - Robust infrastructure to host during emergency situation (lots of bandwidth)

- **What are the best practices for data visualization, discovery, and accessibility?**
  - Broad spectrum of users defines a very complex for data visualization, discovery, and accessibility.
  - We have lots of best practices, some are white papers that haven’t been implemented. Should develop an inventory that can be shared. Identify benchmarking best practices by going to the experts.
  - How do I find a best practice? How do I implement it?
  - Want to be careful when you call something a best practice for lawsuit reasons. Maybe community accepted standards or SOP. Still could be problems with these terms.
  - Have to know what questions you are trying to answer-back to user centered design.
  - Create apps that enable people to capture data
o Need quality statement to go along with data so you can tell how you can use the data.
o Computer mining like on Star Trek- where is that data system that can help you make a
decision. Machine read the data and give me the answer.
o Best practice: Collect it electronically.
o Do best practices have common elements that can be cross walked? Need training and
awareness. Don’t know what you don’t know. Important for planning.
o Where does it go after you collect it?

• What are the best practices for maintaining PII (personal identification, human health, SSN, birth
date, etc.) and Chain of Custody in discovery and accessibility? Human subjects data
protections?
o Follow guidance of OSTP. Best practices of metadata, e.g., instead of name, use a
position title.
o NCDDC documented best practice for chain of custody during DWH – wiring diagram.
o Best practices should be shared.

• How is access to data granted to users given that PII data are available and need to be
protected?
o If there is sensitive data in your data set, how do you decide who has access to it?
o Authoritative data collector can decide.
o Need to see OSTP guidance-need training.
o Get into problems when you use secondary source data-how does that work?
o What about when parts of the data are protected but parts aren’t?
o Interpreting PII to be any controlled data, e.g. marine archeology.
o There is also business data like budgeting data that is under data control (restrictions)
that can’t be accessed even under FOIA. Should make a single list that could be shared
and put in metadata records.

General Discussion:

• Where does analysis fit into the picture?
• How do we link in those things? Whose responsibility is it to link the breadcrumbs back to the
original data? In metadata, it is the responsibility of the product creator to link back to the
original data. Digital Object Identifiers (DOI) can help with this but need more awareness and
training.

“Final Answers”/Conclusions/Summary:

• We don’t have a common data model, and we may never have one, but we need the
ability for the multiple ones to work together (interoperability).
• Discovering the fact that the database exists, then being able navigate the database.
• Keywords, spatial, temporal
• User-centered design
• Good metadata, good metadata training, staff and support
• Should develop an inventory of best practices for data visualization, discovery, and accessibility that can be shared
## Group D: In regards to Discovery and Accessibility:

<table>
<thead>
<tr>
<th>Issues and Challenges</th>
<th>Difficulty in addressing issues</th>
<th>Priority in Addressing Path Forward?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Funding and Staffing (problem across all questions)</td>
<td>R</td>
<td>High</td>
</tr>
</tbody>
</table>

- **Evangelizing**, putting emphasis on data management, bad data means bad decisions that can kill people. Scope the scale of the problem. Write data management into project plans (10-15%). Good data management can save money in the long run. Academic and/or industry investigation of how good data management can save money.

- **Interoperability** between the models through training, awareness, consistency of the existing systems and core elements. Required by Open Data Policy for federal entities to move in this direction.

<table>
<thead>
<tr>
<th>Common Data Model(s)</th>
<th>Y</th>
<th>High (Essential)</th>
</tr>
</thead>
</table>

- Nine core elements plus nine if-applicable elements from Open Data Policy. See http://project-open-data.github.io/schema/. See Common Core elements. Make this information more commonly known through evangelizing, training, publications.

<table>
<thead>
<tr>
<th>Core Parameters for discovery and accessibility</th>
<th>G</th>
<th>High</th>
</tr>
</thead>
</table>

- Limited awareness of core parameters/elements
Limited implementation of core parameters/elements | Y | High

Acknowledge that many members of the public do not have access to the Internet | R | Low

Bandwidth during an incident to both upload and download data | Y | Medium

Infrastructure (hardware) exists for sharing data across entities | G-Technical, Y-Process, R-Security, G-Industry Internally | High

Storage and archiving the data long term so it can be accessible | G-storage; Y-Archive | High

Sharing process and policy information | G | High

Nine core elements plus nine if applicable from Open Data Policy. See http://project-open-data.github.io/schema/. See Common Core elements. Using this information modify existing and new systems. Metadata requirements are extensible.

Use of traditional media to provide information to public; don't rely on Internet/Social Media

Awareness that this is a major problem during an incident. Evaluate alternative communication models such as Wave relay to stand up local infrastructure; UAVs; mesh networks; other technology to address. Communicate what technology and permissions are needed to decision makers. Acknowledge that there is exponential growth in data.

Data sharing agreements and discussions before incidents.

Data centers already exist for archiving issues, but there are issues that exist that go beyond. Recognize that data centers are underfunded. register data with and make it known to use Data.gov and HAZUS.gov

Two pager from federal perspective to list/explain all policies affecting data access; shared broadly as possible.
Developing Best Practices for data visualization, discovery, and accessibility

<table>
<thead>
<tr>
<th>Description</th>
<th>Responsibility</th>
<th>Urgency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementation of policy R High</td>
<td>Information Officer when incident occurs to coordinate data accessibility R</td>
<td>High</td>
</tr>
<tr>
<td>Cultural shift for some agencies. Need decision makers to get behind. Need resources.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incident management handbooks have to be adjusted to include this which is a high level decision.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Lots of command structures in an incident. Need to share data for decision makers timely in common data products/presentations. Y |
Need a common system like DIVER or GII like Homeland Security's which can visualize many streams of data. Need awareness of all the different options. |

Need for data to be checked for quality assurance/sensitivity prior to accessibility during an incident prior to release to media/public Y |
Resource question. Need people who aren't generating data to check it. Joint Information Center if incident is big enough. |
Online metadata training is currently available. Different levels of metadata training for different roles. Figure out which entities need to take it. |
Need metadata training G | High |

Resource question.
| How do users know what words to search with? | Y | Medium | Users want it to work like Google, temporal, geospatial. Learn from people with diverse users. Develop smart user centric smart ware that link associated themes. Give user limited number to help guide user --pull down lists. |
| No awareness of existing ontologies, common vocabularies | G | High | Link existing ontologies; training |
| Implementation of keywords and ontologies for response data by the data generator | R | High | Find out what vocabulary industry uses; dealing with full range of data generators; |
| Multiple entities and agencies come to an incident with their own system. How do you connect those information streams? | Y | Medium | Can be done with communication but security may be a problem. Lead agencies in incidents need to come to an agreement about data sharing. |

Developing Best Practices for maintaining PII and controlled access data during discovery and accessibility

Lack of awareness of definition of PII and controlled access data

G | Medium | See Open Data Policy https://cio.gov/wp-content/uploads/downloads/2012/12/Standardized_Digital_Privacy_Controls.pdf Make this policy well known. Identify all the issues and policies that deal with controlled-access data |
<table>
<thead>
<tr>
<th>Implement best practice for developing access to PII and controlled-access data</th>
<th>High</th>
<th>Resources. Decision maker buy in.</th>
</tr>
</thead>
</table>

**Developing Best Practices for maintaining Chain of Custody during discovery and accessibility**

| Lack of Accountability and Ownership | Medium | Goal is to reach electronic submission. |
| Multiple process for chain of custody depending on collector | Medium | Understanding of litigation hold: General counsel define minimum requirements for litigation hold |

| Transparency of users knowing the data exists even if they can't get access to the actual data | High | Need for synthesis. Need to identify the different processes. |

**Granting users access to data while maintaining PII and controlled access data**

| Who decides which users get access? | Medium | Awareness of the Open Data Policy which gives policy guidance on this issue. Make users aware of why data is being restricted. |
| When request come in for multiple data sets, you don't always have enough information about the data and if it contains sensitive information | Low-In heat of incident as all is sensitive High-in long term | Responsibility falls upon authoritative source. It is the authoritative source's job to know the laws and policy. |

| Same as above. Authoritative source, open data policy, flagged in the metadata. | --- | --- |
Breakout Group D: Discovery and Accessibility (getting data to the users)

Definitions/Acknowledgements:

**Discovery:** User (Levels of priority during incident: Incident command, academia, then media/public) knowing that the data or information exists and then being able to find (by search or other method) the specific data desired. Dynamic continuum priority, priorities shift over time. Acknowledge that many members of the public do not have access to the Internet.

**Accessibility:** Includes how the user accesses the data (by browser, mobile app, or other?) and the level of access (completely public or with credential restrictions). Text messaging can be used after incidents when Internet is down understanding that can’t be preserved where appropriate. Bandwidth can be an issue. Acknowledge varied skill levels of users.

**Data Model:** documents and organizes data, defines how it is stored and accessed, and establishes the relationships among different types of structured and non-structured data. Data modeling techniques and methodologies are used to manage data in a standard, consistent, predictable manner in order to manage data as a resource.

Assumptions:

Large temporal scale – data access in minutes, days, years depending on user and needs

Large user scale – public, Unified Command, NGOs, researchers, media

No common or standard data model

All data system support platform independent, standard file format data exchange (should include standardized metadata). Data sharing agreements where possible, i.e., all parties are aware and agree that interoperability and data access among the entities (see the elements previous).

Access via browser, desk and mobile apps

Integrate security into access and qa/qc
  If user does not have credentials still discoverable but not accessible
  Explain why data/information is not accessible
  Understand security of all entities (Universities)

Don’t conflict with existing statutes, regulations, guidance
  Open Data Policy (OSTO) – metadata, define user access
Information Quality Act

Inventory existing Best Practices (visualization, discovery, accessibility)
See Ocean Exploration and Research (OER)

User Centered Design (UCD)-who is going to be using system, what questions are they trying to answer, how are they going to access

Storage is different from archive. Archive is version controlled, for posterity, and accessible.

**Next Steps:**

**Use Cases:**

1. Multiple datasets collected by different groups that have to be linked together in order to answer a specific environmental question. Digital Object Identifier (DOI) [www.datacite.org](http://www.datacite.org)
2. Amna- conceptual model
3. 

Define user ontology and data vocabulary

Would this be different for each user types?

Keywords, spatial, temporal - parameters

Use case from Amna:

See her diagram photo. Mark took a photo.

See Russ’s DWH Slide. He or I can email.
GROUP D: IN REGARDS TO DISCOVERY AND ACCESSIBILITY

Issue: How to integrate multiple systems with multiple formats with end users

Assumption
- All data system support platform independent, standard file format data exchange (should include standardized metadata). Data sharing agreements where possible, i.e., all parties are aware and agree that interoperability and data access among the entities.
Common Data Model(s)

- Many
- Interoperability between the models through training, awareness, consistency of the existing systems and core elements.
- Required by Open Data Policy for federal entities to move in this direction.

System Requirements for discovery and accessibility

- Limited awareness of core parameters/elements
- Nine core elements plus nine if-applicable elements from Open Data Policy. See http://project-open-data.github.io/schema/.
- See Common Core elements.
- Make this information more commonly known through evangelizing, training, publications.

System Requirements for discovery and accessibility

- Infrastructure (hardware) exists for sharing data across entities
- Sharing process and policy information
- Data sharing agreements and discussions before incidents.
- Two pager from federal perspective to list/explain all policies affecting data access
- Shared information broadly as possible.

System Requirements for discovery and accessibility

- Storage and archiving the data long term so it can be accessible
- Data centers already exist for archiving issues, but there are issues that exist that go beyond.
- Recognize that data centers are underfunded.
- Register data with and make it known to use Data.gov and HAZUS.gov
Developing Best Practices for data visualization, discovery, and accessibility

- Information Officer when incident occurs to coordinate data accessibility
- Incident management handbooks (IMH) need to be adjusted to include this, which is a high level decision.

- Need metadata training
- Online metadata training is currently available.
- Different levels of metadata training for different roles.
- Figure out which entities need to take training.

Developing Best Practices for maintaining Chain of Custody during discovery and accessibility

- Implementation of keywords and ontologies for response data by the data generator
- Find out what vocabulary industry (relevant domain) uses
- Integrate and communicate with full range of data generators

- Lack of Accountability and Ownership
- Goal is to reach electronic submission.
- Need understanding of litigation hold: General counsel define minimum requirements for litigation hold
Developing Best Practices for maintaining Chain of Custody during discovery and accessibility

- Multiple process for chain of custody depending on collector
- Need for synthesis of entity policies.
- Need to identify the different processes.

Granting users access to data while maintaining PII and controlled access data

- Transparency of users knowing the data exists even if they can't get access to the actual data
- Awareness of the Open Data Policy which gives policy guidance on this issue.
- Make users aware of why data is being restricted.

Granting users access to data while maintaining PII and controlled access data

- When request come in for multiple data sets, uploaders don't always have enough information about the data or know if it contains sensitive information
- Awareness of the Open Data Policy which gives policy guidance on this issue.
- Responsibility falls upon authoritative source. It is the authoritative source's job to know the laws and policy.
- Flagged in the metadata.