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.

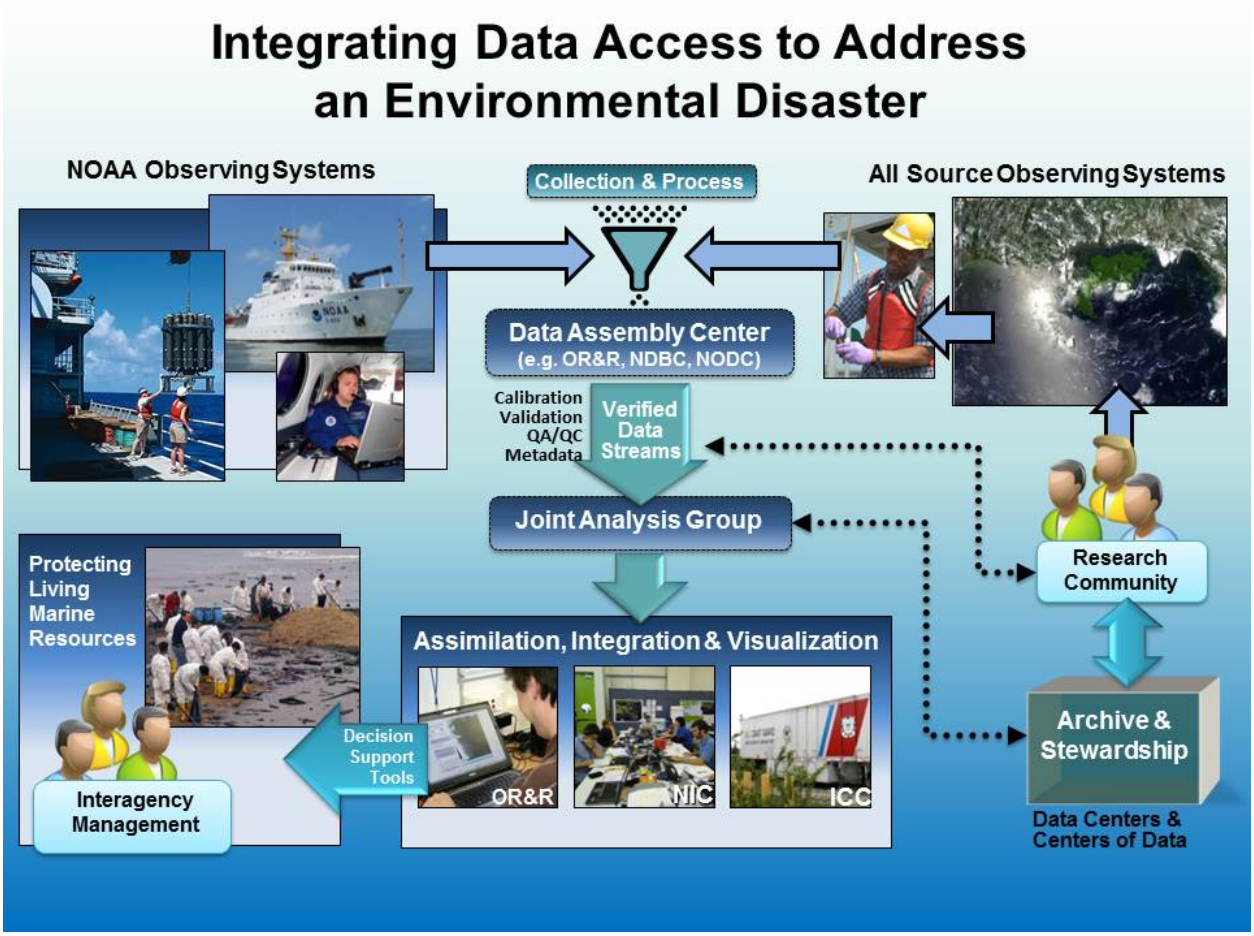

Figure 1. Courtesy of Russ Beard, NOAA, National Coastal Data Development Center

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 sequela. 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.

	Issues and Challenges	**Difficulty**	**Path Forward**

Common data model(s)	Flexibility to adapt	high	Develop interdisciplinary focus. Group/workshop to address.
Core parameters recorded during field collection	Protocols, training, quality assurance, best tools available	medium	Include in funding plan.
Core parameters for metadata	Integrations of citizen and NGO groups collecting core parameters. Using local knowledge/samplers.	medium	Include in preparedness planning. Preparedness and training in advance. Set expectations early about coordination and communication.
Reducing transcription errors	Completeness and accuracy difficult in field conditions	easy	Use electronic entry, when possible. Make it as easy as possible. Do not proceed without filling in all fields. Have automatic data field checks.
			Have appropriate review at appropriate times. Accountability. Real time quality control. Timely review.
			Identify and implement best practices (such as data intake team concept used by NOAA).
			Investigate automated processes, sensors, etc.
	Fatigue	medium	Enhance intake team capacity.
Getting field data into electronic formats	Location Resources: time, money, people	easy	Adopt existing software. Include in drills, plans, and funding.
Maintaining PII	Security, trust, safety, confidentiality	high	Continually upgrade system. Work with security experts. Understand and implement requirements. Follow existing prescribed security processes. Train recorders/collectors.
	Institutional Review Board (IRB) - slows process	high	Have IRB come up with plan for disasters. Blanket IRB that can be implemented during disasters, with pre-approval.
Ensuring accuracy of data	Appropriate QA/QC methods implemented in disasters	medium	Provide training, ensure preparedness
Maintaining chain of custody	Disaster field conditions complicate this	medium	Provide training and supplies, implement procedures
Maintaining data security	Loss or failure of electronic sampling equipment (data integrity), also see PII issues	medium	Implement redundant and robust systems, develop/use best practices for data backup, encryption, training
	Transmission security -integrity	high	Have appropriate systems, encryption
	Transmission security -confidentiality	medium	Have appropriate systems, encryption

3.2 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 on paper and finding the original data recorder to clarify,
- 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.

	Issues and Challenges	Difficulty	Priority	Path Forward
Common data model(s)	Common language (controlled vocabulary)	high	high	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.
	Data structure	medium	high	Create pre-defined forms (e.g., have key tracking terms like keys, ID). Constraint lists (drop down menu - must choose).
	Extensibility & useability	high	high	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.
	Data sharing & ownership	highest	high	Draft memoranda of data sharing agreements so they can be executed at time of disaster. (Group agrees important item, uncertain of best solutions)
Core parameters recorded during data collection	Unique identifier quality: not unique, lengthy, complex	easy	high	Use barcodes to replace long IDs. Use meaningful/logical/sequential IDs so know if something went wrong (alphanumerical order).
-	4-D locations quality	easy	high	Agreement on time zone/reference time and encoding of time. Standardization and training on coordinate system, precision & accuracy, significant figures, calibration, and crossing time zones. Standard operating procedures. Report inconsistencies immediately.
	Parameter measured or observed quality	easy	high	Document the method used.
	Actual values	easy	high	Calibrating equipment, agreement on flag values, significant figures. Checks to make sure the data 'make sense' in the big picture.
	Units	easy	high	Standardize and be explicit.

	Metadata	medium	high	Document instrument used. Zip metadata with data, so it is a core component.
Core parameters for metadata	Confusion regarding definition of metadata	medium	medium	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.
Reducing transcription errors during data formatting/entry	Missing data	easy	high	Validate input. Require all fields.
	Invalid data	easy	high	Inputting techniques (null vs. 0)
	Illogical data (e.g., a male can't be pregnant)	medium	high	Track consistency between fields.
	Typos or inversions	easy	high	Dual entry with cross validation, transcription verify, collect in electronic format, QC after entry (perhaps by field lead or originator)
	Illegible data	easy	high	Have selectable drop down boxes.
	Version control	medium	medium	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.
Getting field data into electronic formats	Resources limitations: equipment & people (analysis takes time)	medium	low	Get more equipment ($) or hire labs with more people and equipment (although can create inconsistencies). Prescreen and certify contractors.
	Time delay between collection and processing, and then loss of information that is needed for a complete record	easy	high	Gold standard with time requirements. Only use electronic. Systems that upload instantly to a cloud. Consider data security.
	Inconsistency in questionnaires, unable to compare groups	easy	high	Have questionnaires available digitally for download.
	Operating equipment in hazardous areas	high	high	Ensure limitations are considered.
	Untrained teams that have different focuses	easy	high	Standardize data entry - ensure team understands forms and variable tested. Field exercises for practice. Data manager accompanies team.

Maintaining data security, PII, and chain of custody	Functionality for user authentication on actual mobile device	high	high	Industry develops necessary technology.
	Inoperability for application within the device (digital signatures)	high	varies per situation	Industry develops necessary technology.
	Something that happens for security adds friction in the field	medium	high	Involve field practitioners in decisions. Minimize security impact. Explain what is required and how to meet it.

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.

	Issues and Challenges	Difficulty	Path Forward
Common data model(s) & core parameters for tracking reliability of data (combined)	Defining metadata standards	easy	Clarify the concept to enable a coalition to develop a project-based approach; leverage existing systems and how they can be adapted; design an easy-reading training/internal outreach strategy
	Adopting metadata standards	high	Engage NOAA and metadata experts to establish a training plan/path forward.
	Implementing metadata standards	high	See above
	Building comprehensive QC plan (validation levels, useability, methodologies, versioning, links to publications, historical and baseline data, links to source, study plan, QAP)	high	Scan, analyze, adapt/adopt; review existing large-scale plans
	Need agreement and active relationships for data before incidents	medium	See above
	Easy translation and communication to public - common language/public outreach on understanding data quality and importance of metadata	medium	Make this a priority and work with incident command structure; forms, job aids, info inserts for incident management handbook & work flows
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	Need for a coalition of government, public, scientific, academia, stakeholders	high	Identify, organize, and deal with the low-hanging fruit; implement plans noted above

Data reliability, QA/QC, and validation when moving data from field to electronic format	Scanning original source data to store alongside electronic data file; transcription verification and validation	easy	Develop best practices for capturing and submitting data types; supply tools and training to enable field personnel
	Physical and electronic objects to be tracked together along with their characteristics (e.g., disposal, location, ID, expiration date, sample identifying information)	medium	Determine importance of sample to set time to be kept; identify potential for legal ramifications
	Robust, flexible, system and processes to move data from field to electronic form	high	Very important for QA/QC, see group A
Informing data generators/users about the status of their data & tracking disparate data sets as they are processed	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.	high	Have provisional pathway built in to data flow
			Status on "push-pull" basis
			Need subject matter expert
			A system to keep generators and users engaged/informed on where the data is in the process.
			Information at a granular 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 provide data, must provide contact – chain of custody

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.

	Issues and Challenges	Difficulty	Priority	Path Forward
Common data model(s)	Many	medium	high (essential)	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.
Core parameters for discovery and accessibility	Limited awareness of core parameters/elements	easy	high	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.
System requirements for discovery and accessibility	Infrastructure (hardware) exists for sharing data across entities.	easy - technical, industry, internally medium - process high - security	high	Develop data sharing agreements and have discussions before incidents.
	Storage and archiving the data long term so it can be accessible	easy - storage medium - archive	high	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.
	Sharing process and policy information	easy	high	Develop a two pager from federal perspective to list/explain all policies affecting data access; share broadly.
Developing best practices for data visualization, discovery, and accessibility	Information officer when incident occurs to coordinate data accessibility	high	high	Adjust incident management handbooks to include this, which is a high level decision.
	Need metadata training	easy	high	Online metadata training is currently available. Different levels of metadata training for different roles. Determine which entities need to take it.

	Implementation of keywords and ontologies used by data generators	high	high	Find out what vocabulary industry uses; across full range of data generators.
Developing best practices for maintaining Chain of Custody during discovery & accessibility	Lack of accountability and ownership	medium	medium	Use electronic submission. Understand litigation hold: General counsel defines minimum requirements for litigation hold.
	Multiple processes for chain of custody, per collector	easy	medium	Need for synthesis. 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 cannot get access to the actual data	medium	high	Raise awareness of the Open Data Policy, which gives policy guidance on this issue. Make users aware of why data is being restricted.
	When request comes for multiple data sets, uploader does not always have enough information about data and if it contains sensitive information.	medium	low during incident as everything is sensitive, high long term	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.

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.

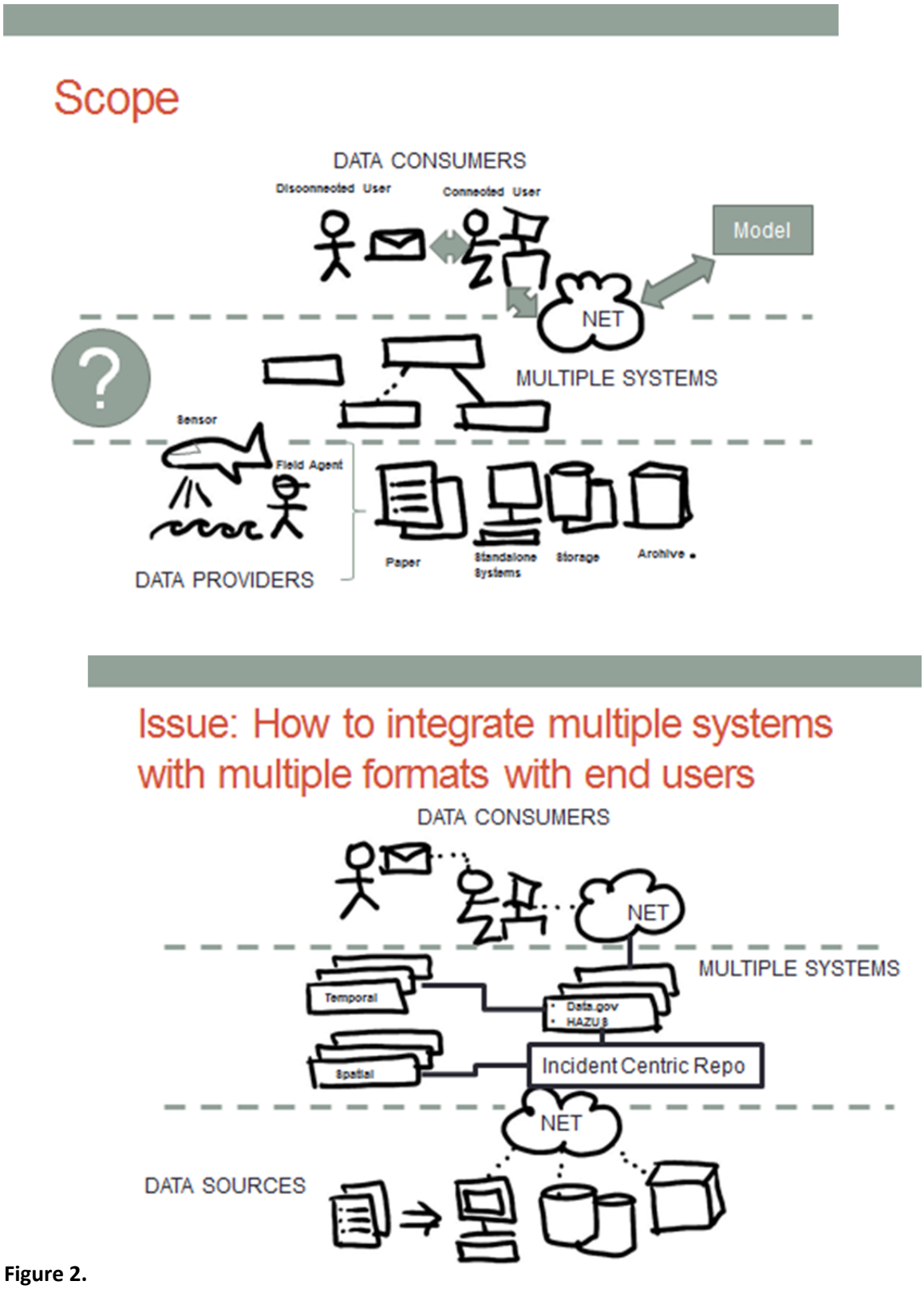

Figure 2.

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.

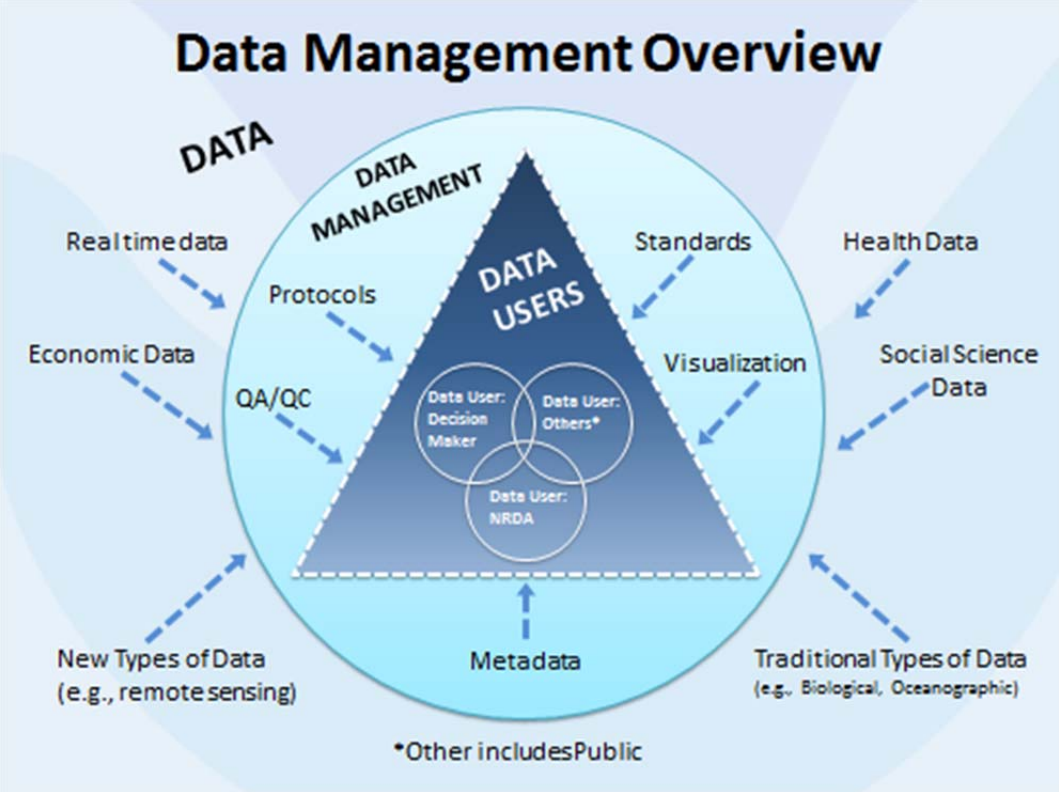

Figure 3.

- **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 Appendices

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