Assessment and Restoration Division Data Management Strategy

Ben Shorr - Lead Author
NOAA OR&R Assessment & Restoration Division
April 2015

This page is intentionally blank.

Table of Contents

1. Executive Summary

The goal of this Data Management Strategy is to support the Assessment and Restoration Division's (ARD) mission and objectives[1] by connecting specific ARD plans and actions with desired data management capabilities. The ability to consolidate and manage existing data holdings efficiently and sustainably ensure that ARD will be able to meet existing challenges and prepare for new ones. As the pace of technological innovation continues to increase, a strategy with the ability to evolve is required to connect data management needs and tools with core scientific questions ARD needs to answer. The vision for ARD data management over the next three to five years is to have coordinated data collection, data management, data query tools, and integrated processes and methodologies (Figure 1) that are driven by ARD's science and case development needs. At the case level, an effective case data management strategy requires that ARD staff quickly recognize the nature of the scientific questions that must be asked, and answered, for decision-making. These scientific questions will determine the types and extent of data/information that must be collected. A fundamental focus of this division-level data management strategy is implementing and promoting existing data standards that ARD uses, and developing or adopting standards to support case data management needs. Incorporation of this strategy into a Data Management Plan for ARD will provide guidance for case teams and project partners. The appendices of this data management strategy document additionally highlight data management requirements and challenges, with a high level of detail, discussion and input from ARD managers and division scientists.

[1] Assessment and Restoration Division's FY15-17 ARD Strategy and Operating Plan

Figure 1. Data Management Strategy Development

2. Introduction

The mission of the Assessment and Restoration Division (ARD) within the Office of Response and Restoration (OR&R) is to protect, assess, and restore NOAA's trust resources and the services they provide that are injured by hazardous waste releases, oil spills, and/or vessel groundings. To measure injury, ARD scientists employ a number of tools, methodologies, and processes developed internally and externally to NOAA. These tools are frequently used by OR&R to assist in environmental response and the natural resource damage assessment process (NRDA) to help protect, assess, and, ultimately, restore public resources that have been injured.

ARD's Spatial Data Branch (SDB) develops many internal tools and methods for the Division's information & data management, data quality assurance, data analysis and interpretation, and warehousing of these data and products. Additionally, the SDB is responsible for conducting training for these tools and methods and for working with the Damage Assessment, Remediation, & Restoration Program (DARRP)[2] case teams to ensure the effective application and transfer of this technology for division-wide use. Following this path, ongoing advances in data management technology coupled with increases in the volume of data being collected on a regular basis, provide ARD with the additional challenges of evaluating which tools should: remain in ARD's scientific toolkit, be combined, and phased out.

The advancement of data management technology combined with the large scale environmental restoration planning and work resulting from the Deepwater Horizon (DWH) oil spill have allowed ARD to develop scientifically defensible methods and tools to facilitate managing data and information. One outcome of the DWH work has been the realization that the development of effective and functional requirements for tools and data standards and the applicability and use of these tools requires DARRP case teams to work closely with SDB staff. As ARD moves forward, it is time to develop a clear Data Management Plan for the division; one that is driven by the needs of its mission and staff, and considers the broader needs of our partners and stakeholders.

To ensure the necessary level of exchange and communication required to develop an effective Data Management Plan, a series of data management calls and workshops were conducted between December 2013 and November 2014. Based on the responses and engagement of DARRP scientists in these calls and workshops, staff clearly understand and value the importance of data management and tools for analyzing, manipulating, and storing data. The outcomes from these discussions indicate that DARRP and ARD scientists are excited to be able to efficiently process increasing amounts of data for cases and want greater communication regarding the tools and functionality available to them. Beyond general awareness webinars, the results of these discussions demonstrate that staff wants, and needs, hands on, real-life scenario training to enable them to effectively use the tools and techniques developed for managing and querying data.

---

[2] The DARRP program is an arrangement among three components of NOAA: ARD of the National Ocean Service (NOS); the Restoration Center (RC) of the National Marine Fisheries Service (NMFS); and the Natural Resources Section of the Office of General Counsel. http://www.darrp.noaa.gov/about/index.html

Data Management Surveys and Workshop Highlights

The Data Management Strategy relied on input from ARD staff from various steps of an 18 month process (Figure 1). In December 2013, ARD held a data management workshop facilitated by the Coastal Response Research Center (CRRC) with representatives from all ARD branches, the Emergency Response Division (ERD), the Marine Debris Program, NOAA's National Oceanographic Data Center (NODC), the NOAA Fisheries Restoration Center (RC), and ARD data management consultants. Prior to the two-day workshop, ARD staff were asked to complete a survey and participate in branch interviews to gather an understanding of the techniques related to data management and analysis currently employed by ARD. A variety of tools and techniques were highlighted and common themes emerged to describe what is currently in use and what staff would most like to see added or improved. The notes from these branch interviews, survey results, and workshop notes are available in Appendix D.

To provide a common frame of reference for the participants, the workshop started with presentations on the importance of data management to ARD and overviews of existing tools. The workshop attendees were divided into breakout groups to tackle themes developed from the branch interview responses. The breakout group themes on the first and second days centered around common types of cases with which ARD is often involved, and specific ARD tools, respectively.

Common themes from branch interviews and the workshop included:

- The need for scalability and adaptability;
- The need to manage data types other than the traditional analytical chemistry and toxicity;
- The appropriateness of existing tools for small cases;
- The concern that legal requirements trump good science as primary drivers for data approaches;
- Tool lifecycles and updates to match those of the cases (3-5 vs. decadal case cycles);
- The need for more hands-on training with exercises geared to ARD/DARRP data and situations; and
- The ability for ARD/DARRP staff to apply and use tools themselves; without relying on experts.

3. Core Capabilities and Recommended Actions

Core capabilities and associated recommended actions, distilled from branch interviews, the workshop, and discussions with ARD staff are presented here. These findings contain references to tools, processes, and methodologies that are discussed at length in the Strategy and Appendices.

Core Capabilities	Recommended Actions[3]
1. Incorporate data science expertise into case management	• Assign Data Coordinator (from Spatial Data Branch) to each DARRP case team • Incorporate data management best practices (e.g., existing data models, common formats and standards, metadata, data and information management plans) into case management
2. Ingest and manage high quality datasets including field-collected samples, laboratory samples, field observations, and spatial data to support assessment and restoration.	• Expand data ingestion, standardization and query tools (DIVER) to serve national needs; coordinate closely with FAST team
3. Make data (and metadata) available to DARRP and partners for analysis, mapping and visualization, decision support and long-term stewardship	• Expand DIVER to serve national needs and leverage ERMA for case team data visualization and communication[4] • Review case data management and create or apply standard folder structures, documentation requirements, data exchange and back-up processes for critical data
4. Facilitate exchange of interim and final analysis products.	• Adopt or develop standards on a national or regional basis; • Develop standard procedures for data exchange • Provide funding and training for implementation. (Refer to secure file transfer protocol (SFTP), file collections and KMS discussion in Strategy and Appendices)

[3] Refer to Appendix for background and description of tools and methods including the Environmental Response Management Application (ERMA), Field Assessment Science Techniques (FAST), Data Integration Visualization Export and Reporting tool (DIVER), Knowledge Management System (KMS), Query Manager, and Habitat Equivalency Analysis (HEA).

[4] ERMA is not an independent data management system- organization and stewardship of spatial data should continue to be a case team collaboration- with the Spatial Data Branch Coordinator and the Regional Resource Coordinator playing key roles in organization, documentation and archive of spatial data files.

Core Capabilities	**Recommended Actions**[3]
5. Capture and share knowledge associated with cases and technical approaches	• Support ongoing development of KM capabilities for ARD science and case teams through collaboration on technology and provision of data and information references
6. Scale core capabilities and effort up/down in NRDA responses and cases (small, medium, large)	• Coordinate ARD Data Management Strategy with ARD's Science Plans and Operating Plans, and Data Management Plans/Data Sharing Agreement Templates (see FAST and Appendix for details)
7. Incorporate data management, information management, and sample management into guidance documents and job aids for on-going and response based activities (FAST)	• Integrate KM and FAST approaches with data management tools and technology • Support ongoing development of FAST capabilities and data management plans and sharing agreements • Provide FAST team training on existing samples, observation, instruments, and photographs common data models and collaborate on implementing for future efforts • Develop a telemetry Data Management Plan including common data model for telemetry data ingest and work with FAST to incorporate into workplan templates • Continue to develop the restoration common data model and build the capacity to link NOAA monitoring efforts to restoration projects through the DIVER system, while leveraging existing field sampling and analysis data models

Core Capabilities	**Recommended Actions**[3]
8. Administer application lifecycles • Develop tools based on needed functionality and requirements • Evaluate performance (meeting needs) and refinement/development • Plan tool/application succession • Train staff on tool use and new functionality	• Identify and demonstrate usability of tools and processes for cases requiring varying levels of effort from ARD (low, moderate, high) • Prioritize and communicate functionality and user interface modifications for ERMA • Review and incorporate photo management (Photologger) into DIVER file collection and work with FAST to include it in the field data collection forms • Conduct usability analysis to determine if processes and applications (tools) meet needs, or require improvement, training, or phase out • Expand training by SDB to include the full suite of internally-maintained tools (ERMA, DIVER, HEA Tools) • Create training on ARD relevant tools and techniques (e.g., Access/Excel/GIS) specifically designed for ARD/DARRP tasks. Host data-related trainings twice per year with priorities based on ARD/DARRP feedback • Have data managers and IT security staff establish procedures for granting access, rules of behavior, and credential refresh
9. Share and report data management capabilities and accomplishments internally and externally	• Leverage the DIVER Explorer tool to deliver summary information on restoration projects, including budget, and support web services of data and mapping outputs that reflect snapshots of time and near-real time information • Contribute data management perspective and techniques to NRDA-related training and workshops • Develop outreach and training materials • Share data-related accomplishments and training opportunities quarterly through OR&R weekly or office-wide email distribution

Core Capabilities	Recommended Actions[3]
10. Track budget and expenditures related to data management tasks	• Develop SDB time and efforts estimates for common tasks • Link budgets and expenditures to data management activities and application development
11. Coordinate Data Management development and processing	• Transition existing capabilities from Query Manager to DIVER capabilities • Use existing standards for management and analysis (e.g., Query Manager for contaminant chemistry and bioassay data, HEA, DIVER developed data models) • Involve regional ARD staff in scoping and development of tools and processes for data management and analysis

4. ARD's Scientific Approach

To protect NOAA trust resources, ARD's work is fundamentally driven by the scientific method (Table 1). This method involves characterization of the spatial and temporal extent of the contamination or impact (e.g., the amount of coral impacted by a vessel), hypothesis development, predictions from the hypothesis, testing or experimentation, analysis and improvement, and confirmation. These principles are applied to the NRDA cases and the priority projects where ARD provides scientific expertise and decision-making including Superfund cases. Other hallmarks of the scientific method are the incorporation of peer-reviewed data and literature and the communication of results and findings. ARD's ability to properly manage data and environmental knowledge is fundamental to its ability to address the types of scientific questions that drive its efforts to protect and restore natural resources. Additionally, ARD's stewardship of data and information that OR&R and DARRP produce or collect from others is critical for supporting injury determination and restoration.

TABLE 1

The Scientific Method in General Terms with Examples (iterative non-linear approach):

Questions
What is the pathway of exposure and extent of the injury? Can the injury be restored in the same area or does there need to be an alternate site? Are the existing samples representative or are more samples needed? If more samples are needed to reduce false positives, then where is sampling required statistically and strategically? What analyses are required and why? What levels of analytical sensitivity (e.g., DLs) are needed to answer the questions? EPA's Data Quality Objectives Process[5] *is often used to help frame environmental data collection efforts.*

Research
Team members study sites by surveying areas of contamination and non-contamination. This may include obtaining latitude and longitude, photography, collecting samples, and sketching areas of potential impact. Sample data are often gathered by the case team from prior studies. Tidal patterns, bathymetry, presence of trust resources, and other data are gathered. The team investigates the probable source/cause for the contamination. All data with location information are plotted on maps in a consistent fashion.

Hypothesis (with Examples)
H_o *= samples collected from the area of suspected contamination and in areas farther upstream of tidal influence are not significantly different in [PAHs] from our reference site.*
H_a *= Samples in both areas have significantly different [PAHs] than our control sites.*

Experiment
A sampling plan is designed to test the hypotheses. Then the plan is implemented (e.g., collect sediment samples per SOPs and sampling plan). Chain of custody best practices are followed by field data uploaded to a data management system and send to the lab for a specific suite of analysis.

Analysis
Initial analysis for the presence of contamination is done by an accredited lab. Samples undergo review including quality assurance/quality control and validation (depending on data quality objectives). Results are transmitted from the lab in a standard format. Field sampling results are plotted on a map and analyzed for spatial patterns and compared to guidelines or ecologically-relevant values. Spatial statistics, such as inverse distance weighting, are applied to provide estimates of concentrations at unsampled locations.

Conclusion (with Examples)
The DARRP case workers report the findings to the case team and suggest the acreage for remediation including the areas above tidal influence (assuming the null hypothesis is disproved). The DARRP case team continues to work on injury quantification and appropriate restoration. Negotiations and possible litigation are based on the interim analyses and findings throughout the whole process.

[5] Quality Management Tools – Systematic Planning, USEPA: http://www.epa.gov/QUALITY/dqos.html

Effective data management and the ability to query data for iterative analyses (and often provide that information to partners and stakeholders) is a core need throughout DARRP's scientific approach.

5. ARD's Science Driven Technology Requirements

5.1. Sources and Types of Data

ARD relies on vast amounts of data for assessment and restoration scaling and general case development. The sources of these data are extremely varied; being able to integrate these data and enable Regional Response Coordinators (RRC) and case teams to access and evaluate data as comprehensive datasets independent of the source is critical. Data sources can include responsible parties/potentially responsible parties (RPs/PRPs), state and federal remedial agencies, state and federal trustees, tribes or native groups, NGOs, and other environmental agencies and organizations. In addition to multiple sources, data are delivered in a variety of formats with varying degrees (or no) sense of quality assurance processes. The ability to receive, store, and make these data accessible for evaluation by environmental scientists is paramount to ARD's mission.

Over the past 30 years, ARD (and precursor organizations) have generally focused on managing contaminant chemistry and toxicity data (primarily in the Query Manager (QM) database structure – discussed in Data Access section and Appendix), Geographic Information Systems (GIS) data, and spreadsheets with calculations. Geospatial data are largely managed in GIS files organized into a fairly consistent file folder structure that varies regionally and by scientist. Before the DWH oil spill, recent ARD NRDA efforts involved managing field-collected information such as shoreline conditions, resource locations, and photographs. In its role as data management lead for the DWH NRDA, ARD has built the capacity to manage a broad array of data types.[6] A key tenet for ARD data management is to define data requirements and be capable of quickly ingesting standard (and new) data formats into the systems and tools. Preferred file formats and potential issues are discussed in more detail in the Appendix.

5.2. Metadata

Metadata is information about the individual data, and is critically important in making data useful for interpretation and analysis. NOAA has a mandate[7] for environmental data to be made available with its associated metadata. FGDC[8] compliant metadata for spatial information are most commonly

[6] See Section 6, Data Management Capabilities and the Appendix for a more detailed discussion of this issue.

[7] NAO 212-15 Management of Environmental Data and Information: http://www.corporateservices.noaa.gov/ames/administrative_orders/chapter_212/212-15.html

[8] Federal Geographic Data Committee (FGDC) metadata is the NOAA standard. The Content Standard for Digital Geospatial Metadata (CDSGM) known as FGDC-STD-001-1998 or simply "FGDC" is the Federal standard, with NOAA encouraging transition to ISO 19115 Parts 1 and 2 (International Organization for Standardization). http://www.ncddc.noaa.gov/metadata-standards/

created by OR&R (e.g., Environmental Sensitivity Index (ESI), SDB spatial analysis). Properly-completed metadata can also be incorporated into data portals or catalogs, making OR&R data discoverable by other users. If metadata is not provided to ARD staff by the author or provider of a spatial dataset used for analysis or mapping, the SDB will generally create a record (internally referred to as "metadata lite") in which it includes as much information as is available attached to a layer in ERMA or in a data package.

5.3. File Sharing/Repository

For ARD assessments and restoration planning, scaling and general case/project development are the collaborative effort of trustees from state and federal agencies and tribes. These data are often subject to litigation hold or have access sensitivities such as restrictions on cultural, proprietary, or sensitive data. Therefore, ARD work requires a secure repository (or repositories) for sharing data among collaborators. Historically, a secure file transfer protocol (SFTP site) administered by OR&R's Business Services Group (BSG) was used to enable this collaboration. This remains a valuable and flexible option with important limitations and liabilities.[9]

Federal Information Technology (IT) security controls must also be addressed for a secure collaborative site. Data managers and IT security staff must establish procedures for granting access, rules of behavior, and credential refresh. The repository also needs to be well organized, so collaborators can easily locate information and have the ability to grant/restrict access to data based on users' need for information or relationship to the case. The "File Collections"[10] approach which couples SFTP for data providers (e.g., laboratory, principal investigators) to deliver packages, flexible on-line forms for gathering field collected information (electronic and scanned), and data files (e.g., GPS, GIS, photos) has the potential to meet ARD requirements for flexible and sustainable case-specific file repository an IT security requirements.

5.4. Data Analysis Required Capabilities for ARD

While ARD has access to powerful spatial and non-spatial analysis tools in support of its scientific analyses, ARD case team members have varying levels of spatial analysis experience. In order to standardize and enhance spatial analysis capabilities and expertise across the regional branches, SDB members are liaisons to each of the four ARD regional branches and tasked to spend 50% of their time on case work for their branch. The goal of this organizational move is to create stronger linkages, through these former SDB staff, between ARD and DARRP regional case team staff and the tools and methodologies being developed and deployed by the SDB.

---

[9] Critical limitations (which is sometimes positive) to SFTP file exchanges are the lack of controls on file movement and organization structure, and the inability to provide sophisticated user-based credentials and privileges.

[10] DWH DIVER: File Collections established for DWH incident (https://logindiver.orr.noaa.gov). See Appendix for more details.

In addition to performing analyses themselves and relying on SDB staff, ARD case team staff also consult with RC staff or obtain external contract support. Examples of spatial analysis often performed by SDB staff in support of assessment and restoration include: spatial joins and attributes, buffer creation around areas of interest, bathymetric analysis to support HEA or restoration design, interpolations of contaminant chemistry for remedial investigations, and the evaluation of the relationships between chemistry and bioassay results.

Data analysis is a broad topic and covers much of the technical work that ARD staff do (in addition to project management). The Data Management workshop summary documents available in the Appendix lists examples of the types of data analysis that ARD staff commonly use in support of assessment and restoration including HEA, statistical modeling and summary, analysis of field and laboratory data compared to ecologically-relevant numbers. Nearly all of the activities that involve gathering, managing, and disseminating data will benefit from a strategy that connects data generators, users, and managers with standards and methods for near- and long-term stewardship. ARD has many partners who rely on exchange of data analysis information for cases across the division and multiple regions. These partners will benefit from a transparent ARD data management processes.

6. Data Management Capabilities

6.1. Data Warehouse Approach and Common Data Models

OR&R's "Common Data Model" approach refers to core sets of fields and schemas for data types (e.g., contaminant chemistry, visual observations, telemetry, photos) to facilitate ingesting, processing, and querying data. The DIVER suite of tools implements data warehouse concepts and common data models[11] to create a consistently updated central data repository for response and damage assessment activities. In order to build a data warehouse using common data models, schemas (or categories of information) must be created and the relationship between the categories and attributes must be detailed. Consistency between data models for common fields enables cross-category queries and analysis, in addition to relationships between data and results and documents such as scanned field sheets or laboratory data packages (often differentiated as structured vs. un-structured data).

6.2. Data Access

The ability to access datasets is critical to DARRP's and ARD's missions. In addition to SFTP and File Collections data repositories, ARD has several custom built tools for managing and accessing data including ERMA, Photologger, QM and DIVER. ERMA is primarily a visualization and mapping tool and also serves as a data discovery and distribution tool. ERMA relies on data providers or generators to manage their own information. The QM application provides standardized queries for contaminant chemistry and bioassay data. The Photologger application consists of an Access database for field photo management and an online tool for photo aggregation and access. The

[11] See Appendix C for an in-depth discussion on data warehouses and common data models.

DIVER suite of tools includes Explorer, a web-based query tool that provides access to contaminant chemistry and photos. More information including ARD survey responses, workshop summary and a detailed review of data management and access for these tools is in the Appendix.

6.3. Interoperability

Interoperability is the ability of tools or processes to exchange data and information efficiently. This can happen within an organization such as OR&R by collaborating on data standards and specifications and then implementing those standards for import and export functionality. Organizations such as the Open Geospatial Consortium (OGC) focus on establishing standards in order to facilitate interoperability between systems. While adopting open standards can promote interoperability, it may also be achieved using custom data formats (example: contaminant chemistry in QM database standard or ERMA data packages between DIVER and ERMA) so long as the system owners share the data format specifications with each other and different systems are able to meet data format requirements.

6.4. Field Data Management

The Field Assessment and Support Techniques (FAST) team's mission statement is to "*…provide national leadership in NRDA science, assessment methodology, and spill response*". FAST's data management goal is to improve knowledge and data management in NRDA field assessment. The FAST team is improving ARD's overall field assessment capability by generating or adopting tools, methods and processes, and expanding staff training. Products include: resource work plan templates, data collection protocols, sampling plan templates for planning and implementing field assessments, archives of field assessment tools, guidance on field assessment techniques and tools (e.g., photography, data management), and Incident Command System (ICS) logistics and field contacts.

The SDB, in collaboration with the FAST team, is developing data management plans, data sharing agreements, and electronic data capture processes to assist in the ingestion and management of field data. [12] As the FAST team is establishing core field data attributes in collaboration with state and federal partners, the SDB data management team is working to establish direct connections to these field forms so that they can be directly ingested into the "File Collection" system. This effort will: (a) reduce the burden and expense on data management in the field, (b) incorporate field team review of data, (c) reduce the amount of time between field information collection and review, and (d) enhance NRDA flexibility for data collection and management.

[12] See Appendix and FAST Library on Google Drive

6.5. Restoration Project Information Management

Restoration is a critical aspect of ARD's and DARRP's mission. Managing these data in the initial stages of restoration planning through the restoration's long term monitoring phases is important to ensure the public is appropriately compensated for the quantified natural resource injuries. Restoration information is often generated concurrent with response and damage assessment efforts, and can be managed with a flexible data model or set of data models. DWH restoration is currently managed in DIVER through a restoration projects data model which is accessed and updated by RC staff, and used to serve information on existing and potential restoration projects. See Appendix for more information.

6.6. Knowledge Management (KM)

OR&R is currently undertaking a phased KM project to facilitate: information-sharing among staff, and more effective access to scientific information. The KM team is conducting a needs assessment to determine priorities for content and functionality of the system. The next steps will be to identify potential systems and tools that could meet those needs and to implement/develop a system.

KMS is expected to relate to data management by providing staff with knowledge on OR&R's data management tools, potentially including pointers to user manuals/guidance documents. KMS may also contain reports describing the data in OR&R's databases (e.g., data collection reports).

6.7. OR&R IT Security

Ensuring security for data, files and applications is a primary responsibility and requirement for ARD, which must be integrated into its processes and approaches to data acquisition and management. As part of OR&R, ARD adheres to IT policies and procedures, which are based on Department of Commerce, NOAA and NOS requirements and policies for reducing security vulnerabilities. ARD is also responsible for acquiring and managing data that meets litigation hold requirements (in addition to NOAA's normal document retention requirements) and satisfies chain of custody protocols. Additionally, software applications that are developed or used by ARD must be in compliance with NOAA IT security requirements. ARD technical leads and development teams should continue to work closely in partnership with the Business Services Group within OR&R to ensure that IT security is incorporated into all phases of process and application development from inception to retirement or archive.

7. Communication and Outreach

Discussions in the branch surveys and the data management workshop in late 2013 revealed (1) low levels of awareness exist among ARD staff regarding the full suite of tools currently available for common tasks, and (2) a desire by staff for enhanced awareness of and training with these tools. The FAST team has made progress toward enhanced awareness by hosting monthly seminars beginning in 2014 on a variety of topics from overviews of case work to presentations on tools and technology being implemented in ARD, OR&R, and/or NOS. FAST includes data management-related seminar topics. In addition to FAST seminars, OR&R is in the process of creating a "tool glossary," available to all staff from a shared location (Google Drive document titled "30 tools."). The KM approach is additionally evaluating mechanisms for discovering literature and previously implemented techniques as resources for current staff work.

8. Training

The data management assessment activities identified a need for relevant and interactive training. People identified the need for tailored hands-on training for ARD-specific case situations and issues. Several people referred to the type of training they would like to see as "give me a word problem and step me through solving it." Additional training on commercial off-the-shelf software (e.g., Excel, Access, ArcGIS) was also highlighted in the data management surveys and interviews. They stress that training should also be made as flexible as possible, in person, and recorded or web-enabled for self-paced/on demand training.

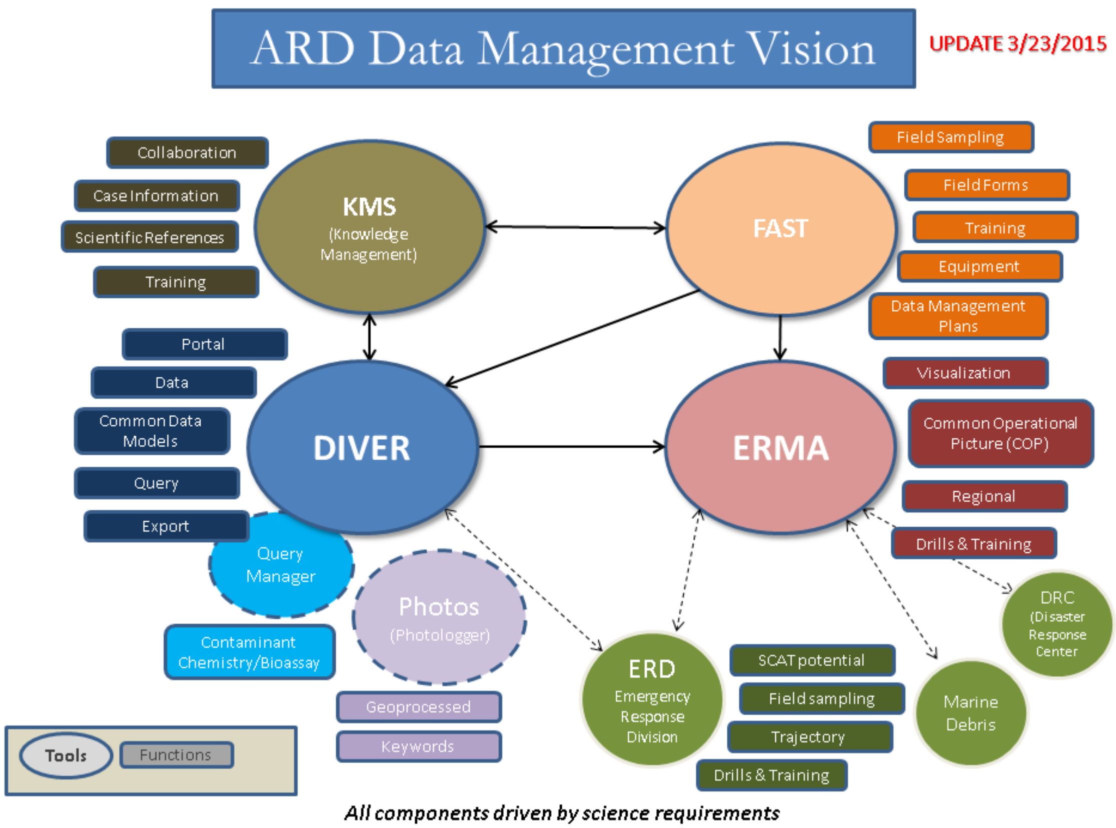

Figure 2. ARD Data Management Vision (Graphic)

Figure 2 captures the vision of related systems, processes and tools within ARD that provide a flexible and scalable infrastructure for managing, accessing and sharing data and information.

ARD Data Management Strategy: References and Appendices

Contents

Appendix A: Abbreviations and Glossary

Assessment and Restoration Division (ARD)

Spatial Data Branch (SDB)

Regional Resource Coordinator (RRC)

Office of Response and Restoration (OR&R)

Emergency Response Division (ERD)

Marine Debris Division (MDD)

Damage Assessment, Remediation, and Restoration Program (DARRP)

NOAA Restoration Center (RC)

Some of NOAA's legal authorities for coastal resources assessment and restoration (more information on www.darrp.noaa.gov/about/laws.html:

- **Comprehensive Environmental Response, Compensation and Liability Act of 1980 (CERCLA or "Superfund")**
- **Oil Pollution Act (OPA) of 1990**
- **Clean Water Act**

Data (as developed during ARD Managers Breakout Group; Oct. 2013 Data Management pre-workshop)

- information which can be tied to a specific point spatially and/or temporally
- single point of information which can be interpreted to answer a question
- individual record-based information that can be tied to location, points in time or space, (observations, concentration measurements, photographs)

Data Management
"Data Management is the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver, and enhance the value of data and information assets."[1]

[1] Data Management Association: Guide to the Data Management Body of Knowledge
http://www.dama.org/i4a/pages/Index.cfm?pageid=3364

Information
As developed during ARD Managers Breakout session; Oct. 2013 Data Management pre-workshop:

- all records collected when working a case - reports, photos, etc.
- not by definition tied to a data point
- everything from documents, reports, photos; information associated with case
- more than just a number or single point in time
- results from data gathered in the Remedial Investigation (RI) process

Working definition developed by OR&R Data Management Workgroup (03/2015): "Information is generated by or the result of data analysis or synthesis".

Damage Assessment, Remediation, and Restoration Program (DARRP)
DARP is a multi-office program within NOAA involving the National Ocean Service, the National Marine Fisheries Service, and the Office of General Counsel. DARRP scientists, economists, and attorneys conduct natural resource damage assessments of and restoration projects for coastal and marine resources injured by oil and hazardous material releases. [2]

Data Integration Visualization Exploration and Reporting Tools (DIVER)
The DIVER suite of tools contains data models with a data warehouse and query tools in support of environmental decision-making. The data warehouse incorporates environmental intelligence – a set of techniques and tools for the transformation of raw data into meaningful and useful information for analysis. More discussion is available in Appendix B.

Environmental Response Management Application (ERMA®)
ERMA is an online mapping tool that integrates static and real-time data, such as Environmental Sensitivity Index (ESI) maps[3], ship locations, weather, and ocean currents, in a centralized, easy-to-use format for environmental responders and decision- makers. [4] More discussion in Appendix B.

HEA (Habitat Equivalency Analysis)
HEA is a method used by NOAA for estimating how much restoration is needed to replace the loss of natural resources from the time they are injured until they are returned to the condition they would have been before the release.[2]

Hazardous Substance
Substances identified as capable of posing "imminent and substantial danger to public health and welfare or the environment." CERCLA has identified more than 800 hazardous substances. The term does not include petroleum or natural gas.[2]

---

[2] Damage Assessment Remediation & Restoration Program: http://www.darrp.noaa.gov/

[3] Environmental Sensitivity Index (ESI) Maps http://response.restoration.noaa.gov/maps-and-spatial-data/environmental-sensitivity-index-esi-maps.html

[4] Office of Response and Restoration website: http://response.restoration.noaa.gov/

Injury
Injury is an observable or measurable adverse change including destruction, loss, and loss of use in a natural resource or impairment of a natural resource service.[2]

Injury Assessment and Restoration Planning
This is the second phase of a natural resource damage assessment. Trustees identify the injuries to natural resources and their services and use that information to determine the need for and amount of restoration.[2]

Electronic Data Deliverables (EDD)
An Electronic Data Deliverable refers to data delivered from a laboratory or analysis in a format or template typically specified by the data owner. An Electronic Data Deliverable can be an opportunity for the data manager(s) to specify the most efficient format and structure to receive data, which may minimize the amount of standardization and transformation necessary to conform to database standards.

Field Assessment Science Techniques (FAST)
The FAST team's mission is to improve ARD's overall field assessment capability by generating or adopting tools, methods and processes and expanding staff training. Data management plans and data sharing agreements for use in incidents (e.g., hazardous materials, severe weather) developed by the SDB are incorporated into the FAST team's efforts.[5]

Knowledge Management System
OR&R is developing a web-based Knowledge System to facilitate information sharing among staff and access to scientific information. In support of this project, the KM team has conducted a needs assessment to identify priority knowledge-related challenges and determine solutions.[6]

MARPLOT
Marplot is a mapping application jointly developed by USEPA and NOAA that is part of the CAMEO software suite, a system of software applications used to plan for and respond to chemical emergencies.[7] A previous version of MARPLOT supported interaction with the Query Manager desktop application for quick viewing of query results. More discussion is available in Appendix B.

Natural Resource Trustees (Trustees)
Government officials who act on behalf of the public when there is injury to, destruction of, loss of, or threat to natural resources as a result of a release of a hazardous substance or discharge of

---

[5] Data Management Plans and Data Sharing Agreements
https://drive.google.com/drive/#folders/0B1PadSHGyfSNZVJTWTJRUnJGTTg

[6] Sept. 2014 Knowledge Management Needs Assessment
https://drive.google.com/drive/?pli=1#folders/0B1nxVqRwgnpBTVFGNU8wQUlfNWs

[7] CAMEO Software suite, USEPA: http://www2.epa.gov/cameo/what-cameo-software-suite and NOAA OR&R MARPLOT: http://response.restoration.noaa.gov/maps-and-spatial-data/marplot.html

oil. Trustees include the U.S. Departments of Commerce, Interior, Defense, Agriculture, and Energy; state agencies; and Native American tribes. NOAA is the lead federal trustee for coastal and marine resources. [2]

Natural Resource Damage Assessment (NRDA)
Investigation performed by trustees to identify and plan the restoration of natural resources injured by oil spills and hazardous substance releases. The goal of NRDA is to restore natural resources. Environmental laws addressing contaminants in the environment include Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA/Superfund) and the Oil Pollution Act of 1990 (OPA).[2]

NOAA Environmental Data Management Committee (EDMC)
The Environmental Data Management Committee (EDMC) coordinates the development of NOAA's environmental data management strategy and policy, and provides guidance to promote consistent implementation across NOAA, on behalf of the NOSC and CIO Council.[8]

Query Manager (QM)
Query Manager is an ARD developed database standard and also a query tool (mature desktop application and beta web version) that can be used to access sediment, tissue, water, and oil chemistry results, and sediment and water toxicity data. More discussion is in Appendix B.

Rapid Assessment Program (RAP)
A capability developed and supported by the DARRP to collect perishable data and readily available information to determine the need for a natural resource damage assessment.[2]

Responsible Parties (RP)
The parties (e.g., individuals, companies, government agencies) responsible for an oil spill or hazardous substance release. This is also referred to as Potentially Responsible Parties (PRPs) under CERCLA.[2]

Restoration
The goal of a natural resource damage assessment, which involves rehabilitating, replacing or acquiring the equivalent of injured natural resources and the services they provided. Restoration includes primary and compensatory restoration projects.[2]

Secure File Transfer Protocol (SFTP)
A network protocol that provides file access, file transfer and file management functionality. There are several advantages to this protocol which have made it critical for use in litigation sensitive data management and exchange.

Natural Resource Trustees
NOAA is a natural resource trustee, acting on behalf of the public to manage, protect and restore coastal and marine resources.[2]

[8] NOAA EDMC: https://www.nosc.noaa.gov/EDMC/

Appendix B: ARD Data Management and Query Tools

Query Manager (QM)

The QM database provides contaminant chemistry and toxicity data in a consistent format including national, regional, and site-specific information. It is a unique national resource used by state and Federal agencies, industry, and the public. QM is a database standard as well as a query tool (mature desktop application and beta web version) that can be used to access sediment, tissue, water, and oil chemistry results, and sediment and water toxicity data. QM data processing for ARD has been done by a team of skilled scientists with database, chemistry, and toxicity background. The QM data process results in organized data sets from multiple studies in a consistent and standardized structure, which improves data delivery and supports interpretation, mapping, and analysis.

Users can sort and examine data in a variety of ways by selecting from a menu of pre-programmed queries. These queries allow the user to evaluate individual contaminants and contaminant groups, make comparisons to common toxicological benchmarks, calculate chemistry totals, and apply toxicity models.

QM has been instrumental in organizing data, and the core concept of common data models and query tools is built upon this software. QM was originally developed as a freely available cross-platform application, and is available for download from the OR&R website[9]. The QM application is developed in Microsoft FoxPro language, which is no longer actively supported; however, QM is still functional. [10]

For more than 20 years, the QM database standard has been used for sediment, tissue and bioassay data. Regional water data was previously managed- but not made available for query and download- through the QM application. The Deepwater Horizon NRDA was the impetus for an expansion of the QM database standard to include a vast amount of water and oil matrix data. Currently, the QM database standard is used for managing the majority of contaminant chemistry collected under the DWH NRDA. Where possible, ARD should continue to use and adapt the QM database standard for contaminant chemistry and bioassay data. The use of data warehouse tools for ingestion and query/export of QM data will support data status reporting and integration with other datasets.

---

[9] NOAA OR&R Download Query Manager and MARPLOT Applications, Data, and Maps: http://response.restoration.noaa.gov/environmental-restoration/environmental-assessment-tools/query-manager-marplot-data-maps.html

[10] Microsoft Support Microsoft Support Lifecycle: http://support2.microsoft.com/lifecycle/search/?sort=PN&alpha=foxpro

Several key QM queries have been ported to DIVER Explorer Guided Queries including station and sample level queries, export of results, single and multiple contaminant queries, Total PAH calculations, and comparisons to guidelines. The Guided Queries paradigm within DIVER Explorer allows development of specialized queries of contaminant chemistry based on ARD current and future needs and requirements.

MARPLOT

MARPLOT is a mapping tool developed by ERD and USEPA for hazardous material response. The QM desktop application had a direct connection with MARPLOT – with easy display of queries on top of simple basemaps. MARPLOT 3.3 integrated with desktop QM. QM is no longer supported, but is still functional and is available for download from OR&R website. ERD does not support interaction with Query Manager with release 5.0 and is seeking to retire MARPLOT 3.3. The DIVER Explorer application has built upon the Query Manager to MARPLOT mapping concept with direct mapping of query results within the Explorer application, and flexible symbology based on user entered values.

Microsoft Access Customization

ARD has developed several Access tools with custom scripting (typically Visual Basic for Applications – VBA). The OR&R Activities database, which is used for: tracking OR&R activities across all divisions, supporting the OR&R website, the DARRP website and the DARRP report; contains custom VBA coding for entering information, integrating data from multiple sources, generating mapping layers, and reporting.

ARD's SDB has also developed custom Access database solutions for managing and presenting restoration options in the Northeast Region/Hudson River.

NOAA's contractor, EXA Data and Mapping[11], has developed custom Access and Excel tools for importing contaminant chemistry data from laboratories (laboratory EDDs) and field collected contaminant chemistry sample information (Field Simple Sheets). These efforts have been instrumental in facilitating data flow from laboratories and field sampling for contaminant chemistry, and are being incorporated into the FAST team's efforts.

ArcGIS Custom Tools

ARD initially developed a set of tools to facilitate spatial analysis and importing contaminant chemistry data in ArcView GIS. These tools have not been updated in several years, and the functionality has largely been incorporated into the core ArcGIS application. More recently, the SDB has been developing and using custom tools and models in ArcGIS using Python to support spatial analysis. As the level of programming expertise increases within the SDB, there is a need

[11] EXA Data & Mapping Services: http://www.exadata.net/

to centralize a repository of code and implement code versioning. This will benefit regional branches with consistent tools and scripting, and provide a common basis for scripting across OR&R.

PhotoLogger

A critical aspect of data collection for response and NRDA efforts is the collection and management of case-related photographs. Photos can be used by staff for the qualitative (visually descriptive such as shoreline oiling) or quantitative (used for counting or classifying amounts) evaluation and documentation of a site. OR&R has developed and uses a custom application called PhotoLogger to support the management process for incident and case photography. The fundamental purpose of PhotoLogger is to collect, archive, organize, "log" or attribute with keywords related to the photograph or field effort and location information, and search and display field photos collected by staff and partners. PhotoLogger consists of a desktop application called APL (Access Based PhotoLogger) and an online application called OPL (Online-PhotoLogger). Photos are generally processed with software that appends GPS locations onto the photos internal metadata prior to uploading and logging into Photologger. Data collection for the DWH NRDA has involved uploading photographs and related documentation including keywords from PhotoLogger into DIVER file collections. The DIVER and FAST teams have also been collaborating on incorporating photographs into field sampling forms. These observational data products provide direct visual evidence of case priorities, documentation for response, and a photo library of OR&R activities.

NOAA HEA Tools

The NOAA HEA Tools began development in 2008, based on discussion and priorities identified in a December 2007 HEA workshop led by UNH-CRRC[12] to address the need for more documented and repeatable HEA's on complex, multi-contaminant cases. The HEA tools consist of custom GIS tools developed in the ArcGIS toolbox which produce standardized inputs in database format, and calculations in a Microsoft Access database application. The GIS tools are a series of steps with some flexibility that develop HEA inputs on a common grid basis. Processing steps are recorded and metadata are developed as the analyst moves through the process. The Access tools consist of a series of steps that navigate the user through inputting thresholds which correlate to levels of injury, key dates and options for calculating HEA, and present an output of discounted service acre years (DSAYs) with limited charting and table output functionality. The HEA Tools are actively used on several cases in the Northwest and have increased efficiency, documentation and collaboration. Issues with the GIS portion of the tools, which are written in Python, stem from the need to update the tools as the ArcGIS software (ESRI) is updated, and underlying code base evolves. The Access portion of the tools has had

[12] Coastal Response and Research Center, HEA Metrics Workshop (December 2007): http://www.crrc.unh.edu/workshop/crrc/hea-metrics-workshop

recurring issues with NOAA customization limits and its ability to handle large sites (to a large number of grid cells for calculation). Continued use and refinement of the HEA tools is needed with a focus on finalizing documentation and providing internal training sessions for NOAA and trustee partners.

Other regional staff continues to use spreadsheet calculations for HEA, which in many cases is built using GIS and QM data. KMS and/or DIVER may provide some structure for management and exchange of case-specific HEA.

Environmental Response Management Application (ERMA®)

The Environmental Response Management Application (ERMA®) enables a user to quickly and securely upload, manipulate, export, and display static and real-time geographic data sets. This allows for high-impact and fine-resolution visualization of data for solving complex environmental response and resource issues.

ERMA is designed to:

- Aid in spill preparedness and planning,
- Assist in coordinating emergency response efforts and situational awareness for human and natural disasters,
- Help define the extent of potential environmental impacts, supporting the NRDA process,
- Support ecological recovery and restoration efforts,
- Provide access to this information from anywhere an internet connection is available,
- Visualize data from a variety of sources, with the ability to include additional media such as photos and links to scientific reports, and
- Tell a story or reconstruct the history of an event using animated layers of information.

The application has been customized with specific data for ten ERMA sites covering the United States and its territories. ERMA sites have secure-access accounts; however, publicly-available environmental and baseline data are openly accessible for all regions.

The application is based on open-source software (PostgreSQL/PostGIS, MapServer, and OpenLayers) that meet Open Geospatial Consortium (OGC) specifications and standards used across federal and international geospatial standards communities. This ensures ERMA is compatible with other commercial and open-source Geographic Information System (GIS) applications that can readily incorporate data from online data projects and avoids licensing costs. Open-source compatibility supports data sharing, leverages existing data projects, reduces ERMA's maintenance costs, and ensures system flexibility as the technology advances.

ERMA is faster and easier to use than some proprietary GIS tools, and because it is open source, it can easily be customized to meet individual requirements. ERMA allows users to download data sets of interest and incorporate them into desktop GIS, Google Earth, or partner GIS internet mapping sites.

ERMA provides non-GIS users with the ability to quickly interact with and analyze information. For example, a polygon can be drawn on the map to communicate a proposed fisheries closure area or a restoration project, and this information is then instantly available to other users.

ERMA draws upon and aggregates existing data to facilitate their interpretation and to display risk scenarios.

Users can draw shapes to select areas (and data) on a map, which they can combine with other spatial data sets and produce a summary report. ERMA also allows the user to measure distance, draw a perimeter to calculate the area of a selected polygon, save certain views on a map, animate data layers, and label points of interest when printing customized maps.

ERMA has a multi-level security structure. Most people can see all publicly-available data layers without having to login, but only authorized users can access, sensitive information (e.g., during a spill, climate change analysis). Views (specific data displays) can also be tailored for particular groups such as responders, restoration specialists, or legislators.

The key to effective emergency response is efficiently pulling current science, information technology, and real-time observational data into the decision-making process. ERMA can act as the common operational picture for response ICS commands or units, keeping decision-makers up-to-date on the latest response information. ERMA is a robust system, proven to handle tens of thousands of data layers, that allows access to a wide array of response-related data.

Data Formats and Types

Some of the common data formats that ARD uses or creates to manage data include, but are not limited to: Shapefiles[13] (de facto GIS standard), CAD drawings, image files (e.g. JPG, TIFF-georeferenced when appropriate), GeoRSS[14], web map services, and KML 2.2 as OGC[15] or ISO[16] standards. For text or database files, ARD prefers files be shared as .csv, .dbf, .xls, .txt, .mdb or other standard file structures where fields can be mapped to the common data models.

[13] ESRI Shapefile geospatial vector format; mostly open specification https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

[14] Geographic Rich Site Summary feed (GeoRSS): http://www.georss.org/overview.html

[15] Open Geospatial Consortium http://www.opengeospatial.org

[16] International Organization for Standardization: http://www.iso.org/iso/home.html

There may be file size limitations or data conversion/storage details (e.g., losing precision) which should be considered when choosing a file type for use in analysis and data management.

Physical, Chemical and Toxicity Data

ARD uses a standard format for ingesting and managing contaminant chemistry and bioassay data for cases and projects spanning remediation and damage assessment. Cases that ARD has been working on for a long period of time typically have chemistry and bioassay data from tens to hundreds of studies that have been gathered from many sources and agencies. These have been compiled by QM.

Some examples by region:

Region	Example Databases
Pacific Northwest	Duwamish River (part of Puget Sound); Portland Harbor;
Great Lakes	All combined in regional Great Lakes database: Kalamazoo; Sheboygan River; St. Louis River; Maumee River
Southwest and Oceania	California; Pacific Islands including Guam
Southeast/Gulf of Mexico	Deepwater Horizon; Historical Gulf of Mexico; Vieques Island
Northeast/Atlantic	Hudson River; NY/NJ Harbor; Delaware River Estuary
National Studies	National Status and Trends (NOAA); EPA EMAP

These databases form the foundation for technical analysis and format all stages of a case from supporting remedial alternatives to establishing damage assessments to settling cases. QM has not been employed for all ARD cases for several reasons including: project partners leading data management, time required for data ingest too great for case schedule, and lack of awareness of the QM database standard. ARD projects and cases often have lifecycles of years to decades and therefore implementing consistent database standards preserves institutional knowledge and better enable continuity of case management and analysis techniques if the case team should change during the lifecycle of the projects.

Data Processing

QM data processing has typically involved conversion and standardization of data from a provider (e.g., state, RP, consultant) where the field sampling information and the laboratory analysis and results are provided together as exports from an established data management system or ad hoc files. This type of data processing often involves coordination with the data provider to resolve naming conventions, valid values and unique relationships between samples and results and their attributes. The level of effort can vary based on the organization and quality of data management, and often involves intense detective work to resolve key data questions (e.g., sampling locations, lab methodologies).

Electronic Data Deliverable

The QM Electronic Data Deliverable (EDD) is a data template that is often provided to laboratories and project partners, and is used as a template for delivering analytical and bioassay datasets for conversion to the QM data standard. The EDD has an effective conversion process where a data provider aligns fields and nomenclature to the data standard. The user is able to iteratively resolve data issues at the beginning of a project and deal with outliers or anomalous data as the project progresses. This is a proven example of using a data template or common data model to initialize data exchange.

Visual Observation

The effort to collect detailed information for the DWH NRDA resulted in the development of an "Observation" or "Visual Observation" common data model within the DIVER system.

In the past, observational data collected in the field under the auspices of a workplan was typically uploaded into the file collection structure and included a scanned copy of the field notes and field sampling forms. A unique "Field Trip ID" was used to relate information for specific field sampling trips such as photographs, GPS files and supporting notes and documentation upon upload to the file collections system. The field sampling forms, which contained details such as marsh observations, marine mammal observations, or oyster sampling details, were used to create an electronic field sampling form in the file collection. A data analyst on the data management team would then transcribe the field sampling form information into electronic format. For quality assurance, a second data analyst would do 100% transcription verification of the electronic data based on the scanned field sampling form.

Within DIVER, the core data elements from visual observations are incorporated into the Visual Observation common data model, with a relationship between a sampling event and the field collection forms based on the sampling trip ID. A user can query and export field collection form information through DIVER Explorer. The FAST team is developing field data collection protocols and processes that build upon and integrate with DIVER data ingestion and management.

Oceanographic data

The Instruments common data model evolved from a tracking database that was developed during the DWH response for managing all vessel-based information including deployed instruments (e.g., CTDs with fluorescence), acoustics, water sampling efforts, and biota collection (generally from towed gear). This database also contained useful information about the roles of personnel who are often critical during post-processing and analysis of oceanographic data. Although this data model was adapted for use in the NRDA, it would be most useful to collaborate with the National Oceanographic Data Center (NODC) on existing standards and methodology for data collection, transfer and management and develop a common data model that supports direct oceanographic data ingestion and overall cruise information management.

Remote Sensing

Remotely sensed information is being used more frequently to examine a variety of environmental concerns from remedial design for hazardous waste sites and damage assessment investigations for casework to longer term monitoring activities and restoration project evaluation.[17] ARD works with other NOAA program offices including the National Geodetic Survey (NGS) and partner agencies such as the U.S. Geological Survey (USGS) on the acquisition of high resolution airborne and satellite-based sensor systems for shoreline and habitat assessment and other site characterization or general mapping activities. DWH significantly increased OR&R's use of remotely-sensed data and has improved its ability to use these assets effectively in support of its casework. Aerial photography was used extensively for the DWH shoreline and Submerged Aquatic Vegetation Technical Workgroup's (SAV TWG) habitat assessment. A variety of satellite-based visible and infrared sensor data are at the foundation of the DWH open water exposure and injury assessment for trust resources.

ARD must develop a specific strategy for ensuring efficient access and maintenance of remotely-sensed imagery, including the ability to acquire or get access to new imagery related to an emergency response or incident.

[17] Graettinger et al., Operational Utilization of Aerial Multispectral Remote Sensing during Oil Spill Response. Photogrammetric Engineering & Remote Sensing, Volume 78, Number 10/ October 2012

Appendix C: Common Data Models, Data Warehousing and File Collections

The development and use of a "common data model" approach by NOAA's OR&R is rooted in the need to clearly and transparently manage core types of information that are collected and used during response and damage assessment activities. This "common data model" approach leverages data warehouse concepts for creation of a central repository of data that is updated with new or appended data, and allows for incorporation of metadata level records which refer to externally stored datasets. The concept involves creating data schemas, which are essentially categories of information and the relationship between the categories and details. An established "common data model" that can be shared with data providers (those collecting information) and data managers (those managing information) provides a collective blueprint for aligning definitions and meaning, in addition to supporting templates for exchanging packets of information. The "common data model" concept is flexible and scalable. As new or different information is received, data managers have the flexibility to evaluate whether that information fits into an existing schema, or if it requires a new or expanded one. These data schemas or organizations are not dependent on a specific database technology or tool, although open source software allows management of routines that extract data from sources, transformation into the common data model, and input into the data warehouse.

Examples of common data models that are created for response and NRDA are:

Data Models (schemas)	Description
Samples	Field-collected information combined with laboratory results
Observational Data	Field observations rather than collected samples (e.g., marsh condition, visual oiling, bird counts)
Photos	Photographs and associated keywords and comments
Telemetry	Tracking of movements of marine mammals, turtles and fish
Instruments	Oceanographic sensor data (e.g., CTD, acoustics) mostly collected onboard ships
Restoration	Details on potential and implemented restoration activities, including budgeting and monitoring

Core Fields

These data models share a common core set of fields in addition to data type-specific fields. The common core set of fields facilitates querying across data models to provide a comprehensive view of the information collected spatially and temporally. The common data models have a hierarchy based on a "workplan" (planning document) which specifies the overarching characteristics of the data (high level) with increasing levels of specificity listed thereafter. This type of document typically establishes the purpose and methodology for collecting information.

Understanding, and in some cases anticipating, the types of scientific questions that will be answered using the data is critically important when structuring and managing the data. This allows for query tools to be built for data interrogation and exporting. Other high levels of organization across the common data models are at the "trip" level – which is typically defined as an event where environmental information is collected, and "Station" or "Site" – which is usually a discrete location or path. Additionally, implementation of a common data model incorporates the geometry associated with the data, which can be single locations (points), multiple locations, complex lines, or polygons (areas).

The following are examples of "Core Fields" with definitions:

Name	Definition
Data Category	General category of data collection (e.g., instruments, photographs, samples, or visual observations)
Workgroup	Technical Working Group under which the field data are collected
Collection Workplan	The workplan under which the field data were collected.
Collection Form	Data submission form used by the field team to submit raw field data
Collection Matrix	Samples or records collected (e.g., sediment, water, photograph, wipe)
Analysis Category	General category of analysis performed (e.g., plankton, nekton, visual observation, contaminant chemistry)
Analysis Type	Subcategory (i.e., type) of analysis performed (e.g., biomass, hematology, genetics)
Analysis Status	Status of samples in the analysis process as reported by laboratories or through results (e.g., archived, results available, in analysis queue)
Image Id	Record identifier for a particular photograph
Case/Activity	Name of the case incident or the activity used to collect data
Data Classification	Purpose for which data was collected within the case incident or activity
Data Source	Originating owner of the dataset
Sharing Status	Extent of data distribution (e.g., publicly available)
Review Status	Extent of data quality review performed.
Trip ID	Identifier for tracking field collection events and the way data files are provided to the data management team (e.g., one Trip ID per file collection, zip file)
Record ID	Identifier for each observation data sheet entered into the NOAANRDA.org database
File Collection ID	Record identifier for the corresponding file collection on NOAANRDA.org
DIVER Dataset	DIVER's internal database table name
Start Latitude	Sample latitude in decimal degrees, or start Latitude for samples collected by trawl, transect or in a line

Name	Definition
Start Longitude	Sample longitude in decimal degrees, or start Latitude for samples collected by trawl, transect or in a line.
End Latitude	End latitude in decimal degrees for trawls, transects or samples collected in a line
End Longitude	End longitude in decimal degrees for trawls, transects or samples collected in a line

Related Information

The environmental data commonly collected has several related types of information that need to be accessed for analysis and interpretation (in addition to satisfying legal requirements for information management). A system called “File Collections” allows for gathering structured and unstructured information and also establishes the relationships between different types of information. The File Collections, which is conceptually a data vault approach, is typically accessed through an on-line upload form (can also be done off-line in bulk) that collects all primary documents (e.g., scanned field sheets, data logs, electronic field sheets or forms, GPS and GIS files), in addition to metadata level information (e.g., Workplan or group that was collecting information, general location and individuals involved). The “File Collections” is flexible to handle uploads of new or unknown information that can be sorted after initial field collection is over, with sections for processing and analysis collections. A major advantage to having an established centralized repository for related environmental data is the available to support queries across data holdings and connection of analytical results with the supporting documentation and source information. For example, a query for contaminant chemistry results will provide not only analytical chemistry output, but also the ability to link to the “File Collections” with scanned field sheets describing the sample collection and habitat, photographs of the field collection, and a link to the CTD cast at the same location that collected fluorescence, dissolved oxygen and temperature profiles. An additional aspect of the File Collections approach is maintaining a complete record of all updates and changes. All files uploaded and information entered is maintained, with a record of updates, edits or replacements (and date/time/user name). This is an important aspect of records management satisfies litigation requirements for information stewardship and custody.

Query Tools

The common data model approach forms the foundation for flexible query tools that can query across all of the data models. A major difference between the data warehouse approach and a simple portal or metadata-level database is the ability to actually query data and results from across many studies with different data models. The common data model and warehouse approach of combining, transforming and standardizing datasets allows for data interrogation and exploration that would otherwise be impossible without significant and potentially duplicative efforts. While a metadata level record and data repository are better than hard-copies, scientists and researchers still face significant hurdles with analysis (and competing versions of datasets) unless a common and cooperative data warehouse is established.

Data View

A basic and significant tool that is part of the DIVER suite is the “user console” or data viewer. This tool, which is a Pentaho[18] user interface, provides direct access to the different common data models (e.g., samples, observations, photos, telemetry, and instruments) for review and querying. There are other industry standard tools that could be used to provide data management and technical staff access to the centrally-managed datasets for review and interrogation. This data view is generally used by technical staff for data exploration and QA/QC with the ability to easily group and filter large datasets in a centrally-held database. Users can save and share data views for review and collaboration. Although this interface contains standard tools for creating charts, generating simple mapping, and filtering, there is no capability for querying across different data models (schemas), and limited ability for incorporating metadata.

Reports and Dashboards

DIVER Reports and dashboards are geared towards providing environmental intelligence in the form of data status and summary. Reports and dashboards are typically created to meet recurring management or data manager questions, and to provide insight into data flow and progress.

Explorer

DIVER Explorer is a custom web-based query application developed by the OR&R. The data query tools within DIVER Explorer allow users to explore environmental datasets across all common data models with a query result that consists of an interactive map, legend, charts, and tables, FGDC metadata which includes data caveats and field definitions, and export capabilities. Users can choose query templates (“Guided Queries” similar to QM explanation queries), which are designed to provide quick access to results and the important details for analysis and interpretation. Alternatively, users can build a custom query by progressively choosing values which refine the query result. The query result also contains quick links to the file collections, which connect the data and results to source information including scanned field notes, data forms, and all updates and edits.

[18] Pentaho Data Integration: http://www.pentaho.com/

Appendix D: ARD Data Management Survey and Workshop

ARD Commonly Used Data Types and Sources

ARD Data Analysis Methods

ARD Staff Typical/Current Activities

Compiled Regional Feedback

ARD Tools Recommendations

ARD Scientific Questions

ARD Toolkit Uses

ARD Staff Typical/Current Activities

- **CERCLA**
    - **Sediments that have been contaminated by releases of CERCLA subatances**
    - **Removal and remedial**
- **OPA**
- Oil Spills
- Chemical Spills
- Ship Groundings
    - Groundings in Coral
    - Pacific Islands
    - Often related to or a potential oil spill
- Waste Sites
    - chronic
    - acute
    - remediation and toxicity
    - **Large hazardous waste sites**
        - **Risk and Injury Assessment**
        - **Restoration Opportunities**
        - **Contaminants of Concern**
        - **Contaminant Effects**
        - **Toxicity assessments**
        - **Tissue residue effects**
        - **Example Cases**
            - **Hudson River (NRDA and remedial)**
            - **Passaic**
            - **Berrys Creek**
            - **Newtown Creek**
            - **Cosco Busan**
    - Small and Medium Cases
        - Site assessment
            - Preliminary
            - Risk – Ecological Risk Assessment
            - Injury
            - Damage
        - Settlement Negotiations
        - requests for information and general outreach (public, academic, other govs, NGOs, co-trustees),
        - remedial & restoration monitoring
        - **Example Cases**
            - **Alcoa Grasse River remedial**
            - **Troy Chemical (remedial will be proposed to NPL Nov 2013)**
            - Raritan River watershed

ARD Staff Typical/Current Activities

            - Marathon Battery – marsh restoration under the remedy NL NRDA – initial effort could begin 2014
            - Restoration of Housatonic River in CT
            - Urban watershed restoration and planning
            - Portland Harbor
            - Regional Clean-up
            - Duwamish
            - Port Angeles Harbor
            -

- evaluation of natural resource injury for resolution of claims within global settlement context
    - assessment
        - pathway evaluation
        - injury determination
    - restoration project development, and scaling damages;
    - identify/develop restoration projects to compensate public for losses to navigation and transportation services
    - commercial and recreational fisheries, and supporting habitats
    - assist in developing Consent Decree and supporting documents
- Restoration
    - examples: dam removal, fish passage, habitat equivalency analysis (HEA)
    - habitat mapping/classification - gulf of mexico example for ranking/prioritization
    - monitoring and other data involved.
    - **Restoration projects database**
    - Lakes Restoration Initiative activites
        - OR&R both generates and manages data under this initiative
        - evaluate potential injury and to show status and trends
- Partnerships with EPA and states
    - coordinating, developing, and implementing multidisciplinary proposals and remedial and removal investigations
        - monitoring well placement/design/construction
            - sampling plans
            - health and safety plans
            - community relations plans,
        - contaminant extent and magnitude
        - contaminant fate and transport
        - risk assessments
        - cleanup criteria development
        - feasibility studies
        - proposed remedies
        - remedial and removal design plans
        - response action implementation
        - long-term monitoring, evaluation and oversight of mitigation and restoration activities necessary to resolve site related injuries
        - community relations

        - legal documents
    - Identify and evaluate cost-effective strategies to reduce injury to natural resources and sensitive ecosystems
        - negotiate with responsible parties
        - coordinate with EPA and co-trustees to maximize restoration of natural resources by integrating restoration into the remedial process.
- Planning
    - Arctic
        - NRDA
        - Oil Spill
- Technical support
    - NRDA
    - Remedial
    - Analytical Chemistry Database Systems
    - QA&QC
- **DWH**
    - **Case work**
    - Analytical data management
    - Analytical data management contract
    - Contracting – oversight – contract representatives
    - Have been interviewed by RC staff on restoration for DWH and monitoring data
- National projects such as development of models predicting sediment toxicity from sediment chemistry
- Contracting
    - General contract management
    - ARD General Watershed data management contract
    - contracting - oversight - contract representatives
    - statements of work for any projects needing contracts/grants
- General Administrative activities,
- Incident - oil or chemical spill or ship grounding
    - making maps -
        - trajectory - where is oil
        - operational
        - observations of oil in the environment (photos, field, SCAT, overflight, wildlife)
    - information management
    - display of information
- Incident - extreme weather event
    - hurricanes
    - marine debris
- Training and exercises
    - field sampling
    - tools and technology
    - drills and exercises

Questions to initiate the Scientific Method

- What questions do we need to ask from a higher level?
  - if you take all from cradle to grave – best way to capture is to conceptually walk through a case and see what types of questions we ask
  - What types of questions need to be asked from the following perspectives:
    - Engineering?
    - Cost?
    - Technical?
    - Legal?
  - There are a myriad of case types we deal with and while there are basic questions you don't really know until you know some characteristics of the case.
    - **Questions vary depending on response/ injury assessment phase, how much time you have, and whether using existing data or designing data collection, or a combination**
    - hard to know what is missing in laundry list – depends on where you are in the case – depends on case type –
  - The science questions I ask are very broad, since they are not for a specific case, and are intended to help focus our planning efforts. Some examples of broad science questions that I am looking at are:
    - **What are the exposure and injury endpoints for resources at risk of being impacted by an oil spill in the arctic that should/could be studied for NRDA in the event of a spill?**
      - ensuring response actions will control source and remove residual injury (ultimately, though may be initially risk-based),
      - Does sediment/tissue/water chemistry provide information that can be used in source identification?
      - Which of these endpoint are the highest priority (in terms of maximizing potential injury quantification and restoration) for ephemeral data collection?
      - What is the variability in these endpoints? and how can we achieve desired power to detect effects (comparing to baseline and/or comparison sites)?
  - Things that arise to drive the case are:
    - Legal questions arise
    - sometimes we evaluate, and analyze regs and statutes and present interpretations to DOJ in order to build a case
    - keep in mind to a large extent we are thrown into a regulatory science framework rather than a strict science
- Common Questions
  - How are we determining our strategy?
    - what is feasible within the response or SOL timeframe?
    - what is the appropriate scale; are response and restoration efforts achieving targeted goal--if not how can actions be modified

- What types of restoration are practicable to compensate for site-related injuries;
- what projects are available in the appropriate timeframe;

- Have NOAA trust resources been injured? (**Do we have evidence of exposure?)**
    - For how long
    - What potentially affect resources are present
    - **What is our injury to trust resources (or what is the ecological risk)?**
    - **How can we measure exposure to trust resources?**
    - **Have we exceeded identified thresholds?**
    - What concentrations would be protective of NOAA trust resources?
    - Do we have baseline information about the resources?
    - **What are the primary contaminant exposure pathways to resources of concern? What is the fate and effects of contaminants in the system of concern? What is the potential for adverse biological effects and measures of injury?**
- **What are the levels of toxicity that cause effects (acute and chronic) in various species**
- What information is needed to answer the questions at hand?
- What additional data is needed?
- Can data acquisition be combined with current or already planned data collection efforts in a defensible manner?
- How do I account for climate change in my long term planning?
- **What information do I need for sampling design, field sampling plan (temporal relevance, data quality objectives)**
    - Will data collected under a given sample plan support be representative of the parameter being measured?
    - will sample plan provide enough power to support testing conclusions?
    - are planned sample locations appropriate?
- **How do I collect and manage various field information (examples: chemistry, observations, non-analytical, physical, bile)**
- **What is my Spatial reference (datum, coordinate system)?**
- **Do I have metadata about information so it can be used for analysis (when collected & source)?**
- **What spatial analyses may help?- generalize from samples to larger area (interpolation)**
- What does they data look like on a map?
- Contamination
    - Are natural resources exposed to hazardous substances?
    - Can adverse effects to natural resources be measured as a result of exposure to hazardous substances?
    - **what level of contamination will be left behind by cleanup?**
    - **what concentration best represents sitewide contaminant concentrations?**

Questions to initiate the Scientific Method

- What type of contamination are we dealing with? (What contamination is present?)
- How do I account for background contamination and multiple stressors?
- What is the geographic extent of the contamination?
- Get samples if possible, or records of discharges.
- Can the contaminant(s) in the affected areas be linked to a source or sources?
- what samples are needed to prove pathway?
- Are there confounding contaminants / sources?
    - Get samples or records of other discharges
- **Can we get data to show resources have been exposed?**
- What are adverse effects that can be caused by contaminants for each resource, at what concentrations?
- Are levels of contamination high enough to cause injury based on literature toxicity values? (includes water or sediment concentrations, concentrations in prey, and potential to bioaccumulate or biomagnify contaminants)
- Can adverse effects be measured amid natural variability?
- Are there dead / injured organisms? gather; can we test them to show cause of death or injury?
- **What are the potential secondary adverse effects of the contamination (e.g. ecological interactions that affect organisms higher in food web). Can we scale impacts?**
- What is the categorical severity of contamination?
- Do measured variables (e.g. abundance, cover) within a given categorical severity class differ compared to baseline or control?
- **What background data is available for these endpoints of interest in Arctic resources?**
- What is the difference between impacted areas and control areas for a given measured variable?
    - What are appropriate and defensible quality control measures for sample collection that deviates from standard protocols?
- Have hazardous substances been released?
- Is the analyte/parameter list appropriate?
- What are/were the target receptors--are they appropriate?
- Are there Contaminated sediments?
    - **What do we estimate historic sediment, water, tissue data were for data gap years based on current data and estimates of sediment deposition rate, decay rates, changes in source conditions, etc. comparisions of remedial alternatives- how reasonable are the projections developed by a PRP or EPA?**
    - are sediment/tissue concentrations declining, remaining same, increasing? what is the filtering capacity of mussels in a given area?

            - **how best to evaluate uncertainties how to manage data? deal with missing data? what data is required to answer questions What is the nature and extent of contamination in sediment and biota?**
- o NRDA
    - one of the first questions should be – "is restoration even possible?
        - This is one of the criteria for moving forward.
        - What restoration can be done to compensate for injury?
            - How much?
            - For how long?
        - what types of restoration are we targeting?
            - What are the opportunities?
            - Have opportunities been identified?
            - How can we assess/monitor & scale? benefits? can they be implemented?
        - **how many acres are impacted/benefit gained by action? what are the trends in biomonitoring data?**
    - **what are the benefits of a proposed restoration project?**
    - **which resources/species best represent the site as a whole for quantifying injury?**
    - **how long will natural recovery take?**
            - **Is site risk/injury greater than natural variability**
            - **What is the estimated rate of recovery with or without remediation?**
    - exposure/pathway
        - sub-questions to identify exposure
        - documenting exposure –
            - commonly think will get from Response but that's not necessarily true.
            - quantify exposure
            - identifying resource(s) at risk
            - How are they getting impacted (physical pathway, dermal exposure).
        - Are there pathways that may result in exposure to natural resources?
        - What are the sublethal impacts, how can they best be measured and how do they translate to percent service loss.
            - o
    - injury
        - population?
        - information on resources being assessed? are they present?
        - Can we translate the data collected during the assessment into injury? These are all pretty standard / generic.
            - On small spill cases uncertainty estimates and the negotiation is as important as the data.
        - what is population trend?

- other biologic info on resource
- depends on resource
- size of event
- chronic waste site vs. spill
- assessing past and residual injury—
  - what are receptors, contaminant-residue effects concentrations, component sources--viz., contaminant fate and transport—
- physicochemical properties affecting bio-uptake, concentration, magnification—
- foodweb; contaminant extent and magnitude intersection with receptor home ranges and preferred habitats during various life stages;
- **what are the ecosystems services provided by a given habitat before and after an action?**
- TOC, lipid normalization; site-specific tox data--were controls appropriate--was the study design robust enough (pwr) to merit analyses; sed/soil/ water screening value comparisons—
  - again, how these are used and the questions that follow are often determined by available time/resources
  - decision tree will dictate the questions under the specific scenario
- **what is the PCB or dioxin congener pattern in fish vs water vs sediment vs soil? Has it changed over a spatial or temporal scale? how to predict fish whole body tissue concentrations from fillet concentrations or vice versa especially if don't have site specific lipid data?**

- Other Questions
  - How do DQOs for existing data align with current question set?
  - are GW wells constructed properly
  - DWH specific questions are a different question and not as related to daily ops.
  - **What is the spatially averaged surface/subsurface sediment concentration?**

Economic questions

- how do we value resources/affected resources those sorts of ideas – from surveys of people
- looking for metrics that would indicate losses –
- instead of abundance of cover – very case specific as to what approach – reveal preference, direct losses or if we try and value
- start off and think what component of human losses are measurable
  - human use survey
- do we need to collect data or is there existing data we can leverage

Questions to initiate the Scientific Method

Legal Questions that affect approach

- how do we allocate injury to various parties
- what does case law say about how we do things
- what level of certainty do we need to have, etc.

ARD Data Sources

Other case team members
EPA
Probable Responsible Party/ Responsible Party
Field work
Consultants
Query Manager
Values from literature
Google Map/Earth
Internet Search (Google)
Toxicity Search
Google Scholar
Difference between getting data from EPA project manager (ECOTOX)
USACE Tissue Chemistry DB
Specific external gov't databases that we go to
National Status and Trends (NOAA)
other consortiums (SFEI)
European Databases, particularly for international cases
American Petroleum Institute (API) database
Census
NPS Data
Cal spill watch – OSPR
Facebook page communicating out initial information.

- Get email with alert takes to Calif. Facebook page.
- Not numerical data but information (photos, maps)

CA state toxic site website "GeoTracker"– has links to sites and data;
Water Boards.

- Mainly document management (electronic data may mean PDF)
- copies of permits

Other data portals for NOAA?
CSC data explorer (charts from OCS)
Sanctuaries

- have own ways of managing data
- local sanctuaries have data collection/storage
    - (Gulf of Farallones) as beachwatch with historical data transcribed to digital.
    - Geospatial and queryable.
    - Use for baseline conditions before/after spill.

ESI maps used (product)
spot trackers (safety and can verify locations- accountability)

ARD Commonly Used Data Types & Sources

- **environmental-contaminant chemistry**
  - sediment
  - soil
  - tissue
  - oil
  - water (ground and surface)
  - toxicity data
    - Mortality
    - sub-lethal endpoints
    - genomics
    - DNA sequences
    - When large amounts of data are generated/collected
      - how we manage is important
  - evidence of presence of contaminants at site(s) being investigated
    - biota
    - other samples
    - analysis
    - Analysis / interpretation of investigations,
      - Necropsies
      - fingerprinting of samples
- bioassay/toxicity samples
- biomass (how much stuff falls on a plate, stuff from a tow)
- bio-diversity (whatever)
- bioassay/toxicity samples
- biomass (how much stuff falls on a plate, stuff from a tow)
- bio-diversity (whatever)
- habitat
  - substrate
  - grain size
  - vegetation
  - soil-sediment type
- contaminant tissue-residue effects
  - literature
  - site-specific
  - conventional parameters
- Supporting Chemistry
  - TOC
    - Size
    - weight
    - Age
    - Sex
    - Reproductive condition
  - Lipids
    - Size
    - weight

        - Age
        - Sex
        - Reproductive condition
- instrument
    - CTD
- evidence of release
    - physical samples
    - results of sample analyses
    - photos
- Response records such as SCAT reports
    - Categorical classification of oiling levels along shorelines
- Physical/Environmental parameters
    - Temperature
    - Wind
    - Storm surge
    - Tropical/hurricane
    - Wind
    - Radar
    - precip
    - CTO
    - CTD
    - Buoys
    - Fluorescents
    - Data captured under the water
    - dissolved oxygen
    - Organic Carbon (what is always there vs elevated)
    - currents
    - turbidity
    - UV light penetration
    - Salinity
    - Mined data from other sources (NOAA and others)
        - Weather (NWS)
        - Remote sensing
        - Tides
        - UV penetration
- FCA
- Topography
- Bathymetry
- flow velocities (seasonal and episodic)
- river stage
- Hydrography
- Nav channel—raster
- Shoreline
- ESI
- ESN
- land cover

ARD Commonly Used Data Types & Sources

- land use
- internal site boundary delineations
- stream barriers
- construction material/labor costs
    - engineering efficacy
    - cost efficiency
- case documents/references
- cost doc data to support strategy and recovery
- Data is high quality
    - contaminant data collected under DQOs w/ agency oversight
    - QCed goes into QM
    - site-specific bathy data collected under DQOs
    - layers from USGS/EPA/RPs under oversight
        - contour maps
    - state GIS shops
    - manufacturing and industrial processes with respect to uses of specific chemicals/compounds
    - **Sediment, water, tissue, Maybe study specific data**
        - depends on the case
    - Documents
        - Data sheets etc. gps data etc. toxicity data?
        - DWH it's maintained externally but not smaller cases.
        - key references? more KMS but related to you guys.
    - **sediment chemistry and sediment toxicity**
        - **(QM, remedial agencies, PRPs)**
        - **injury study results (e.g. fish tumors)**
        - **(remedial agencies, cotrustees, PRPs)**
        - **Environmental parameters when samples were collected**
        - **Concentration values over time and space**
    - **descriptions/maps of proposed restoration projects**
        - **(PRPs, local organizations)**
        - **Restoration opportunities**
    - **contaminant toxicity data**
        - **(scientific literature, site-specific studies)**
    - **<Chemistry data>**
    - **Physical, biological (e.g., habitat) data, etc**
        - **biological evaluations of resources**
            - **surveys of sites to enumerate organisms**
            - **condition of organisms including presence of contaminants on/in organisms**
        - acoustic
        - Animal Tags
            - Telemetry
            - Satellite
        - Distribution and abundance data for plants and animals likely to be impacted by a release.

        - Includes historical data and data collected specifically for a given case.
        - Categorical or continuous
- interactive maps
- Geospatial boundaries
    - Sanctuaries
    - fish advisories etc.
    - geopolitical and resource specific areas
- Effects levels
    - Threshold levels
    - TRVs, etc....
- **Data on resources at risk**
- **Baseline data about resources and chemical characterization**
- Baseline information about other potential sources of contamination or environmental stressors Information about methods
- modeling and data use in NRDA
    - This data comes from a wide range of sources and varies widely in its format, quality, relevance, geographic scope, etc.
- Photos
    - site photos
- Aerial imagery
    - Lidar
    - Satellite
    - Aerial photography
    - elevation surveys
    - other things not in a standard excel spreadsheets
- historical data
    - (looking at footprints from industry to data on past usage to any range of historical data – what was the use 30 years, 50, 100 years ago)
- human use
    - subsistence harvest data,
- Species Abundance data, baseline covers a huge spectrum of data – would take forever to list everything
- Lists of responsible parties
    - contact info
    - public meeting participants and contacts
    - legal precedence
    - digital files of public meetings
        - (for Fed register)
    - Mass media releases, tribal council meetings and consultations
- NGO watershed analyses and reports
- Remedial footprints/plan information EPA
    - might not always be in GIS
- Inundation analysis
- National surveys that we point to

        - MRIP (NMFS)
        - NSRE
        - national survey of fishing hunting, etc (FWS)
        - there are more but these are the big ones
    - National maintained data bases
        - EMAP
        - stated databases
        - fisheries independent databases
        - fisheries databases
        - Mussel databases
        - NST (national status and trends)
    - Lists of data used in similar cases in other regions (need better insight on this)
    - Lists of contacts for certain types of cases
    - Economist Data
        - Surveys
        - Human use
        - Models
        - HEA calculations
        - Recreational use data
            - lifeguard counts at a beach
            - car counters at beach entrance points or on the ground surveys we've done
            - photos
            - census demographics
            - American community survey on
                - Income
                - population, etc.
    - Administrative Data
        - Budgets (need software)
        - Time (need software other than web T&A to be effective)
        - Reimbursement tracking
        - Cost documentation
        - Contracting managment

ARD Data Management Toolkit & Primary Usage

Tools for Data Management

- Query Manager (online)
- Query Manager (desktop)
- Query Manager (database)
  - Visual FoxPro
- Excel Spreadsheet
- Microsoft Access
- Google Earth
- ArcMap
- Quantum GIS (QGIS)
- File Cabinet
- Stacks on my desk
- Computer Harddrive
- STATA
- GAUSS
- Garmin Mapsource
- Doodle Poll
- Visual Hea
- HEA Spreadsheets
- HEA spatial tool
- Wiki
- Google Docs
- Online Photologger
- Desktop Potologger
- GPS Photolink
- Trac

Tools for Data Analysis

- DIVER
- Query Manager (online)
- Query Manager (desktop)
- Statistical Analysis Software
  - SAS
  - STATA
  - S (contractors)
  - R (contractors)
  - SPSS
- Sample Design
- Interpolation
- Excel
- ERMA

ARD Data Management Toolkit & Primary Usage

- ArcMap
  - Spatial Analyst Extension
  - Geostatistical Analyst Extension
- Access
- Endnote

Tools for Data Visualization

- Google Earth
- Google Map
- ESRI ArcMap
- MARPLOT
- CSC Sea Level Rise Tool
- Graphic Software
- ArcGIS online
- AOOS
- Mussel watch
- CSC data displays - digital coast
- Sea level rise viewer
- Nautical chart viewer and download app
- Shorezone
- LOSDMS – LOSCO data management
- gulf spill restoration sites – data portal type for DWH
- Cal Spill Watch
- Responselink

ARD Data Analysis Methods

- HEA calculations
  - Mapping of HEA
  - Scaling
- Organizing and tagging of photos and videos
- QA/QC (bad data, duplicates)
- reporting out categorical data (habitat types, assessment/status of targets, response metrics)
- trending
- charts/graphs display (pie charts, line graph)
- Comparison of field collected data with controls
- Lab analysis of field samples
- Response curves LCx (20,30,40, etc)
- Fingerprinting of samples
- WOE--cleanup level derivation
- Mass-balance Modeling
- Plotting data – trends over time (e.g. marsh)
  - Historic distribution mapping
    - General Additive Model (GAM)
- visualizing meaningful numbers (maps, tables, intervals)
- Mapping
- spatial analysis
  - depending on data you draw a space and analyze within space
  - correlations (proximity, joining)
  - interpolations
  - digitization
  - visualization
  - geo-processing (buffers)
  - cartography
  - interpolations like kriging or area weighted averages
  - mashing layers into one spatial context
- ecological risk assessment
  - food web modeling
  - stock assessment
- Animal behavior
  - not an easy answer for examples
  - predator prey interaction after exposure to contaminant
  - avoidance behavior after exposure
  - burrowing behavior after exposure
- Statistics
- Population modeling (contractors)
  - looking at trends over time
  - survivability of species
- hot-spot analyses in support of removal
- Monte Carlo analysis

  - RP using (example - also others)
- Types of statistical analysis
  - parametric/ non-parametric t-test
  - ANOVA (parametric and non-parametric)
  - Linear Regression
  - Co-variant regressions
  - Multi-variant analysis
  - Cluster analysis (hot spot?)
  - Principle component analysis
  - Nearest Neighbor
  - Computing thresholds
  - Power analysis for sample design (statistical)
  - model fit statistics (eg root mean square error, etc)
  - Box and Whisker Plots
    - Mead
    - Standard deviations
  - cluster analysis approach
  - ordination techniques
  - various logit specifications
    - restricted to stats packages GAUSS or Matlab,
  - descriptive statistics
  - Distribution testing and looking at outliers (treating)
- summarizing literature (e.g. swim speeds, oxygen uptake)
- cost
  - for both response and restoration actions
- all appropriate QM queries exported to spreadsheet for additional analyses;
  - lots of data exploration using QM/Marplot to refine questions and subset data for detailed analyses in AV
  - QM-Direct used to conduct rapid on-the-fly analyses during meetings--no other tool has this capability.
- Decision analysis (not statistical) NRDA cases
- decision tree
  - the specific analyses conducted are dictated by project phase, available time

ARD Tools Needing Attention

DIVER

- I find I do not always follow the flow of steps for running queries correctly
  - it makes sense when I watch someone else do it
  - when I do it myself I either forget a step or go down a wrong pathway
- Queries aren't completely intuitive
- Freezes at times
- NOAAnrda.org – struggle to find things, I need to follow a link to get to certain places I can't navigate to – would like to have quick links on my desktop version of noaanrda, knew how to do it in classic and can't on modern and vice versa. Trustee vs trustee/rp access – I know where things are and how I should be able to get to them but I can't and its makes me crazy
- Difficult to use with Internet Explorer
- is more complex than how we previously screened data in QM.
  - You need a level of training to fully use DIVER appropriately.
  - There are tricks to query creation and flow of selection that can change

ERMA

- it makes sense when I watch someone else do it
- when I do it myself I either forget a step or go down a wrong pathway
- DWH has so much data it can be hard to navigate
- works well for certain basic tasks.
- Printing with proper placement and labels is tricky
- querying of layers could be more robust
  - however this may also be better accomplished in DIVER depending on user need.

Query Manager Desktop/On-Line

- requires you to look for things in a limited range of ways (e.g. only subsurface sediment chemistry; or must know study name)
  - doesn't always allow me to find what I'm looking for
  - even though it is there
- sometimes give me more than I'm looking for – bring into excel to compensate
- definitely someone needs to walk a new user through their first few experiences.
- QM-Direct, as the tool has not been updated for 10x, requires use of 3x, then operations in 10x for that functionality.
- desktop – if you shut down QM before Marplot – things get squirrely – long standing known issue
- there is currently not a way to select multiple user-selected bioassay endpoints,

- the option is either one endpoint or all endpoints.
- To work around this requires exporting all the endpoints into ArcMap and then turning layers off and on and eventually deleting the ones not needed.
  - It is a very time consuming work around.
- For tissue data queries, I can select the exact species I want to look at and am not required to only select only one or all of them.

Photologger Online/Desktop

- Don't use photos and videos as much as could because overwhelming to dig in
- perhaps not caught up with tools but typically everyone with camera uploads into folder structure
- Need to be diligent about tagging them for them to be useful. Hardest part.
- Processing and making useful is difficult. Quickly can generate many photos.
- FinBase application – takes a LONG time because need to go through all photos.
- Challenge – easy to take photos but not processed with keywords or photos not useful.
- Finding useful photos out of many bad ones is difficult. Might tag whole roll.
- Issues with GPS and camera alignment and don't really use
- also need a tutorial (if we are still planning to use this software-- maybe it is obsolete now that cell phone pictures can automatically record locations?)
  - I have Photologger on my computer but I am not sure how to get updated versions
  - mine crashes because it is not the latest version
  - I don't know how to connect it to see other people's pictures besides my own
- Need more/tagging – easier searching
  - If I know what photo I want I have to dump through google chrome so I don't have to go pic by pic

HEA

- quirky and I have to go to the developers for that often

ORR.db

- "bad" data has to be modified to be included in the database
  - Incorrect data
  - Incorrect spatial information
  - Incorrect column names
  - Missing data
  - Missing columns

ARD Tools Needing Attention

Areas without tools, need tools for efficiency

- Stratus did a survey back when and I had definite input on this – contaminant and study list for example
- Restoration planning tools I would like to use that aren't covered
  - an empirical tool from an Army Corps paper for sizing channels for restoration
  - tools looking at watershed /landscape restoration
  - connectivity of habitats and their value and scaling
- Tools could probably be developed in ArcGIS that would make it easier for us to do certain analysis ourselves such as changing projections, nothing else comes to mind at the moment
- better understand ARD is one group we have such different clean-up concentrations across the country
  - help to explain different number form say same type of clean-up in Seattle vs Boston
  - bring us together rather than all being in different directions in our clean-up agreements with EPA –
  - different background concentrations - this should be part of the knowledge base tool
- some sort of project management tool relative to the process of what are we really trying to accomplish here
  - surprised to be jumping right into data management rather than a higher level on making sure
- Observational data
  - fits in category of labor intensive process.
  - Field teams recording observations (unstructured)
  - trying to QA and manage difficult.

- Early on with DWH
  - the more free-form data entry = more difficult.
  - Helping with multi-select; checkbox; circle helps with QA
- Sample IDs –
  - barcoding (random number)
  - aliases (alternate IDs)

- Laboratories
  - don't use same system to manage inventory and flow of sample processing.
  - Different commercial info management systems (LIMS).
  - Some use excel.

ARD Tools Needing Attention

- Lack of standardization is difficult.

- Sample tracking-
  - COC's aren't necessarily trackable electronically.
  - Lab would like to get electronic COCs first (save effort and transcription errors). Can't avoid all issues though…
  - If going to trouble on COC's- having way to track better would be good.
- Photo process – not systematic tagging of the photos – photo dump approach
- Journal articles – being able to stay up on the state of the science is important and very frustrating when I can't get something I'm interested in
- Sharing and editing of documents for lit sensitive and non-sensitive cases – sometimes a ftp site and other times gets emails –
  - just because you get notification of ftp site data doesn't mean people actually access it –
  - Google docs is OK for small docs, but nervous for lit sensitive data –
  - if you don't have a direct link you can very rarely find it despite all searching – not sure what - searching Gmail is the same
- Sharing information, data, etc is not as smooth as I would like it to be – really important for a division spread across the country, time zones, etc.
- converting shoreline segments to aerial extent based on a stratified random sampling design that then uses a shoreline and buffer to depict and calculated aerial extent categorically.