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

Environmental Disasters Data Workshop

September 16-17

Ben Shorr
NOAA's Office of Response and Restoration

Overview

Discuss "Data Warehouse" approach

Actual framework and processes; flexible and scalable

Common Data Models

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

Data Query and Delivery

Requirements that drive development of data discovery, query, reporting and export tools

Data Warehouse and Business Intelligence

Data Warehousing concept and reality

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

NOAA | Office of Response and Restoration | Assessment and Restoration Division

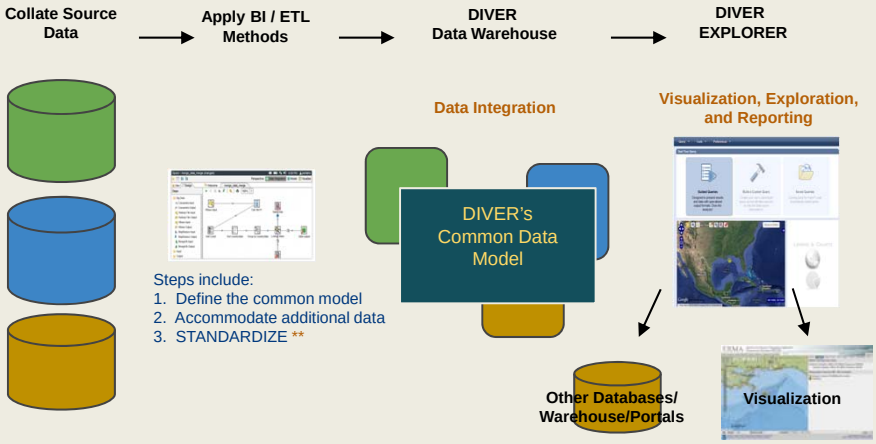
Data Warehouse and Business Intelligence

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

Data Warehouse and Business Intelligence

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

Data Warehouse and Business Intelligence



Our Approach: Promoting Common Data Models

Common Data Model Examples (schemas)

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

NOAA | Office of Response and Restoration | Assessment and Restoration Division

Common Data Model Examples (schemas)

- **Core Fields**

  Higher level across data models e.g. Analysis Type, Data Source, Status, Spatial

- **Data Specific**

  Results, Methodology, Units

- **Related Information**

  raw data, field information, source data packages, unstructured documents (reports, graphs, charts etc…)

Common Data Model: Contaminant Chemistry

Samples

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

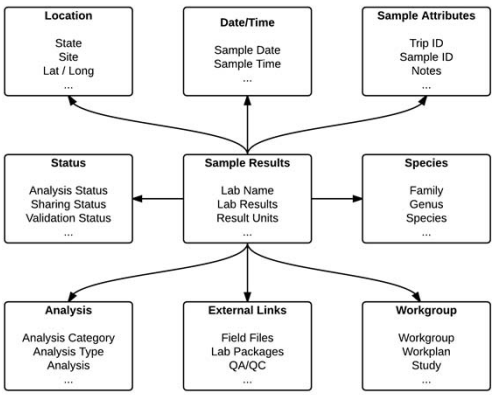
NOAA | Office of Response and Restoration | Assessment and Restoration Division

Common Data Model: Contaminant Chemistry

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

Common Data Model: Samples (Includes Contaminant Chemistry)

Conceptual View



Common Data Model: Samples (Includes Contaminant Chemistry)

Sample Fields			
Age (years)	Cruise Leg	Latin Name: Class	QM Reporting Standard
Analysis	Cruise Name	Latin Name: Family	QM Sample Details
Analysis Category	Data Category	Latin Name: Genus	QM Sample ID
Analysis Detail	Data Classification	Latin Name: Kingdom	QM Site ID
Analysis Method	Data Source	Latin Name: Order	QM Station ID
Analysis Result	Date	Latin Name: Phylum	QM Study ID
Analysis Result Unit	Day or Night Sample	Latin Name: Species	Qualifier Code
Analysis Status	Depth Category	Latin Name: Subphylum	Record ID
Analysis Type	Detection Extent	Length (cm)	Reporting Limit
Analyte Detection Details	Detection Limit	Link to Lab Data Files	Result
Case/Activity	DIVER Dataset	Location_Geom	Result (0.5 DL)
ChemCode	DV Qualifier Reason	LOSDMS Workplan ID	Result (0 DL)
Client ID	DV Qualifier Reason Code	Measurement Basis	Result (-1 * 0.5 DL)
COC ID	DV Qual Reason	MeasuresLevel	Result (-1 * Full ASA)
Collection Form	DV Qual Reason Code	Minutes of Hour	Result (-1 * Full DL)
Collection Matrix	End Latitude	Month	Result (Full DL)
Collection Method	End Longitude	Month (Numerical)	Result Notes
Collection Study Name	File Collection ID	Month Short	Result Type
Collection Workplan	Fingerprint Class	NRDA Grid	Review Status
Common Name: Class	Fingerprint Class Source	Number Below Detection Limit	Sample Delivery Group
Common Name: Family	Gulfspill Workplan Name	Number in Composite	Sample Depth Unit
Common Name: Genus	Habitat Type	Number Measured	Sample ID
Common Name: Kingdom	Hour of day	Oil Presence	Sample Lower Depth
Common Name: Order	Image Id	Oil Presence Screening Class	Sample Notes
Common Name: Phylum	Lab ID	Oil Presence Screening Source	Sample Size
Common Name: Species	Lab Name	Percent Lipid	Sample Size Units
Common Name: Subphylum	Lab Replicate	Photo URL - Midsize	Sample Type
Composite/Part Sample	Lab/Result Matrix	Photo URL - Original	Sample Upper Depth
Composite Sample ID	Lab/Result Matrix Detailed	Photo URL - Thumbnail	Sex
	Last Update Date	QM Matrix	Sharing Status
			Source Type
			Species
			Start Latitude
			Start Longitude
			Station/Site
			Survey Notes
			Tissue Code
			Tissue Type
			Total Organic Carbon (pct)
			Trip End Date
			Trip ID
			Trip Start Date
			Validation Level
			Weight (g)
			Workgroup
			Year

...Just a few favorite fields

DIVER (Data Integration Visualization Exporting and Reporting): Explorer Tool

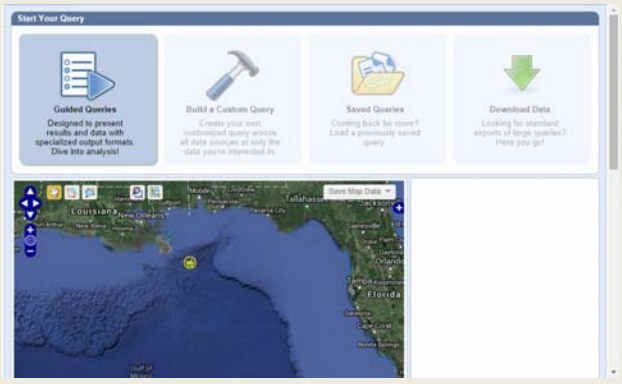
- Objectives & Requirements
  - Flexible query and export of all data including NRDA collected and external datasets
  - Documented lineage and connections to data holdings
  - Metadata, Metadata, Metadata
  - Export for analysis, visualization and processing

13

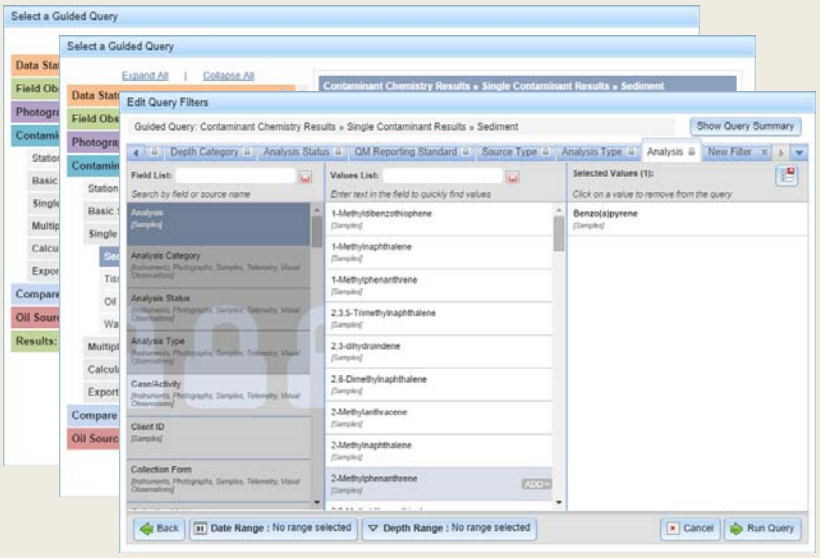
NOAA | Office of Response and Restoration | Assessment and Restoration Division

DIVER (Data Integration Visualization Exporting and Reporting): Explorer Tool

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

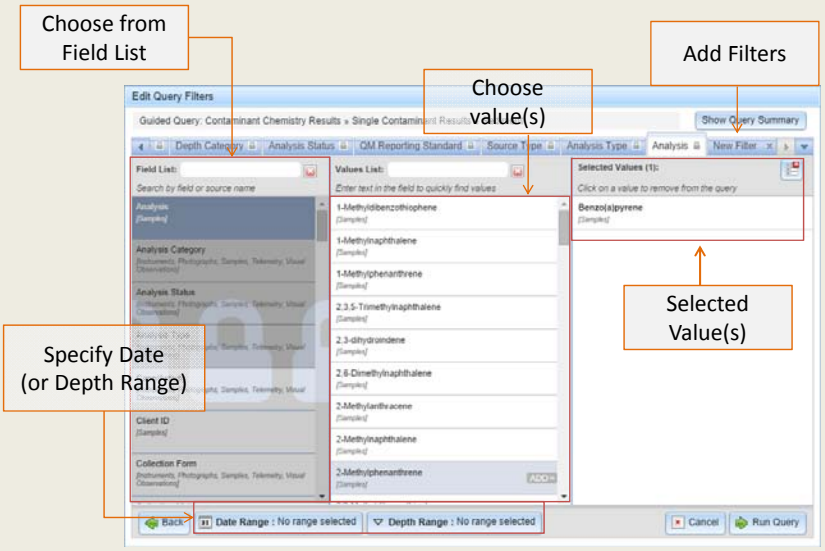


Guided Queries



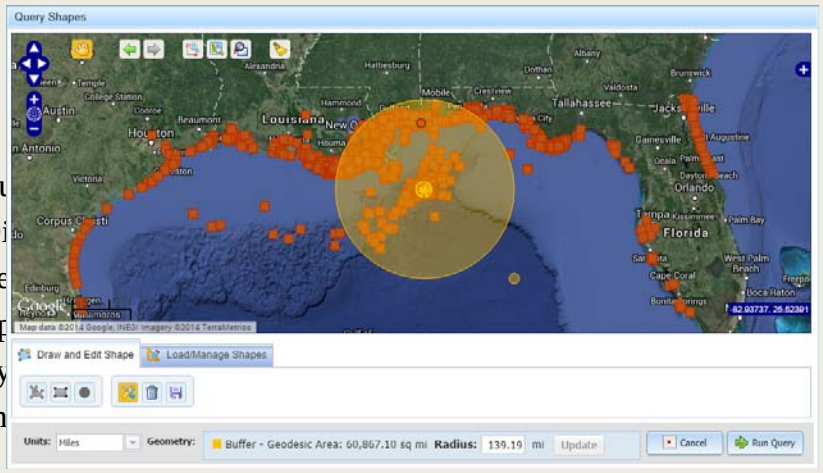
15

Query Filters



Requ

- Di
- Se
- Sp
- Sy
- In



*NOAA's Environmental Response Management Application

17

NOAA | Office of Response and Restoration | Assessment and Restoration Division

Query Results: Summary

18

Query Results (Table)

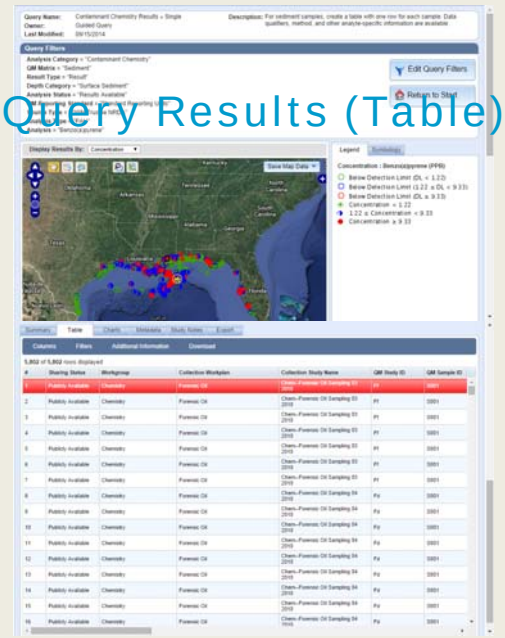
Requirements and Functionality:

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

19

NOAA | Office of Response and Restoration | Assessment and Restoration Division

Query Results (Table)



20

Charts

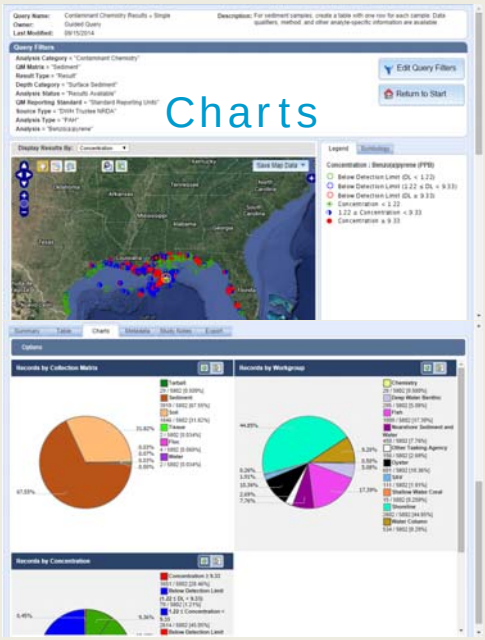
Requirements and Functionality:

- Provide overview summary of query results
- Interactive - click on charts to show filtered data
- Flexible - built to handle new information

21

NOAA | Office of Response and Restoration | Assessment and Restoration Division

Charts



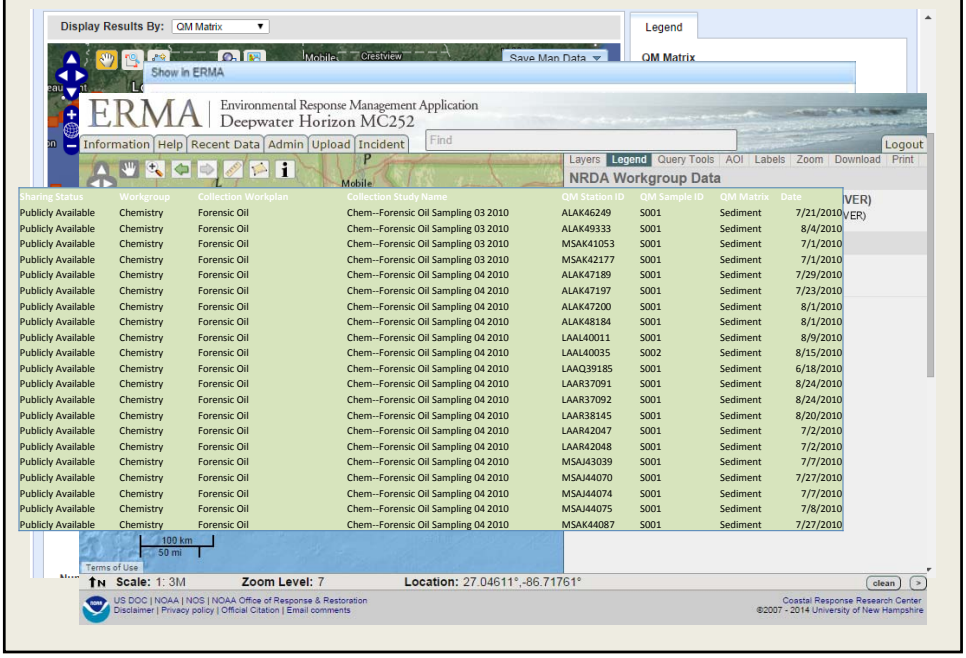
22

Export

Requirements and Functionality:

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

NOAA | Office of Response and Restoration | Assessment and Restoration Division



Sharing Status	Workgroup	Collection Workplan	Collection Study Name	OM Station ID	OM Sample ID	OM Matrix	Date
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 03 2010	ALAK46249	S001	Sediment	7/21/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 03 2010	ALAK49333	S001	Sediment	8/4/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 03 2010	MSAK41053	S001	Sediment	7/1/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 03 2010	MSAK42177	S001	Sediment	7/1/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	ALAK47189	S001	Sediment	7/29/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	ALAK47197	S001	Sediment	7/23/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	ALAK47200	S001	Sediment	8/1/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	ALAK48184	S001	Sediment	8/1/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	LAAL40011	S001	Sediment	8/9/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	LAAL40035	S002	Sediment	8/15/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	LAAQ39185	S001	Sediment	6/18/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	LAAR37091	S001	Sediment	8/24/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	LAAR37092	S001	Sediment	8/24/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	LAAR38145	S001	Sediment	8/20/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	LAAR42047	S001	Sediment	7/2/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	LAAR42048	S001	Sediment	7/2/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	MSAJ43039	S001	Sediment	7/7/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	MSAJ44070	S001	Sediment	7/27/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	MSAJ44074	S001	Sediment	7/7/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	MSAJ44075	S001	Sediment	7/8/2010
Publicly Available	Chemistry	Forensic Oil	Chem--Forensic Oil Sampling 04 2010	MSAK44087	S001	Sediment	7/27/2010

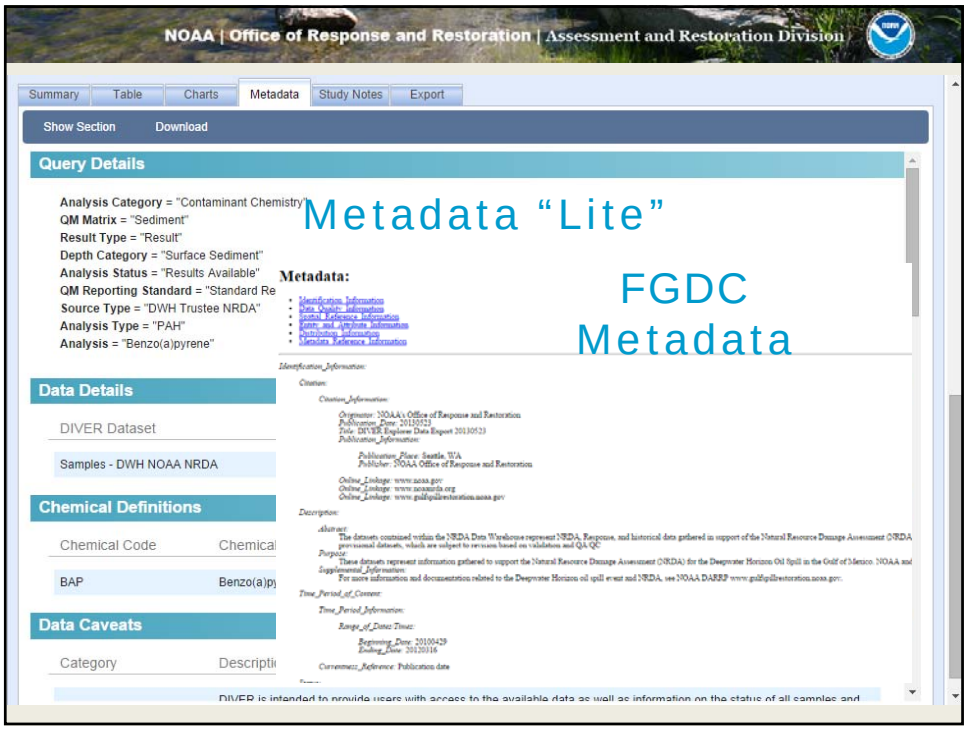
Metadata

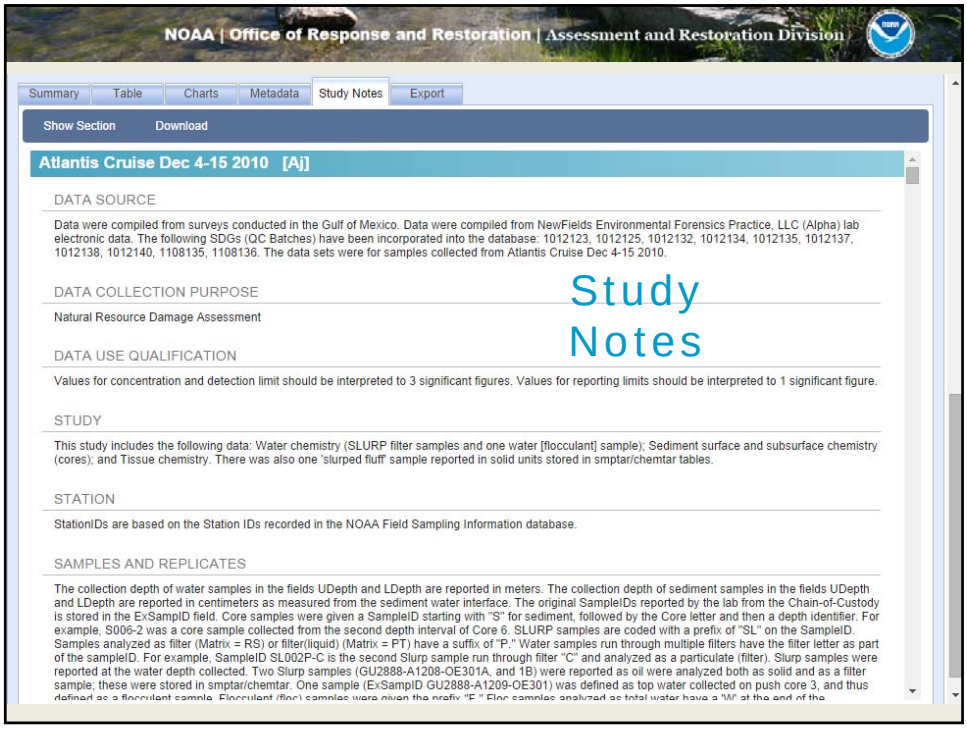
Requirements and Functionality:

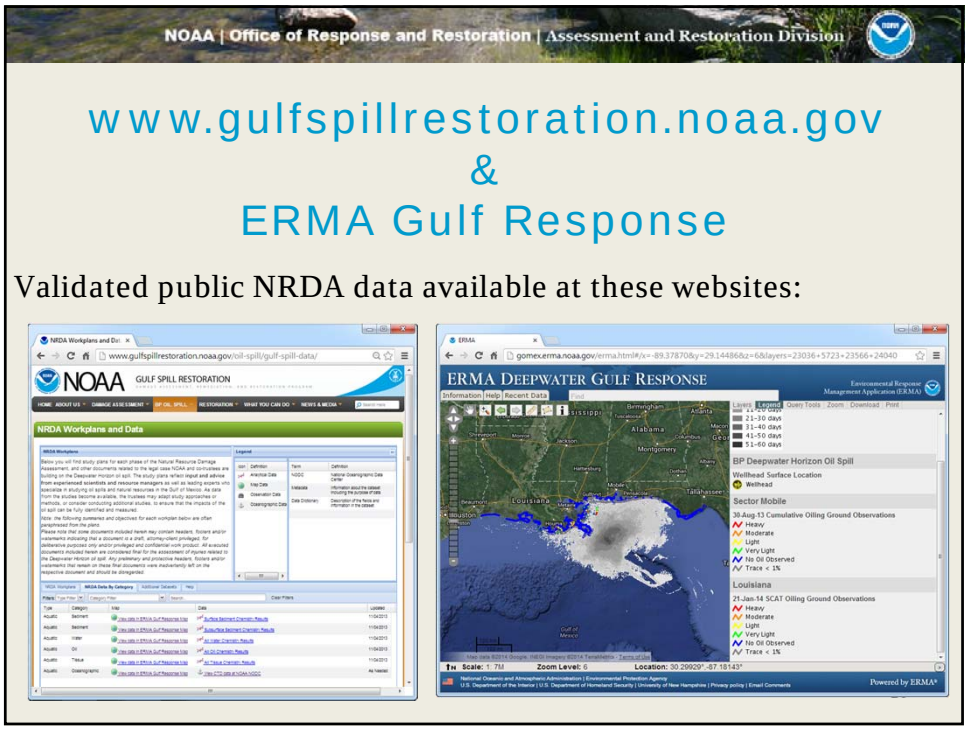
- "Lite" version with key information (HTML)
- FGDC compliant metadata (XML and HTML)
  - moving to ISO 19115
- Query Details: fields and values chosen
- Data Details: when were datasets updated?
- Data Caveats: notes about data
- Field Definitions

25

NOAA | Office of Response and Restoration | Assessment and Restoration Division




NOAA | Office of Response and Restoration | Assessment and Restoration Division
Summary
Table
Charts
Metadata
Study Notes
Export
Show Section
Download
Atlantis Cruise Dec 4-15 2010 [Aj]
DATA SOURCE
Data were compiled from surveys conducted in the Gulf of Mexico. Data were compiled from NewFields Environmental Forensics Practice, LLC (Alpha) lab electronic data. The following SDGs (QC Batches) have been incorporated into the database: 1012123, 1012125, 1012132, 1012134, 1012135, 1012137, 1012138, 1012140, 1108135, 1108136. The data sets were for samples collected from Atlantis Cruise Dec 4-15 2010.
DATA COLLECTION PURPOSE
Natural Resource Damage Assessment
DATA USE QUALIFICATION
Values for concentration and detection limit should be interpreted to 3 significant figures. Values for reporting limits should be interpreted to 1 significant figure.
STUDY
This study includes the following data: Water chemistry (SLURP filter samples and one water [flocculant] sample); Sediment surface and subsurface chemistry (cores); and Tissue chemistry. There was also one 'slurped fluff' sample reported in solid units stored in smptar/chemtar tables.
STATION
StationIDs are based on the Station IDs recorded in the NOAA Field Sampling Information database.
SAMPLES AND REPLICATES
The collection depth of water samples in the fields UDepth and LDepth are reported in meters. The collection depth of sediment samples in the fields UDepth and LDepth are reported in centimeters as measured from the sediment water interface. The original SampleIDs reported by the lab from the Chain-of-Custody is stored in the ExSampID field. Core samples were given a SampleID starting with "S" for sediment, followed by the Core letter and then a depth identifier. For example, S006-2 was a core sample collected from the second depth interval of Core 6. SLURP samples are coded with a prefix of "SL" on the SampleID. Samples analyzed as filter (Matrix = RS) or filter(liquid) (Matrix = PT) have a suffix of "P." Water samples run through multiple filters have the filter letter as part of the sampleID. For example, SampleID SL002P-C is the second Slurp sample run through filter "C" and analyzed as a particulate (filter). Slurp samples were reported at the water depth collected. Two Slurp samples (GU2888-A1208-OE301A, and 1B) were reported as oil were analyzed both as solid and as a filter sample; these were stored in smptar/chemtar. One sample (ExSampID GU2888-A1209-OE301) was defined as top water collected on push core 3, and thus
Study Notes


NOAA | Office of Response and Restoration | Assessment and Restoration Division
www.gulfspillrestoration.noaa.gov
&
ERMA Gulf Response
Validated public NRDA data available at these websites:
NOAA GULF SPILL RESTORATION
NRDA Workplans and Data
ERMA DEEPWATER GULF RESPONSE
BP Deepwater Horizon Oil Spill
Sector Mobile
Louisiana

Next Steps & Challenges

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

Internet Security

Discussion? Happy Hour?

29

NOAA | Office of Response and Restoration | Assessment and Restoration Division

Acknowledgements

DIVER Data Management Team: Mike Jackson, Dan Hudgens, Jim Anderton, Ann Jones, Amy Merten, Ben Shorr, Kevin Kirsch

NOAA: Amy Merten, Kevin Kirsch, Jay Coady

IEc: Kate Doiron, Ann Jones, Jess Fydenkevez, Lena Flannery, Neal Etre, Amy Anderton

Sirius Solutions: Vincent Luzzo, Nicole Williams, Brian Thompson