



A Data Warehouse Approach To Managing and Querying Environmental Data

Ben Shorr

Assessment & Restoration Division | Spatial Data Branch

NOAA | National Ocean Service | Office of Response and Restoration

National Environmental Monitoring Conference 2015
Chicago, IL



The information in this presentation reflects the views of the author, and does not necessarily reflect the official positions or policies of NOAA or the Department of Commerce.



Presentation Overview

Background on NRDA and Data Sources

Data sources, standards and needs

Data Warehouse and Common Data Models

Overview of flexible/scalable framework; Data Models and standards; related information/data. Information Management Portal and DIVER

Data Query and Delivery

Requirements that drive development of data discovery, query, reporting and export tools- supporting scientific analysis and reports for a Damage Assessment case



Natural Resource Damage Assessment (NRDA)

Preassessment Screen (exposure assessment)

Restoration Planning

- Field Studies
- Data Evaluation
- Modeling
- Injury
- Quantification

Restoration Implementation





Marsh Assessment



Shoreline Data



Toxicity Data



Oyster Collections



Water Column



Telemetry Data



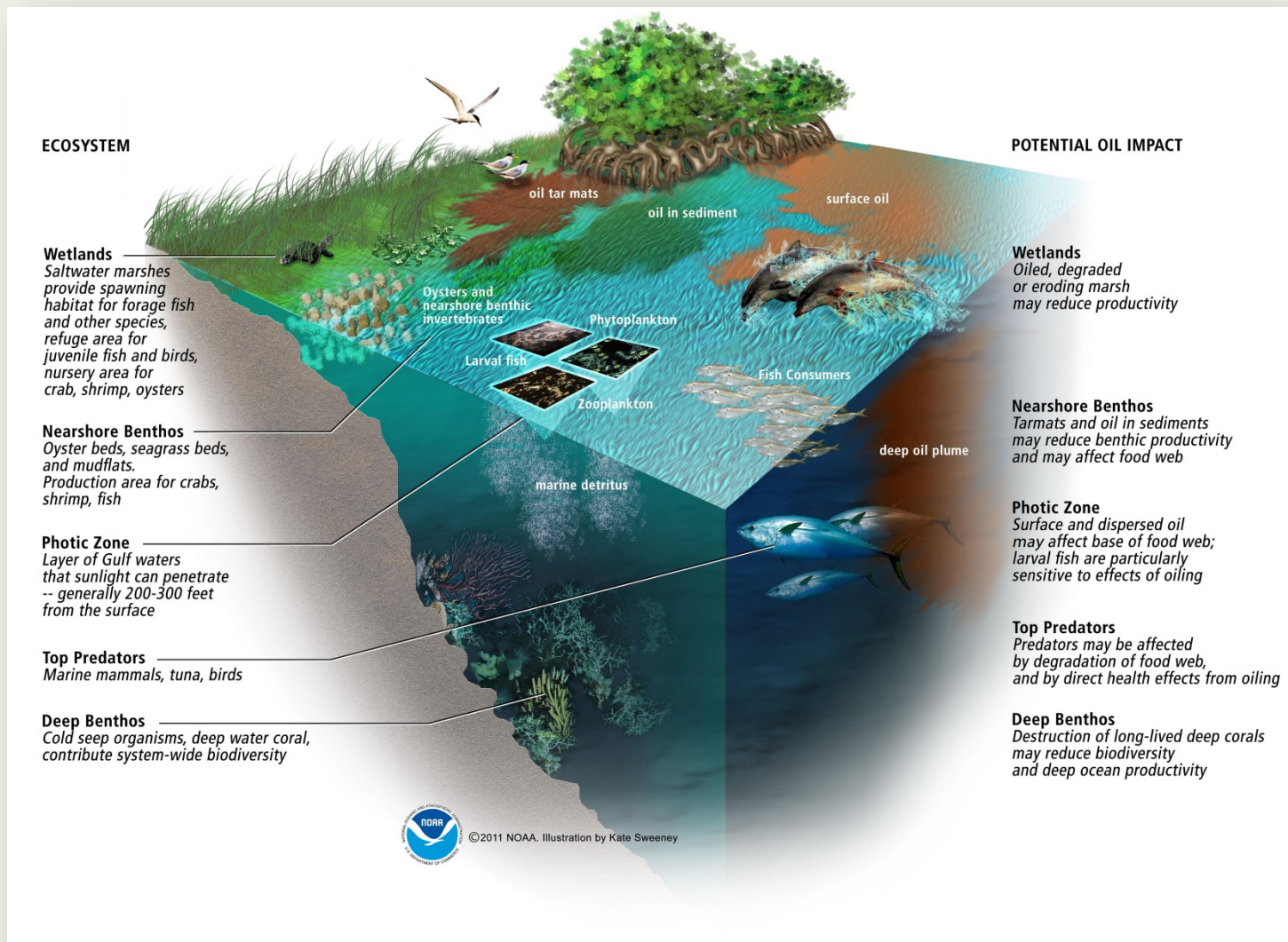
Seafood Safety



Marine Mammal & Turtle Assessment



Ecosystem Overview





Natural Resource Damage Assessment Data

- Vast amount of NRDA and Response data collected under different authorities, different formats, different destinations and management
- NOAA OR&R and partners were part of key NRDA and Response data streams early and adopted/adapted/created data flow processes.

File Collections (aka NOAANRDA website)

[illegible]

1.5 years into NRDA case

Preliminary Measures and Dimensions; Priority of "Questions to Answer"; Data Sources that can be used to Answer Questions																
Priority of Measure Implementation	Measures (The Question to Answer)	Dimensions (Ways to slice the Question)													Current Data Warehouse Pulling Data From:	
		Time	Spatial	Depth	Sample Type	Habitat	Site	Study	Workplan/Method	Instrument Type	Oiling	Species	Hypothesis	Status		Lab
1	Contaminant Lab Results	X	X	X	X	X	X	X	X		X	X	X	X	X	Validated EDD, QM
2	Observation Data	X	X	X	X	X	X	X	X		X	X	X	X		nn.org
3	Additional Lab Data Results	X	X	X	X	X	X	X	X		X	X	X	X	X	nn.org, 50+ labs
4	Response Activities Count and Duration	X	X	X		X	X		X					X		Spatial Data Team
5	Species Count	X	X	X		X	X	X	X		X		X			Observation (nn.org), Telemetry
6	Instrument Results	X	X	X				X	X	X		X	X	X	X	Multiple: NODC, Source, Database per Instrument (Currently does not exist)
	TBD:															
	- Photographs (Photologger)															
	- Video Clips (TBD-"Kaltura?")															
	- Acoustic Clips (TBD-"Kaltura?")															



Data Warehouse Approach

Ingest Data

- Bring in data from different sources; flexible and scalable
- Adopt or adapt existing standards; develop and document new standards
- Manage **structured** and **unstructured** data/information
- Litigation quality
- Documented processes

Relate Information

- Examples: samples and observations; field data and photographs; laboratory packages



DIVER Overview

Data Integration, Visualization, Exploration and Reporting

DIVER is a **data warehouse** and **query** application developed by NOAA. The DIVER approach **integrates** and **standardizes** datasets so users can query across data holdings and download information and results.



Data Deliverables

Laboratory data example

- Full Laboratory packages
- Validation packages
- Either receive in template (Electronic Data Deliverable) specified by QAPP or coordinate with laboratory
- Sample cross-walk may be necessary to resolve Field/Chain of Custody/Laboratory identifiers

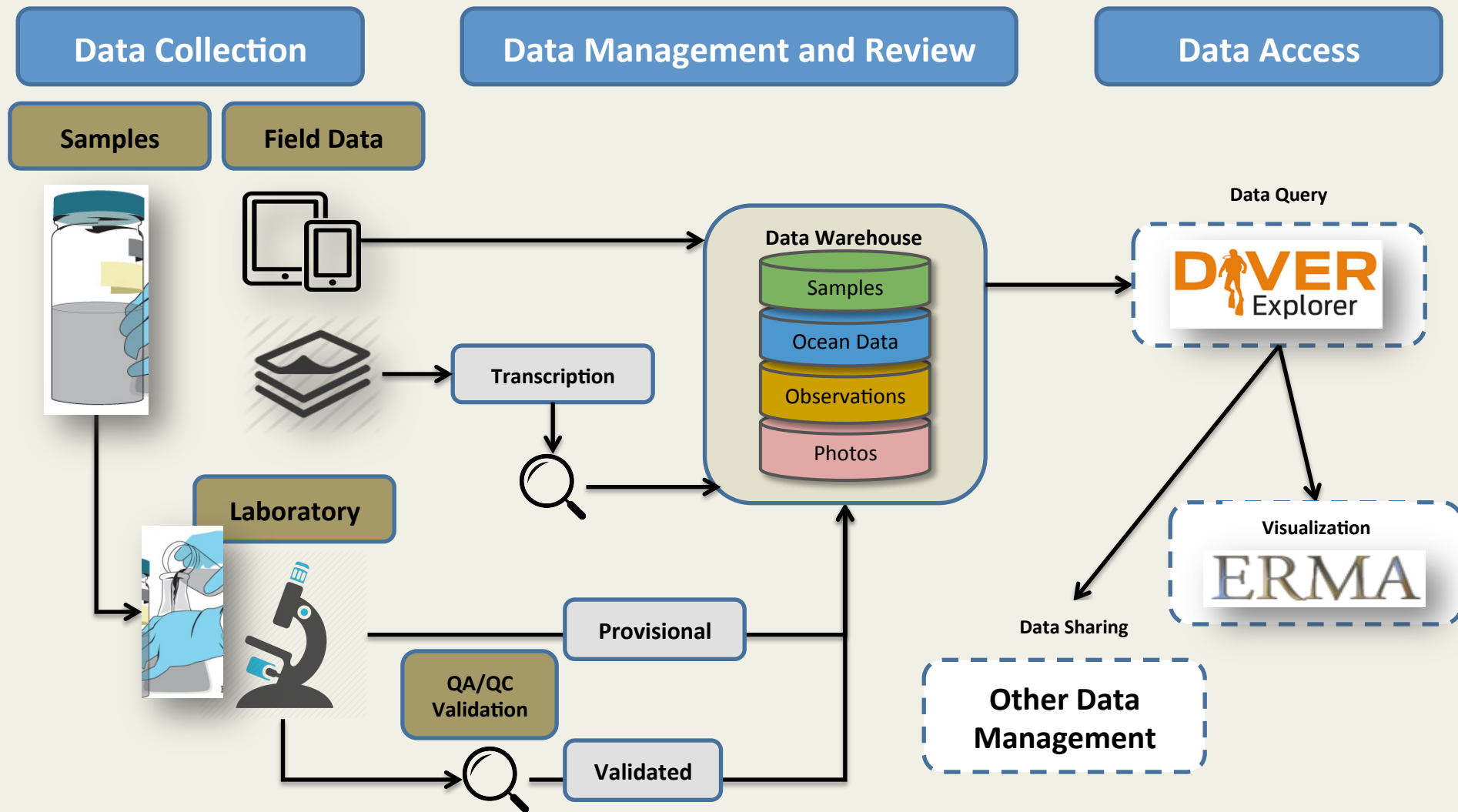


Common Data Model Examples (schemas)

- **Samples:** chemistry (NRDA, Response-EPA Scribe, historical), biological data, forensics...
- **Oceanographic:** cruise-collected sensor data
- **Observations:** shoreline, marsh, birds and mammals
- **Telemetry:** location tracking devices
- **Photographs:** keywords, location
- **Restoration data:** potential and implemented projects

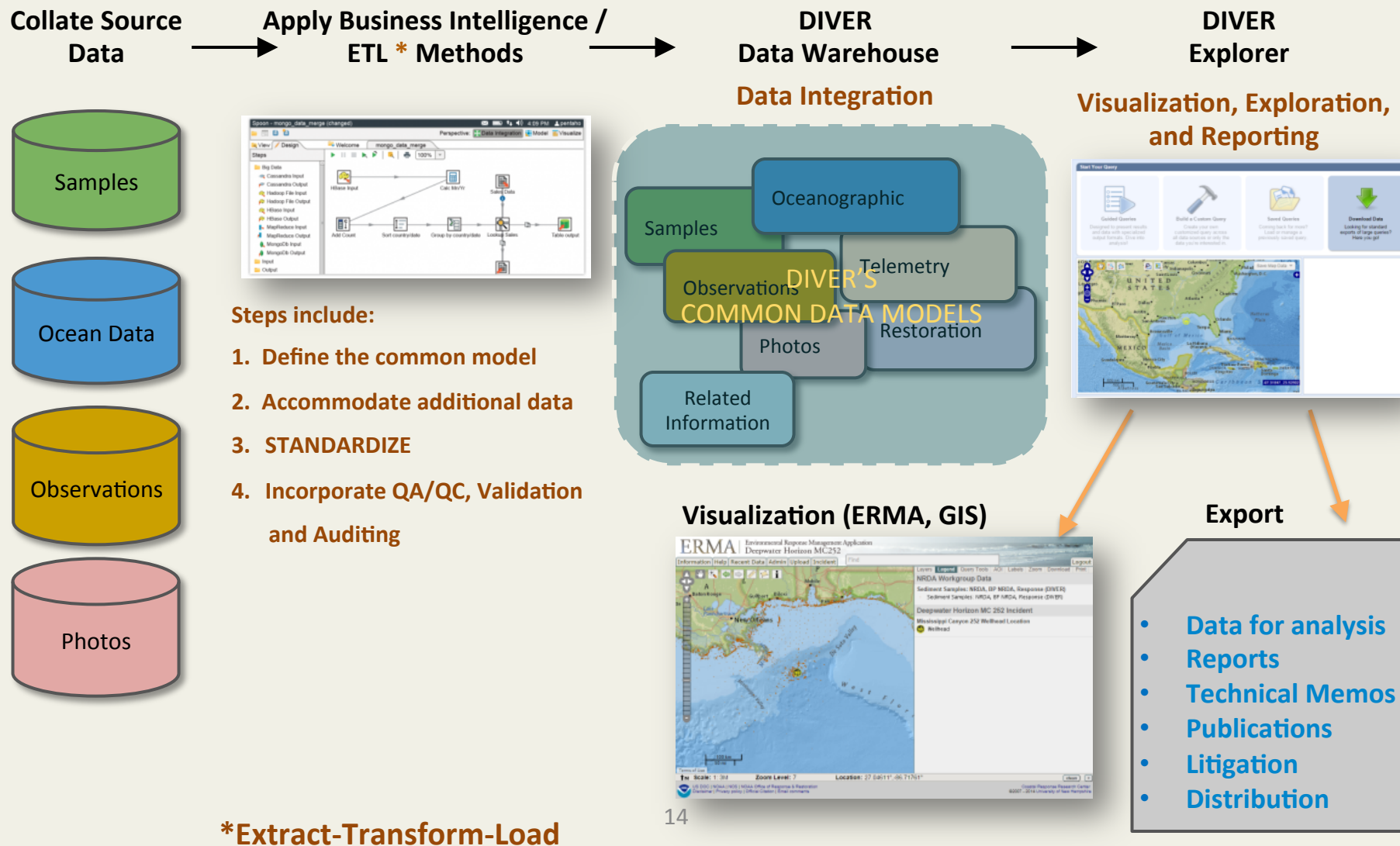


Simplified Data Flow: Samples





Data Warehouse and Standardization





DIVER: Information Management Portal

- Centralized location for case/incident information management
 - Evolving to meet needs
 - Uses industry-standard tools including Open Source Liferay Portal Web Platform
 - Liferay Portal:
<http://www.liferay.com/products/liferay-portal/overview>
 - Also used by NWS VLab (Virtual Lab Development):
<https://vlab.ncep.noaa.gov/>. We are talking monthly.



DIVER: Information Management Portal

Deepwater Horizon
Information Management Portal

Welcome, Benjamin Shorr | [Change Password](#) | [Sign Out](#) | [Feedback/Questions](#)

Home Communities Find Data and Documents Contribute Data and Documents News Help

Information Management Portal » Home

Quick Links

- DIVER Explorer**
Query and export available data and metadata. Interact with the data through maps, charts and tables.
- DIVER Reports + Dashboards**
Generate on-screen views of the data in chart or table form based on pre-set queries.
- Resource Catalog**
Access documents and information-sharing pages. To access Workplan information, visit the [Workplans List](#) page.
- Help Documents**
Find documents that provide assistance for DIVER and the DWH Information Management Portal users.
- Litigation Hold Training**
Access training materials, PowerPoint Slide Show, and Staff and Contractor Survey.

Announcements

All Data Incident System

Data Updates - Summary Report [Mark as Read](#)

Trustee All Jan 28, 2015

DIVER Explorer updates made on January 27, 2015 include new Contaminant Chemistry results from Query Manager, and revisions made to the following Workplans:

- Fish, Blue Crab *Megalopae* Plan: 9 samples with new results.
- Nearshore Sediment and Water, Nearshore Sediment and Water Preassessment Sampling for Louisiana: 32 samples with new results, one sample had updates made to the upper and lower depth values.
- SAV, Tier 2 Workplan: 6 samples with prior results had qualifier code values updated.
- SAV, Tier 3 Workplan: 15 samples with prior results had qualifier code values updated.

DIVER Data Updates - Summary Report [Mark as Read](#)

Trustee All Jan 9, 2015

DIVER Explorer updates made on January 8, 2015 include changes made to the following Workplans:

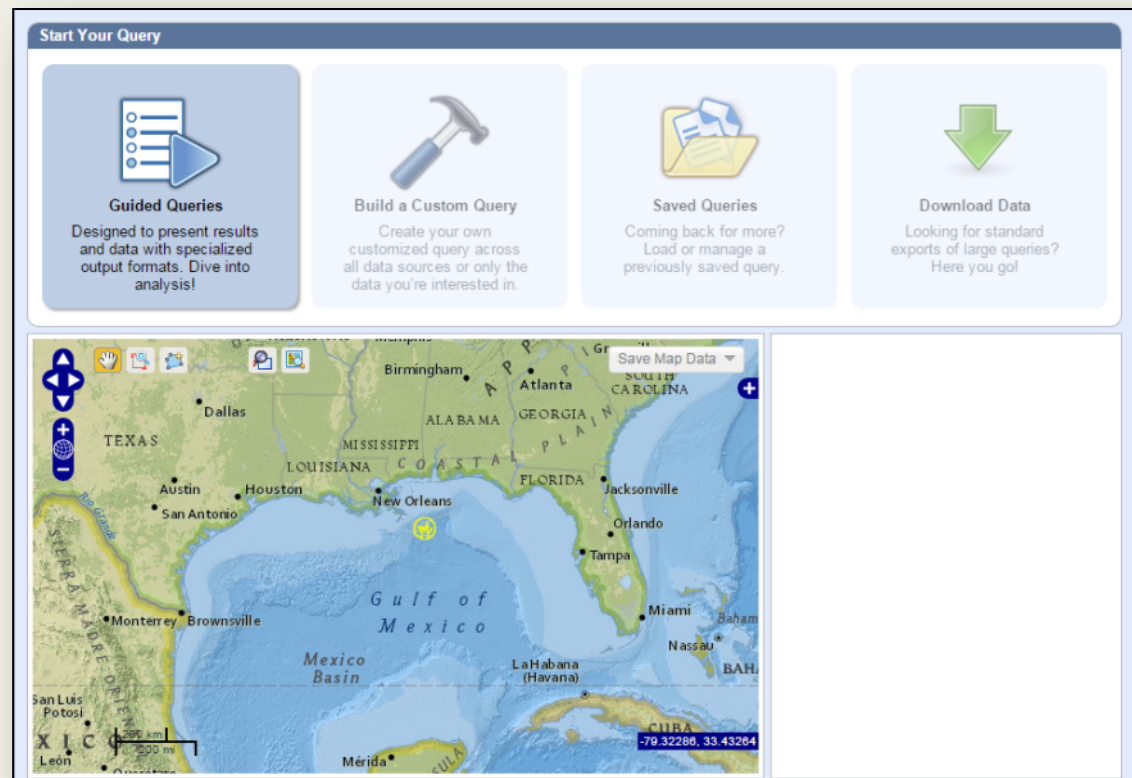
- Marine Mammals and Turtles, Dolphin Health Assessment 2013: 164 samples with new results.
- Oyster, Oyster Sampling Plan: 1 sample with new results.

- “Wiki”-type communities for team Collaboration
- Communicate Updates and Schedule
- Data Submittal (laboratories, analysis, processed)
- Data Query and Reporting Tools
- Help and Training



DIVER Explorer Application

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





DIVER Explorer: Guided Queries

Select a Guided Query

Select a Guided Query

Expand All Collapse All

Edit Query Filters

Guided Query: Contaminant Chemistry Results » Single Contaminant Results » Sediment Show Query Summary

Depth Category Analysis Status QM Reporting Standard Source Type Analysis Type Analysis New Filter x ▶ ▼

Field List:	Values List:	Selected Values (1):
Search by field or source name	Enter text in the field to quickly find values	Click on a value to remove from the query
Analysis [Samples]	4/9-Methylphenanthrene [Samples]	Benzo(a)pyrene [Samples]
Analysis Category [Instruments, Photographs, Samples, Telemetry, Visual Observations]	9-Methylphenanthrene [Samples]	
Analysis Status [Instruments, Photographs, Samples, Telemetry, Visual Observations]	Acenaphthene [Samples]	
Analysis Type [Instruments, Photographs, Samples, Telemetry, Visual Observations]	Acenaphthylene [Samples]	
Case/Activity [Instruments, Photographs, Samples, Telemetry, Visual Observations]	Anthracene [Samples]	
Client ID [Samples]	Benzo(a)anthracene [Samples]	
Collection Form [Instruments, Photographs, Samples, Telemetry, Visual Observations]	Benzo(a)fluoranthene [Samples]	
	Benzo(a)pyrene [Samples] ✓	
	Benzo(b)fluoranthene [Samples]	

Back Date Range : No range selected Depth Range : No range selected Cancel Run Query



DIVER Explorer: Dashboard Approach

Data Summary

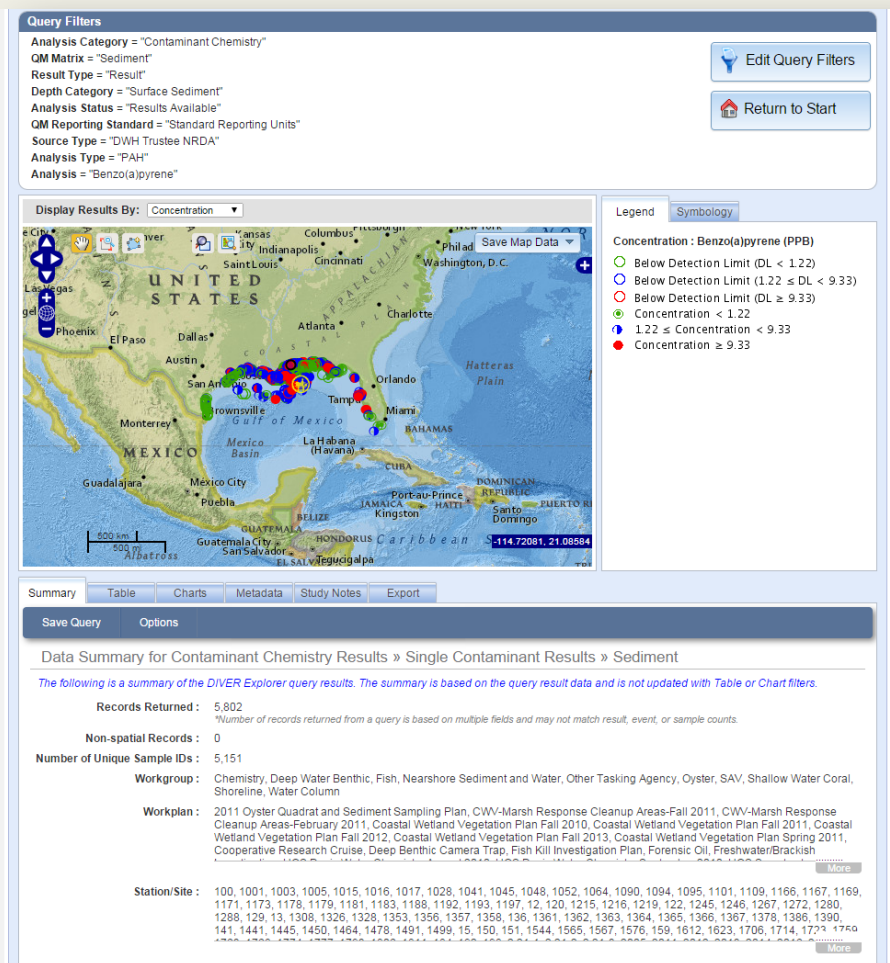
Data Table

Charts

Metadata

Study Notes

Export





DIVER Explorer: Dashboard Approach

Data Summary

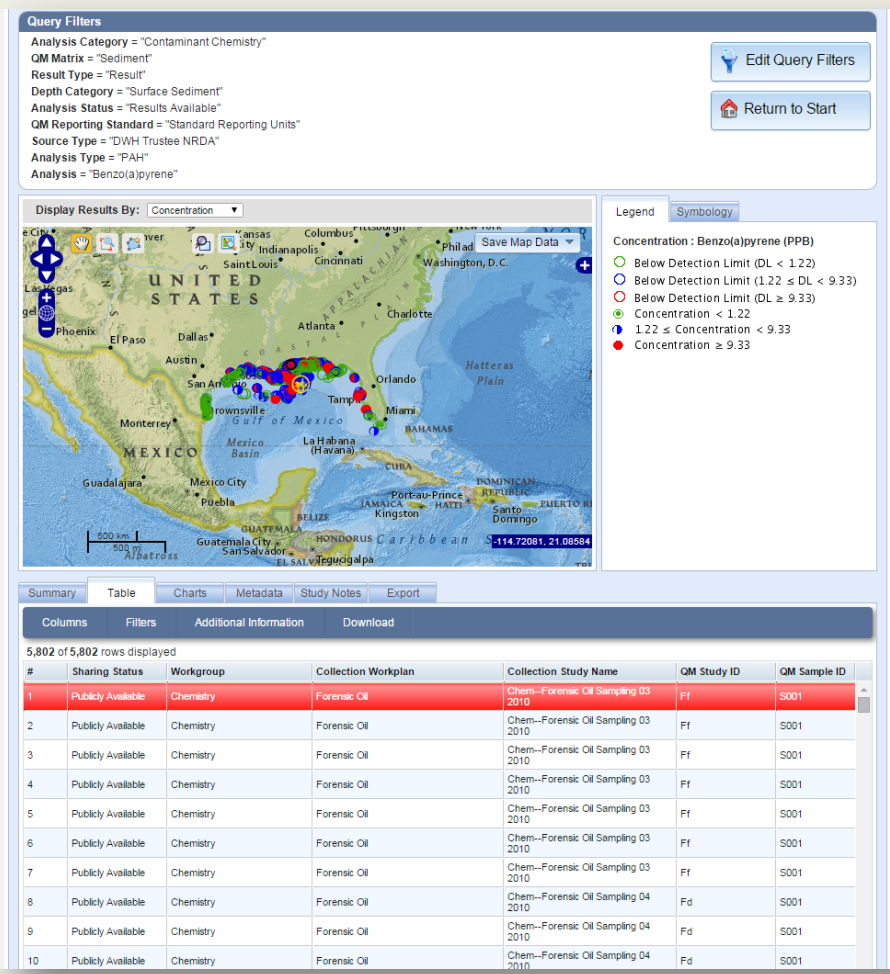
Data Table

Charts

Metadata

Study Notes

Export





DIVER Explorer: Dashboard Approach

Data Summary

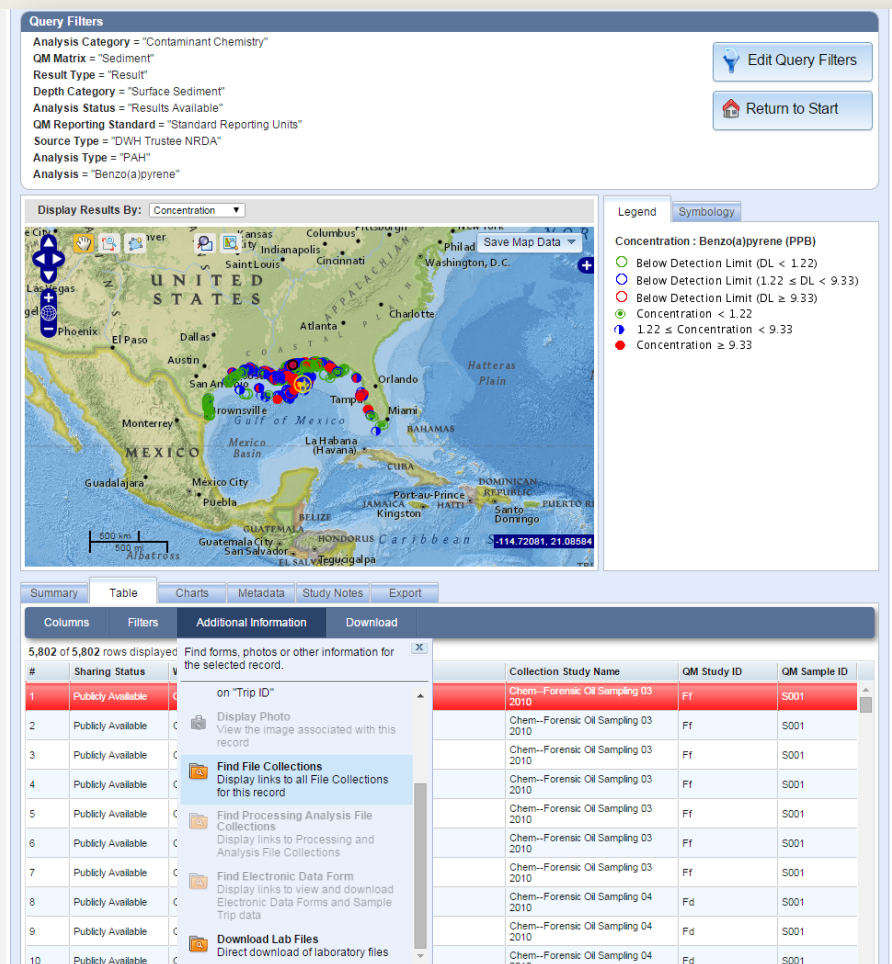
Data Table

Charts

Metadata

Study Notes

Export





DIVER Explorer: Dashboard Approach

Date Summary

D

N

St

Export

Query Filters
Analysis Category = "Contaminant Chemistry"
QM Matrix = "Sediment"
[Edit Query Filters](#)

File Collections for Sample Trip ID 1388

The following file collections are available for sample trip id 1388.

Source	Trip Date	Status	List of files	Cruise	Leg
Samples	10-30-2010	Inserted FSF database	2010_1030_Richardson_Alec_FieldCollectionForms_DM.xls_QAQC.rtf 13.1 KB 2010_1031_Richardson_Alec_FieldCollectionForms_DM1.xls 937.0 KB 2010_1031_Richardson_Alec_Logbook_DM1.pdf 323.5 KB 2010_1031_Richardson_Alec_Waybills_DM1.pdf 86.2 KB 2010_1031_Richardson_Alec_FinalCOCs_DM1.pdf 3.9 MB 2010_1030_Richardson_Alec_FieldCollectionForms_DM.xls 1.0 MB 2010_1031_Richardson_Alec_AccessioningReadMe_DM1.pdf 44.4 KB 2010_1031_Richardson_Alec_FieldCollectionForms_DM.xls 937.0 KB 2010_1031_Richardson_Alec_FieldCollectionForms_DM.xls_QAQC.rtf 4.1 KB 2010_1030_Richardson_Alec_SampleReadMe_DM.pdf 34.8 KB		
Visual Observations	10-30-2010	Unvalidated	2010_1031_Richardson_Alec_FieldDataForms_DM.pdf 1.7 MB 2010_1031_Richardson_Alec_NonanalyticalCOC_DM.pdf 740.3 KB 2010_1031_Richardson_Alec_GPX_DM.gpx 303.5 KB 2010_1031_Richardson_Alec_GDB_DM.gdb 55.7 KB 2010_1031_Richardson_Alec_Photos_DM1.zip 105.9 MB 2010_1031_Richardson_Alec_Photologger_DM1.pdf 45.9 KB 2010_1031_Richardson_Alec_ReadMe_DM1.pdf 48.1 KB		

[Close](#)

		Analysis File Collections					
7	Publicly Available	C	Find Electronic Data Form	2010	Chem--Forensic Oil Sampling 03 2010	Fr	S001
8	Publicly Available	C	Display links to view and download Electronic Data Forms and Sample Trip data	2010	Chem--Forensic Oil Sampling 04 2010	Fd	S001
9	Publicly Available	C	Download Lab Files	2010	Chem--Forensic Oil Sampling 04 2010	Fd	S001
10	Publicly Available	C	Direct download of laboratory files	2010	Chem--Forensic Oil Sampling 04 2010	Fd	S001

Dat

The

D

San

Vis

M

St

Ex

DIVER Explorer - Information x <https://files.noaa.noaa.org> x

[https://files.noaa.noaa.org/workgroups/twg-chemistry.nsf/0/8DCB96892A79C02C862577D90080BA9F/\\$FILE/2010_1031_Richard%20](https://files.noaa.noaa.org/workgroups/twg-chemistry.nsf/0/8DCB96892A79C02C862577D90080BA9F/$FILE/2010_1031_Richard%20)

22

10-31-10 NOAA NROA Team 7 (Forward Sample) Saturday p1 of 6

Marine launch: Small Craft Harbor, Biloxi, MS

Site # 334, 336, 786

Team members: ① Alec Richardson (CMI/NOAA) - 214-683-3381
② Gary Jacob (Shaw-LA)
③ Shane Cantrell (Entria)-

Instruments:-

Garmin GPS- BPGPS-62

Digital cameras- BPCAM68 and BPCAM-32

SPOT Tracker- NOAA NROA SSN 14

Radio- 700MHz RADIO-44 and NOAA NROA VHF3

Box- NOAA NROA 22

Satellite phone- BPSAT-5

Hotel departure- 05:00

Marine arrival- 06:30

Marine departure- 07:50

Site #

Anchor

Grid

Lat

Long

Arrival

Departure

334 LAAM44 29.95049 -88.83037 10:50 10:50 13:40

336 LAAM44 29.94774 -88.83069 none 13:50 17:00

786 LAAM44 29.94269 -88.82861 90-1006 17:30 18:20

Anchor

18:45

Photologger-

#

Info

Time

79 GPS unit 07:37

273 GPS unit 07:38

100 C-1(45) 11:17

101 C-1(40) 11:17

C-2(45) A.R.

102 C-3(45) 11:20

103 C-3(40) 11:20

106 P-1(45) 11:25

107 P-1(40) 11:25

P-2(40) A.R.

104 P-3(45) 11:23

105 P-3(40) 11:23

108, 109 P-1(60)(2) 12:30

P-2(60)(2) A.R.

110, 111 P-3(60)(2) 12:28

GPS Waypoints

#

Info

Lat/Long

Time

32 SS 29.95083 11:04

33 IS 29.95082 11:05

34 C-1 29.95084 11:14

35 C-2 29.95083 11:17

36 C-3 29.95083 11:17

37 P-1 29.95081 11:23

38 P-2 29.95081 11:21

39 P-3 29.95081 11:21

Junk photos-

cruise

Leg

Close

Ff	S001
Fd	S001
Fd	S001
Fd	S001



DIVER Explorer: Dashboard Approach

Data Summary

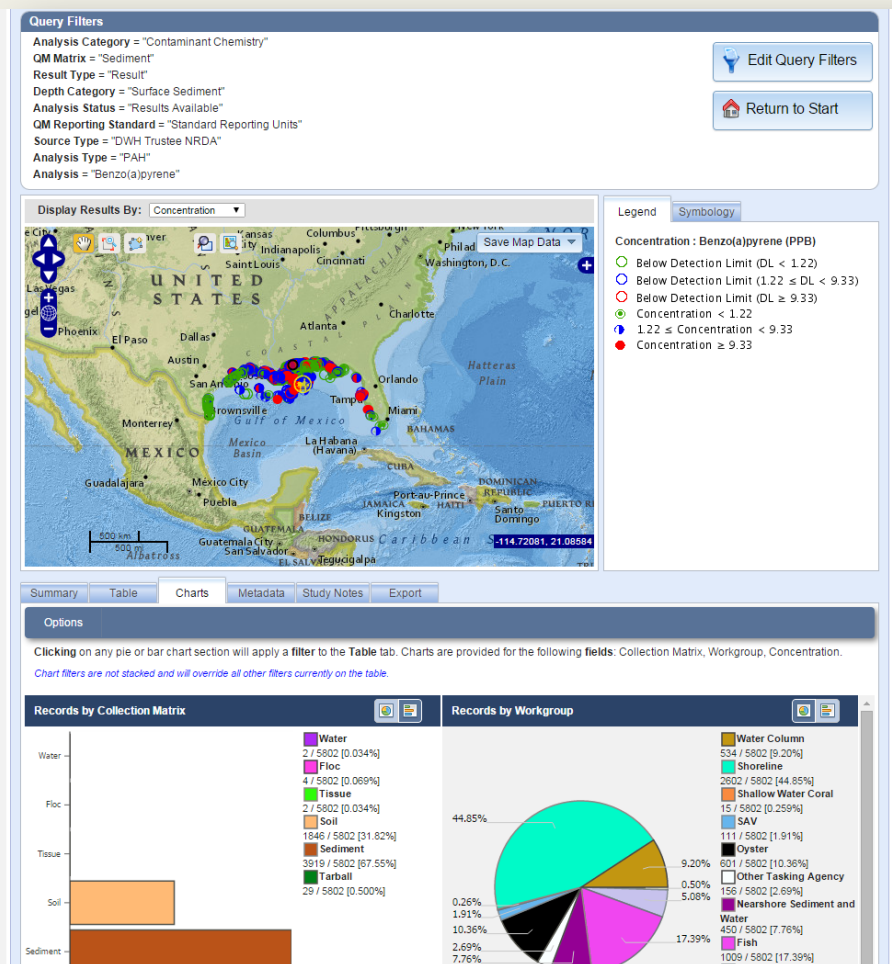
Data Table

Charts

Metadata

Study Notes

Export





DIVER Explorer: Dashboard Approach

Data Summary

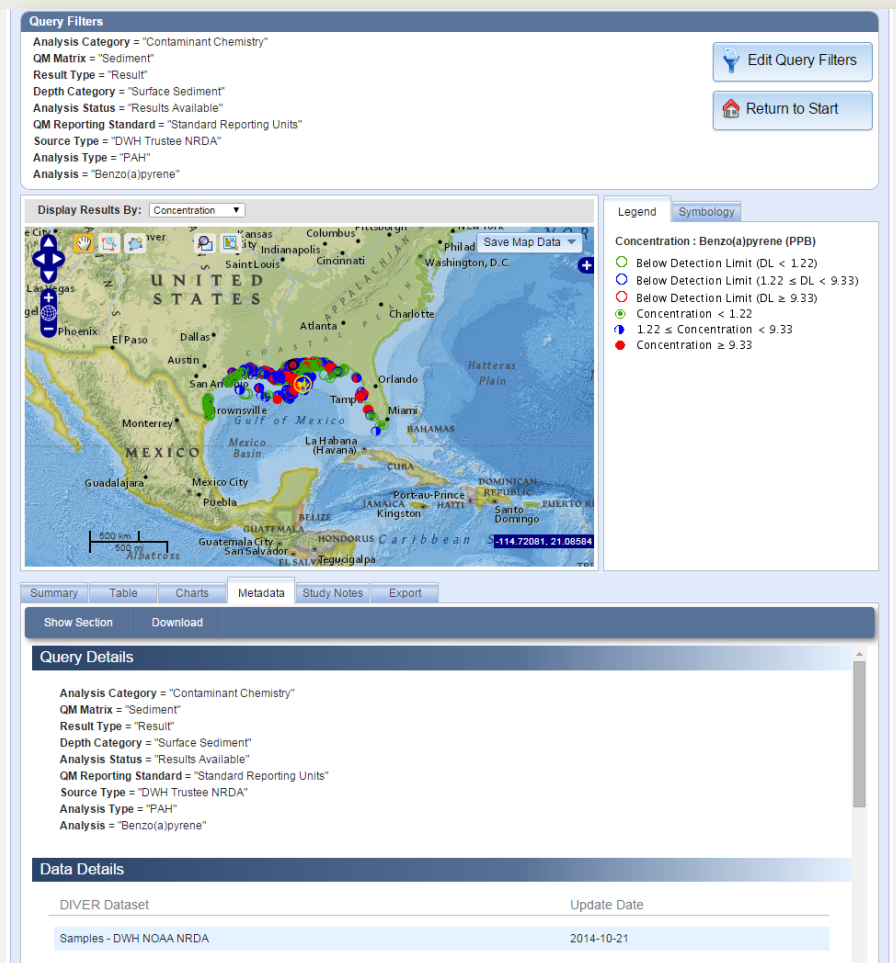
Data Table

Charts

Metadata

Study Notes

Export





DIVER Explorer: Dashboard Approach

Data Summary

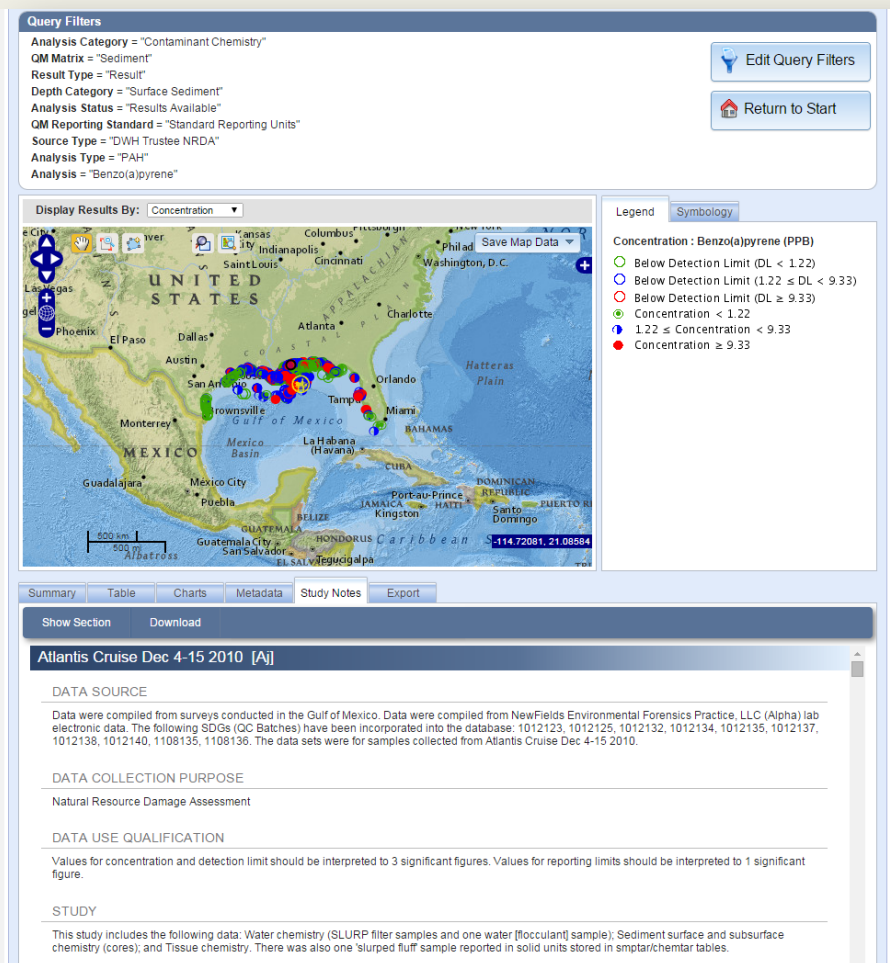
Data Table

Charts

Metadata

Study Notes

Export





DIVER Explorer: Dashboard Approach

Data Summary

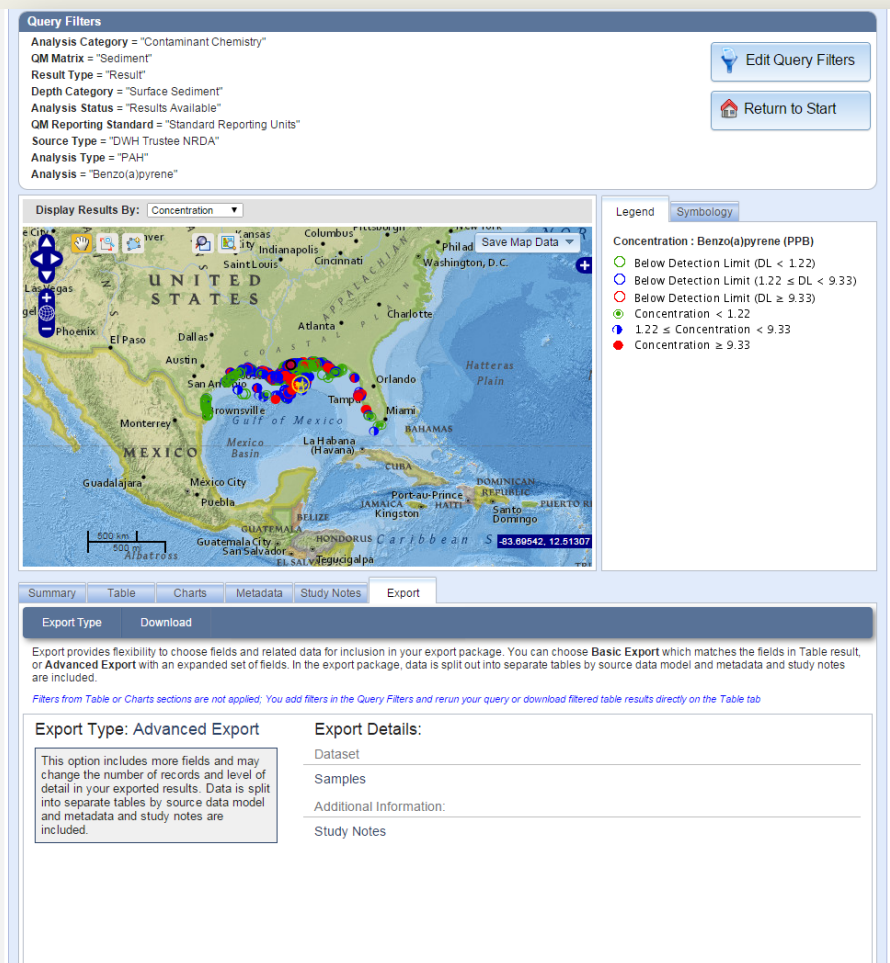
Data Table

Charts

Metadata

Study Notes

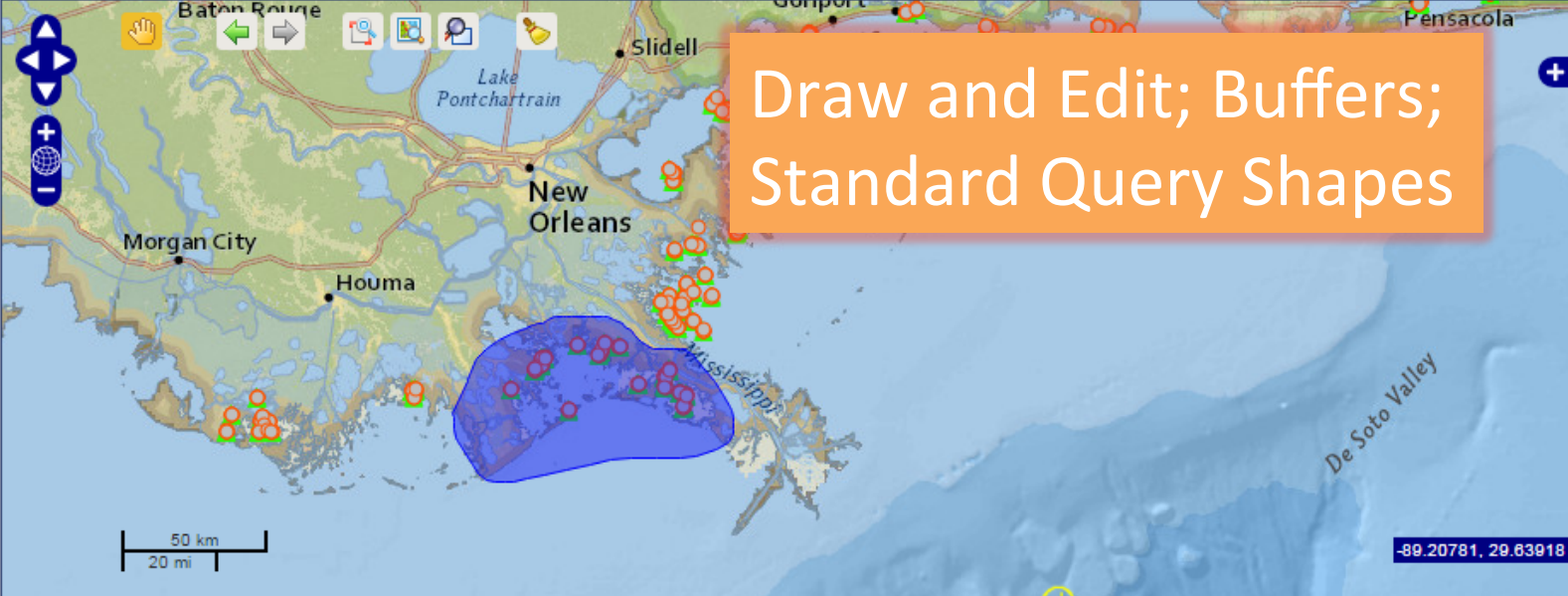
Export





DIVER Explorer: Query By Shape

Query Shapes



Draw and Edit; Buffers;
Standard Query Shapes

Draw and Edit Shape Load and Manage Shapes

Units: Miles Geometry: Polygon - Geodesic Area: 1,270.17 sq mi

Cancel Run Query



DIVER Explorer: Export

Query Filters
Analysis Category = "Contaminant Chemistry"
Source Type = "DWH Trustee NRDA"
QM Matrix = "Field Blank" | "Oil" | "Sediment" | "Tissue" | "Water"

Edit Query Filters
Return to Start

Display Results By: QM Matrix

Summary | Table | Charts | Metadata | Study Notes | Export

Save Query | Options

Data Summary for Contaminant Chemistry Results » Station Locations

The following is a summary of the DIVER Explorer query results. The summary is based on the query result data and is not updated with Table or Chart filters.

Records Returned : 17,395
*Number of records returned from a query is based on multiple fields and may not match result, event, or sample counts.

Non-spatial Records : 171

Workgroup : Chemistry, Deep Water Benthic, Fish, Marine Mammals and Turtles, Nearshore Sediment and Water, Other Tasking Agency, Oyster, SAV, Shallow Water Coral, Shoreline, Water Column

Workplan : 2011 Oyster Quadrat and Sediment Sampling Plan, 2012 Intertidal Oyster Quadrat Sampling Plan, 2013 Nearshore Oyster Sampling Plan, 25ft Parker UV Radiometer Fall 2010, American Diver Plankton Aug-Sep 2010, American Diver Plankton/Water Jul-Aug 2010, Blue Crab Megalopae Plan, Bunny Bordelon Plankton/Dispersant May-Jun 2010, CWV-Marsh Response Cleanup Areas-Fall 2011, CWV-Marsh Response Cleanup Areas-February 2011, CWV-Marsh Response Cleanup Areas-Spring 2011

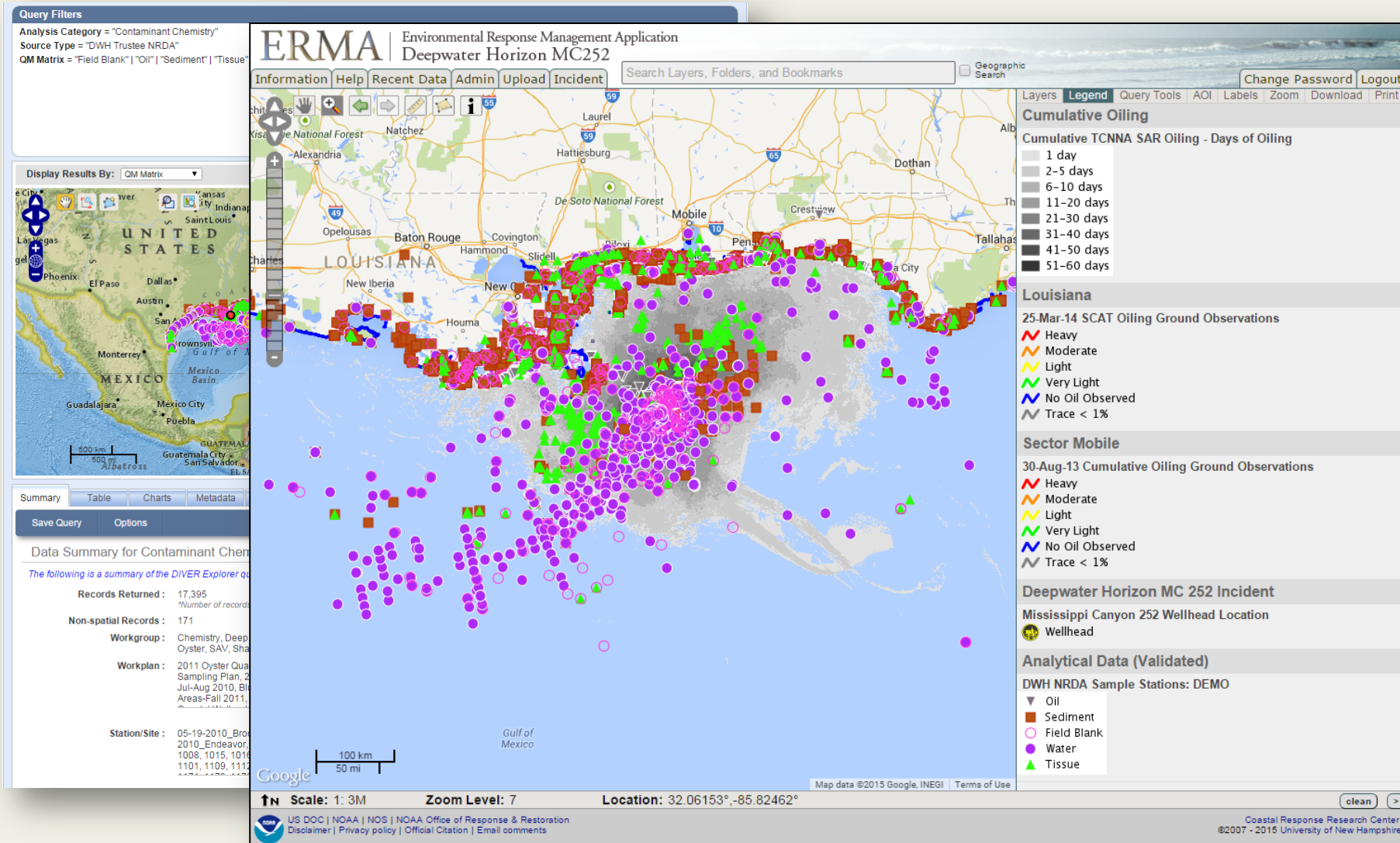
More

Station/Site : 05-19-2010_Brooks McCall, 05-27-2010_Ocean Veritas, 06-03-2010_Thomas Jefferson, 06-15-2010_Thomas Jefferson, 06-17-2010_Endavor, 06-29-2010_Ocean Veritas, 08-01-2010_Henry Bigelow, 08-25-2010_HOS Davis, 1.D1, 100, 1001, 1003, 1005, 1008, 1015, 1016, 1017, 1021, 1024, 1028, 1035, 1037, 1041, 1045, 1048, 1052, 1064, 1070-01, 1090, 1094, 1095, 110 Loop, 1101, 1109, 1112, 1114, 1116, 1124, 1128, 1129, 1144, 1145, 1148_1149, 1150, 1151, 1153, 1155, 1156, 1166, 1167, 1169

More



DIVER Explorer: Export





DIVER Explorer: Export

Export Packages:

- Includes full FGDC Metadata (Federal Geographic Data Committee)
- Spreadsheet; Shapefile (GIS); KML (Google Earth)
- Electronic field data (spreadsheets)
- Automated output of updated data to:



Gulf Spill Restoration

<http://www.gulfspillrestoration.noaa.gov/> and



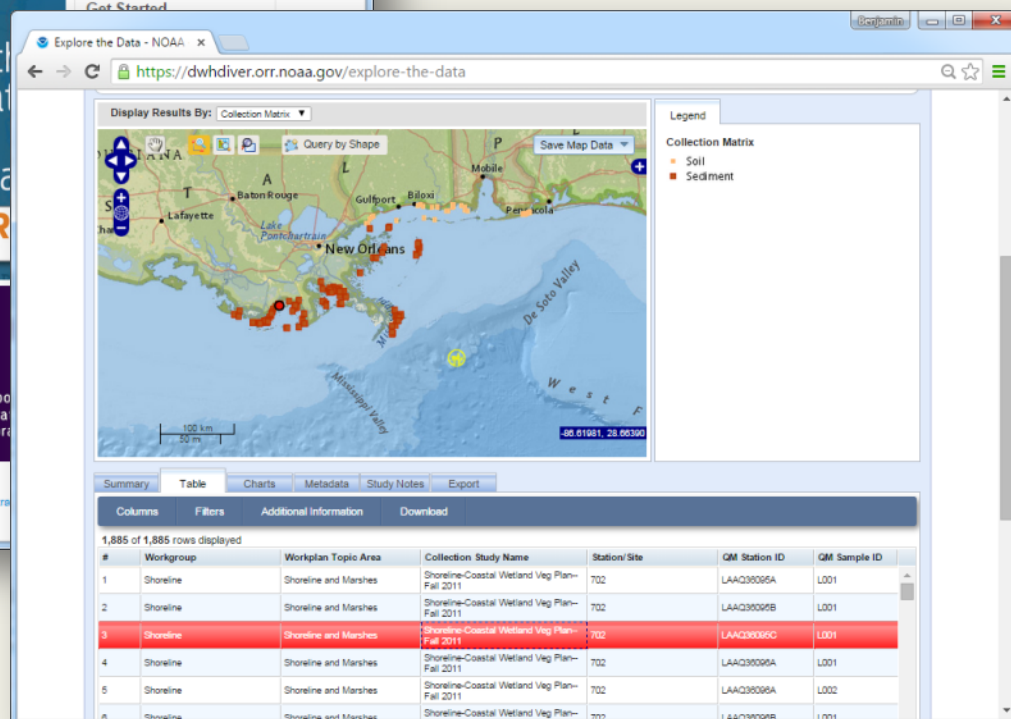
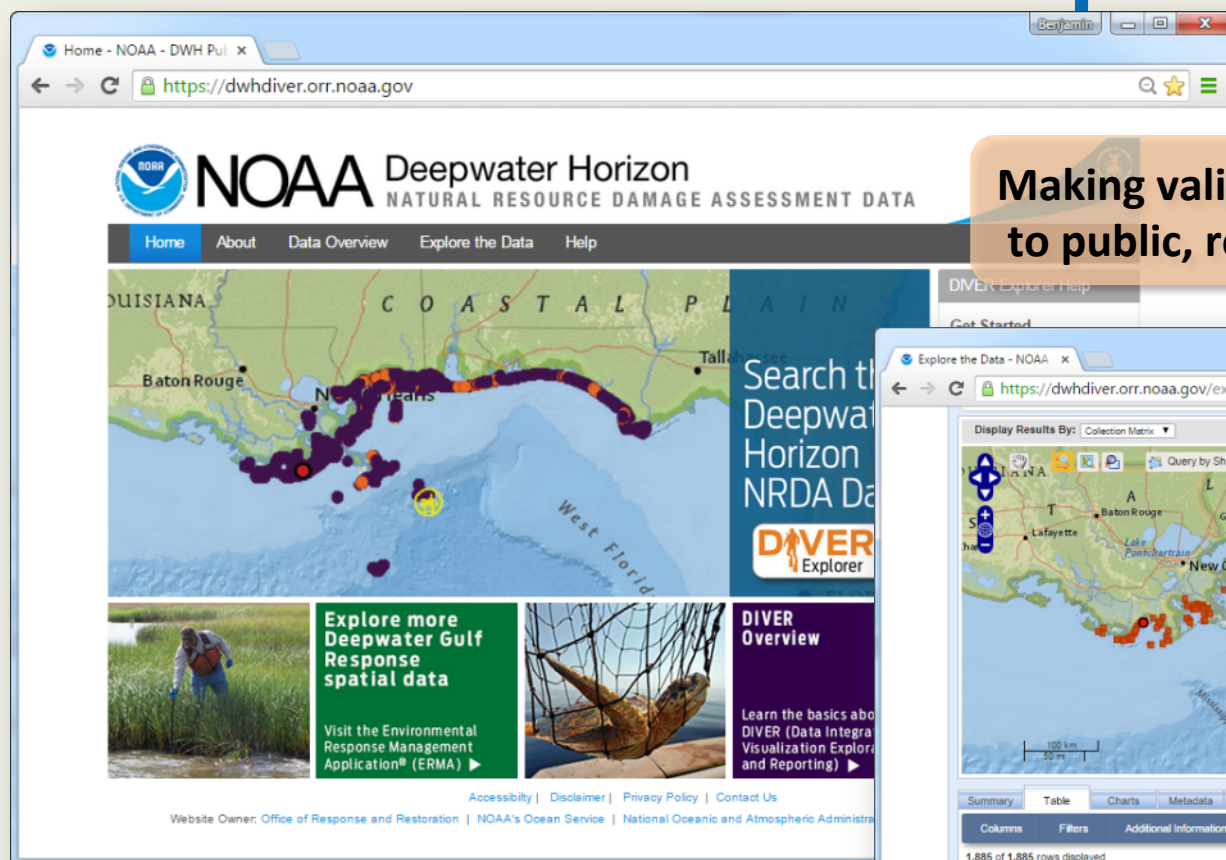
ERMA Gulf Response:

<http://gomex.erma.noaa.gov/erma.html>



Public DIVER for Deepwater Horizon

Making validated scientific data available to public, researchers and academia



Released May 4, 2015

<https://dwhdiver.orr.noaa.gov/>



DIVER Strategy for Data Management and Query Tools

National DIVER

- OR&R developing Regional DIVER sites, Contaminant Chemistry, Observational data, Photos, Restoration, Response & Restoration data models
- Flexible, Scalable, Adaptable
- Developing field data collection capability based on DWH techniques and tools- and ongoing OR&R response, assessment and restoration mission

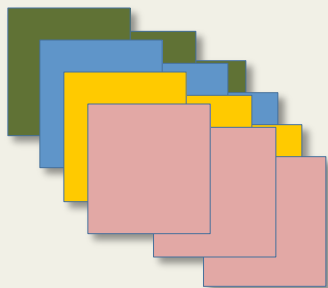


DIVER Strategy for Data Management and Query Tools

National DIVER

- Regional Data Focus
- Templates for data ingestion (flexible; many data types)
- Data Transcription
- Automated Ingestion (ETL)
- Quick review and download (sharing)
- Connection to raw data and documents

Current Capabilities

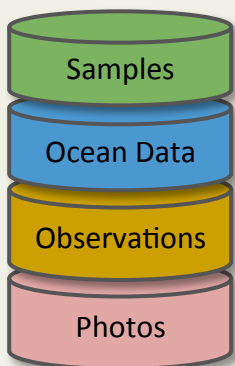


File Collections

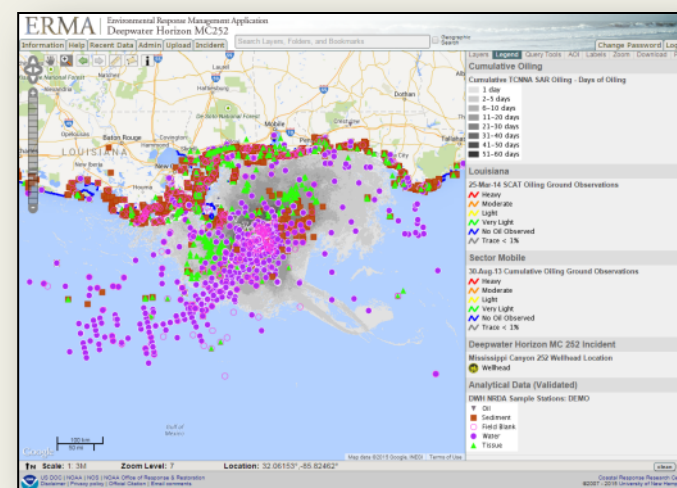
Ingest and manage data from many sources; structured and unstructured

Data Warehouse (ETL)

Integrate and standardize data from multiple sources (e.g. field-collected data, laboratory data, analysis)



Query and Export Tools





Santa Barbara: Refugio Beach Oil Spill

NOAA DIVER *Workspace* Beta

Admin My Sites 0 Ben

Home Regions Help Activity Configuration

National / Southwest/Pacific Islands / Refugio State Beach - American Pipeline / File Collections

File Collection

Show 100 entries

Category	Description
Ephemeral Data	ED Team 1 Subtidal Sampling
Ephemeral Data	Bird and beach photo observations
Ephemeral Data	ED Team 1 Pore Water Sampling
Ephemeral Data	ED Team 2 Pore Water Sampling
Ephemeral Data	Field observations of the cormorant colony.
Recreation Use	Recreation Use Site Photos
Ephemeral Data	ED Team 1 Water Sampling
Ephemeral Data	Bird and beach photo observations
Ephemeral Data	ED Team 2 Field Observations
Ephemeral Data	ED Team 5 Tarball Sampling
Ephemeral Data	ED Team 1 Tarball Sampling
Ephemeral Data	ED Team 3 Tarball Sampling
Ephemeral Data	ED Team 4 Tarball Sampling
Recreation Use	Recreation Use Site Photos
Ephemeral Data	ED Team 2 Tarball Sampling
Ephemeral Data	EPA Sea Urchin Fertilization Bioassay Results

File Collection

FILE COLLECTION
[EDIT FILE COLLECTION DEFINITION](#)

Start Date: 05/26/2015 End Date: 05/26/2015 ID: 2103

Description: ED Team 5 Tarball Sampling Data Classification: Assessment

Category: Ephemeral Data

Lead Sampler: Location: Organization:

Notes/Comments:

DATASETS FOR INTEGRATION INTO DIVER

Type	File Name	File Size	Upload Date	Uploaded By
Refugio DIVER Template - Trustee	2015_0526_2103 - Samples	3.9 KB	06/01/2015	Karmen Cheung
Refugio DIVER Template - Trustee	2015_0526_2103 - Photographs	1.5 KB	06/01/2015	Karmen Cheung

Select a Template... SELECT DATASET (.CSV) **UPLOAD**

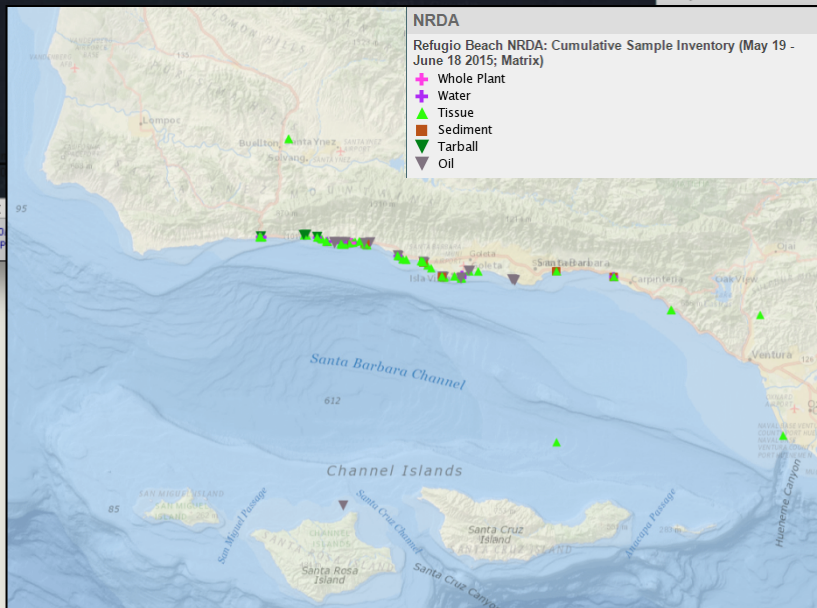
OTHER ASSOCIATED FILES

Type	File Name	File Size	Sharing Status	Review Status	Upload Date	Uploaded By
Chain of Custody	2015_0526_COC_ED5.pdf	1.1 MB	Trustee Only	New	05/27/2015	Neal Etre EDIT DELETE
Field Sample Form	2015_0526_FSF_ED5.pdf	4.6 MB	Trustee Only	New	05/27/2015	Neal Etre EDIT DELETE
GPS File (gpx)	2015_0526_ED5_GPS.gpx	99.2 KB	Trustee Only	New	05/26/2015	Ben Shorr EDIT DELETE
Original Image Files (zip)	2015_0526_ED5_Photos.zip	162.8 MB	Trustee Only	New	05/26/2015	Ben Shorr EDIT DELETE



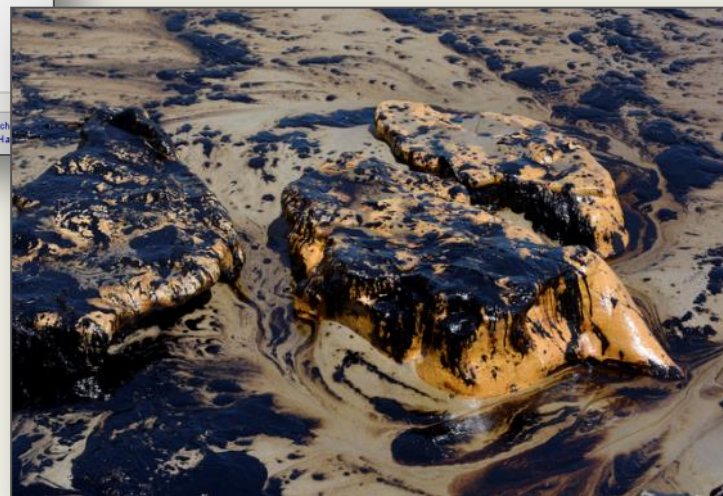
Santa Barbara: Refugio Beach Oil Spill

ERMA | Environmental Response Management Application
Southwest



9 - 2013)

Coastal Response Research
©2007 - 2015 University of New Ha





DIVER Technology and Processes:

Open Source Tools

- **ETL Process:** Pentaho Analytics
 - Schemas; PostgreSQL/PostGIS;
- **Portal:** Liferay
- **Explorer Query Tool**
 - Custom Java API and query engine
 - Mapserver/OpenLayers
 - Dojo Toolkit Javascript library
 - MDX query language – Pentaho Community Edition
 - Infobright (Hadoop integration)-fast column-oriented db



30 minutes goes quickly when you're talking big data!

Technical Details:

Amazon AWS; FedRAMP;
NIST 800-53 security standards compliant
Liferay Portal
Pentaho Data Integration tools
PostgreSQL/PostGIS
Infobright (Hadoop integration)
Mapserver/OpenLayers
Dojo Toolkit Javascript library
Custom Java API and query engine
Agile development approach:
(data management and tool development)
IBM Domino (phasing out)

Senior Team (and co-authors):

- Ben Shorr (Spatial Data Branch)
- Dr. Amy Merten (Spatial Data Branch Chief)
- Dan Hudgens (IEc Inc.)
- Neal Etre (IEc, Inc.)
- Jim Anderton (Solea Consulting)
- Jerry Bower (Sirius Computer Solutions)