Managing and Sharing BP Oil Spill Data from the Gulf of Mexico

Sampling and Monitoring Data

Jeffrey White
Tetra Tech, Inc.
April 22, 2010

The damaged oil rig sinks two days after the explosion, and a 1-by-5-mile oil slick appears in the water, according to Coast Guard Rear Adm. Mary Landry.

It is not known whether there is a leak on the rig or the well it was connected to underwater. David Rainey, vice president of BP, which leased the rig from Transocean Ltd., said “it certainly has the potential to be a major spill.”

June 18, 2010

Sixty days after the explosion on the Deepwater Horizon rig, up to 60,000 barrels a day continue to leak into the Gulf of Mexico.

September 19, 2010

Officials formally declared an end to the worst oil spill in U.S. history. After a weekend of pouring cement into the base of the ruptured well, pressure tests conducted early Sunday confirmed the seal was holding, former Coast Guard Adm. Thad Allen announced. The Interior Department agency that regulates offshore drilling pronounced the well dead at 6:54 a.m. ET.
Rapid Response Goals

- Water and Sediment samples collected daily
- New sample results available daily.
- New data required screening on a daily basis.
- Report back detections and exceedances of oil related compounds on a daily basis.
Overall Goals

I. Summary analysis of oil related chemical contaminants
   a. Number of Water and Sediment Samples.
   b. Number of Detections.
   c. Number of Chronic and Acute exceedances by chemical against benchmarks.

II. Migrate data used in analysis to the EPA STORET data warehouse.
   a. Data Validation
Work Flow for Data Analysis

1. Files (data) received daily from EPA.
2. Imported to Access Database for ‘Rapid’ Screening.
3. Data Returned to EPA the same day Screened and Summarized.
4. Data Moved to Oracle database (EDAS2) for Security and Additional Analysis.
Incoming Files

- Files received daily from EPA.
Initial ‘Rapid’ Data Processing

1. Import from Excel
   - Only the first two will be imported.
   - renames file with date (YYYYMMDD_Sample_Data_Water_Sediment)
   - be careful of import errors (data types), redo manually if have errors

2. Update “Current” data table: contains only the most recent data
   - Current Data Table - Delete Old
   - Current Data Table - Append New
   - double check number of records

3. Export data - Directory
   - Set Export Directory to Network Location with Date
   - Create Export Directory based on above location

3.1. Export data as Excel
   - check records, if no detects then don’t have to export
   - Export Detects - WATER
   - Export Detects - SEDIMENT
   - Export Detects - Sediment (Soil)
   - Export Detects - Sediment (Silt)
   - Export Detects - OTHER

WATER
- 06_Detects_WATER__List_of_Surface_Water_SampleIDs
- 02_ResultsDivided_Water_AquaticLife_0IRelatedCompound
- 02_Narrative_WATER_AquaticLife_OIlRelatedCompound
- 03_Water_AquaticLife_Chronic_Exceed
- 02_Narrative_WATER_AquaticLife_Metals
- 03_Water_AquaticLife_Metals_Exceed
- 02_Narrative_WATER_HumanHealth
- 03_Water_HumanHealth_Exceed

SEDIMENT
- 06_Detects_SEDIMENT__List_of_Sediment_SampleIDs
- 06_Detects_SEDIMENT
- 02_ResultsDivided_SEDIMENT_AquaticLife
- 02_Narrative_SEDIMENT_AquaticLife_Metals
- 03_SEDIMENT_Metals_Exceed

Open Query Design - Append Previous

1A. Update CAS #s in Import
   - Import - CAS# - Update - Sediment
   - Import - CAS# - Update - Water
   - Query_Update_Water_CAS#
Queries and Maps in EDAS2
Viewing Bench Marks using the Map
Viewing Sampling Locations

Monitoring Location Search Results

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Description</th>
<th>MUC 8</th>
<th>MUC 12</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Tag</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR25</td>
<td>IR25</td>
<td>ROADWAY</td>
<td>30.692254997</td>
<td>88.117459933</td>
<td>Alabama DHM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCA10-1284</td>
<td>NCA10-1284</td>
<td>ROADWAY</td>
<td>29.06728115</td>
<td>86.944287935</td>
<td>Florida DHM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCA10-1285</td>
<td>NCA10-1285</td>
<td>ROADWAY</td>
<td>29.91751602</td>
<td>84.117666864</td>
<td>Florida DHM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCA10-1297</td>
<td>NCA10-1297</td>
<td>ROADWAY</td>
<td>30.09956278</td>
<td>84.30774502</td>
<td>Florida DHM</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Monitoring Location Search Results

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Description</th>
<th>MUC 8</th>
<th>MUC 12</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Tag</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR25</td>
<td>IR25</td>
<td>ROADWAY</td>
<td>30.692254997</td>
<td>88.117459933</td>
<td>Alabama DHM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCA10-1284</td>
<td>NCA10-1284</td>
<td>ROADWAY</td>
<td>29.06728115</td>
<td>86.944287935</td>
<td>Florida DHM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCA10-1285</td>
<td>NCA10-1285</td>
<td>ROADWAY</td>
<td>29.91751602</td>
<td>84.117666864</td>
<td>Florida DHM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCA10-1297</td>
<td>NCA10-1297</td>
<td>ROADWAY</td>
<td>30.09956278</td>
<td>84.30774502</td>
<td>Florida DHM</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The STORET Data Warehouse is a repository for water quality, biological, and physical data and is used by state environmental agencies, EPA and other federal agencies, universities, private citizens, and many others. [http://www.epa.gov/storet/](http://www.epa.gov/storet/)

The Water Quality Exchange (WQX) is a framework that makes it easier for organizations to submit and share water quality monitoring data over the Internet. [http://www.epa.gov/storet/wqx/index.html](http://www.epa.gov/storet/wqx/index.html)

STORET is the home for data submitted through WQX.
Scribe Database

- Scribe is a database developed by the USEPA's Environmental Response Team (ERT) to assist in the process of managing environmental data\(^1\).
- Scribe is the official database for managing and distributing analytical sample data collected in response to the Oil Spill. Scribe captures sediment, water, and biota sampling data\(^2\).

\(^1\)http://www.epaosc.org/site/site_profile.aspx?site_id=Scribe
1. Tetra Tech used a web service (Scribe.NET) connected to a local copy of Scribe to maintain and keep water and sediment sampling data up to date.

2. Create a staging database with Scribe tables and tables containing valid WQX domain values and functions to generate XML.
EPA SCRIBE to EPA STORET Data Migration Steps

3. Run scripts from UpdateDomainValues.sql file which updates tables with valid WQX domain values.

4. Create procedure to generate an valid XML file from SCRIBE for import to WQX\STORET.

```xml
<Payload Operation="Update-Insert">
  <WQX xmlns="http://www.exchangenetwork.net/schema/wqx/2"
       xsi:schemaLocation="http://www.exchangenetwork.net/schema/wqx/2/1/WQX_WQX_v2.1.xsd">
    <Organization>
      <OrganizationDescription>
        <OrganizationIdentifier>DWH_SCRIBE1082</OrganizationIdentifier>
        <OrganizationFormalName>BP Deep Water Horizon Oil Spill</OrganizationFormalName>
        <OrganizationDescriptionText>Surface water and sediment sampling collected in response to Spill</OrganizationDescriptionText>
      </OrganizationDescription>
    </Organization>
    <Project>
      <ProjectIdentifier>R04DW</ProjectIdentifier>
    </Project>
  </WQX>
</Payload>
```
Sharing Data – Scribe to WQX / STORET

Consolidated EPA Data Project file published to Scribe.NET

EPA Regional Project File (1082)

Data Auditor/ Data Specifications

SCRIBE.NET

Error checking per data specifications

XML Generation and Parse and Load

STORET Data Warehouse

ETL

WQX ODS

WQX Staging Tables (SCRIBE format)

Data Access: Query Application Web Services
Data Validation and Cross Mapping

- Are Compounds (Alkyl PAHs and Dispersants) in WQX/STORET or will they need to be updated in WQX/STORET.
- Are all the Data Elements (fields) required in WQX (e.g., Sample Fraction) in SCRIBE?
- Are there missing values?
- Do the Domain Values (Valid Values) match between SCRIBE and WQX (e.g., Monitoring Location Type, Activity Type)?
- Configuration file - Transform values as needed.
**SCRIBE to WQX / STORET**

- Cross map SCRIBE tables and fields to appropriate WQX data elements

<table>
<thead>
<tr>
<th>SCRIBE.Samples</th>
<th>WQX.Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samples-&gt;Samp_No</td>
<td>*ActivityIdentifier</td>
</tr>
<tr>
<td>Samples-&gt;SampleType</td>
<td>*ActivityTypeCode</td>
</tr>
<tr>
<td>Samples-&gt;Matrix</td>
<td>*ActivityMediaName</td>
</tr>
<tr>
<td>Samples-&gt;SampleDate</td>
<td>*ActivityStartDate</td>
</tr>
</tbody>
</table>
SCRIBE Data Validation

- No apparent Domain List to enforce consistent naming.
- No or minimal required fields to minimize missing data.

<table>
<thead>
<tr>
<th>WQX Requirement Data Requirement</th>
<th>SCRBIE Table</th>
<th>Default Value submitted to WQX</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SampleFraction</td>
<td>Total or Dissolved from LabResults</td>
<td>Total or Dissolved, WQX valid values</td>
<td>WQX valid value are Total or Dissolve, will transform T or D from SCRBIE, many values in SCRBIE are null</td>
</tr>
<tr>
<td>CharacteristicName</td>
<td>Analyte from LabResults</td>
<td>Only Valid characteristic names have been mapped,</td>
<td>Waiting on input/decision on non matching names.</td>
</tr>
<tr>
<td>MeasureUnitCode</td>
<td>Result Unit from LabResults</td>
<td>Valid mapped WQX domain values</td>
<td>Some unit codes in SCRBIE are not valid in WQX</td>
</tr>
<tr>
<td>ResultStatusIdentifier</td>
<td>Not in SCRBIE</td>
<td>‘Accepted’, a WQX valid value</td>
<td>Domain list in WQX is Result Status</td>
</tr>
<tr>
<td>ResultValueTypeName</td>
<td>Not in SCRBIE</td>
<td>‘Actual’, a WQX valid value</td>
<td>TIC/TRG/MS/Null values in SCRBIE are not valid in WQX</td>
</tr>
</tbody>
</table>
Generating a WQX Compatible XML File

- Processing Report
- XML File

```
- <Payload Operation="Update-Insert">
  - <WQX xmlns="http://www.exchangenetwork.net/schema/wqx/2">
    - <Organization>
      - <OrganizationDescription>
        <OrganizationIdentifier>DWHSCRIBE1082</OrganizationIdentifier>
        <OrganizationFormalName>BP Deep Water Horizon Oil Spill</OrganizationFormalName>
        <OrganizationDescriptionText>Surface water and sediment sampling collected in response to the BP Gulf of Mexico oil spill</OrganizationDescriptionText>
      </OrganizationDescription>
      - <Project>
        <ProjectIdentifier>R04DW</ProjectIdentifier>
        <ProjectName>Deepwater (R04)</ProjectName>
        <ProjectDescriptionText>Deepwater (R04)</ProjectDescriptionText>
      </Project>
      - <MonitoringLocation>
        - <MonitoringLocationIdentity>
          <MonitoringLocationIdentifier> R4DART </MonitoringLocationIdentifier>
          <MonitoringLocationName> R4DART </MonitoringLocationName>
          <MonitoringLocationTypeName>Ocean</MonitoringLocationTypeName>
          <MonitoringLocationDescriptionText>Near Shore</MonitoringLocationDescriptionText>
        </MonitoringLocationIdentity>
    </Organization>
  </WQX>
```

Processing Report

Transaction ID: _c366863e-edfa-49e4-a54a-a6f3c9e65d0
Status: Completed

Software Information

Component Version
WQX Code: 2.10
WQX Dataset: 1.20

Summary Information

# Errors: 0
# Warnings: 0

Items Successfully Processed:

<table>
<thead>
<tr>
<th>Projects</th>
<th>Insert</th>
</tr>
</thead>
</table>
| 1

<table>
<thead>
<tr>
<th>Monitoring Locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insert</td>
</tr>
</tbody>
</table>
**Destination STORET**

**STORET**

**U.S. ENVIRONMENTAL PROTECTION AGENCY**

**Results by Project (stormod_)**

<table>
<thead>
<tr>
<th>Project ID</th>
<th>Project Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>R04DW</td>
<td>Deepwater (R04)</td>
</tr>
<tr>
<td>R06DW</td>
<td>Deepwater (R06)</td>
</tr>
</tbody>
</table>

**Step 1:** Select a Single Organization from the List

<table>
<thead>
<tr>
<th>ORG ID</th>
<th>ORGANIZATION NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>DWH Scribe1082</td>
<td>BP Deep Water Horizon Oil Spill</td>
</tr>
</tbody>
</table>

**Step 2:** Select a Single Project by Clicking “Look Up”

[Look Up]
Success, the sampling data are now in STORET and downloadable by the public.
Conclusions and Data Needs

- Require use of domain lists (valid values) to:
  - To reduce data entry errors or omissions in the metadata (such as information on sample location and depth) associated with sediment and water samples collected for PAH (and other) analysis.
  - New values may be added to domain lists.

- Data fields that require unique names, for example station or sample identifier.

- Minimizing inconsistencies at the time of data entry will reduce the need to re-analyze the data.

- Submission of data from multiple labs.
References:

- Scribe website:

- Summary Report Sub-Sea and Sub-Surface Oil and Dispersant Detection: Sampling and Monitoring, Dec 17, 2010

- STORET \ WQX
  [http://www.epa.gov/storet/wqx/index.html](http://www.epa.gov/storet/wqx/index.html)
Acknowledgements

- **Tetra Tech:**
  - Jerry Diamond
  - Vladi Royzman
  - Liejun Wu
  - Sunitha Sajjan
  - Erik Leppo

- **EPA:**
  - Charles Kovatch
  - Treda Grayson
  - And others that provided the daily data feed

Jeffrey.White@tetratech.com