Evaluation and Comparison of Tables from Scribe Import Files and QM DBF files

# From Response Scribe Databases to Import Files to Query Manager (QM) Files

The Response import files were generated from data collected and entered into the Scribe database format by several states, federal agencies and BP contractors. The initial Scribe database and import file versions are listed in Table 1 and are archived in the NOAA Deepwater Horizon sftp site at Deepwater\_Horizon\_Ext/Scribe\_to\_QM\_Mapping/FinalReport/ in files: Scribe\_Databases.zip and Response\_Import\_Files.zip, respectively.

The NOAA\_to\_QM-Update.accdb import file was developed by matching the more accurate and complete station and sample information in NOAADW.mdb with the chemistry data in Envstd\_Sampling\_Analytical\_2.mdb. The unmatched data were reported in ESI\_to\_QM\_Revised\_w\_Append\_6-26-12.accdb. The data collected with onboard laboratories and sensors was extracted from NOAADW.mdb and reported in NOAADW\_OnBoard\_Update\_with\_RyanChouest.accdb. The total numbers of records in the Query Manager files are shown in the Record Cnt column.

No data were imported from CTEH\_DW\_Monitoring.mdb, TS\_DW\_Monitoring.mdb, and LDEQ\_DW\_Sampling\_Analytical.mdb because all the data were monitoring and/or air quality data. No data were imported from BP\_DW\_Sampling\_Analytical.mdb because the collection dates and times or matrices did not match supposed split samples reported in EnvStds\_DW\_Sampling\_Analytical.mdb. Other data were not imported because they were records for tentatively identified compounds, sieve-grain analyses that were converted to percent soil types, waste samples, bioassay results, calculated values, repeated data, non-environmental samples, or problems with the units or matrices. The process for determining and documenting non-imported data is described in Appendix A. Data Not in QM.

The data that was not imported to QM are archived in the NOAA Deepwater Horizon sftp site at Deepwater\_Horizon\_Ext/Scribe\_to\_QM\_Mapping/FinalReport/Data\_not\_in\_QM.zip.

The process for creating the import files from the source files is described in Appendix B. Creating Import Files from Scribe Source Files.

# Creating Study-Media Tables from Import and DBF Tables

The fields in the Scribe source databases, the Access import databases and the Query Manager (QM) FoxPro™ tables are shown in Table 2. The Chem, Samples, and Stations tables in the import files were used to compare to the data in the dbf files. The ChemQC, COC, EDD, and SampleTags import tables were not used because they contain information in fields that were not compared or fields that were duplicated in the Chem, Samples, and Stations tables.

Three dbf tables, eddchem.dbf, eddchmsp.dbf and smpmstr.dbf contain fields that link to the import files. These fields are the identifiers for studies, stations, and samples. These dbf tables also contain fields not in the dbf station, sample and chemistry tables. A table was created from the pertinent fields in these three tables using the following procedure:

1. Create Access database file
2. Link eddchem.dbf, eddchmsp.dbf and smpmstr.dbf to the file
3. Create extracted information from the smpmstr.dbf table using qry\_Extract\_from\_smpmstr and make the table, Ext\_smpmstr.
4. Create table called EDD using qry\_Create\_EDD which contains pertinent fields and joins Ext\_smpmstr eddchem.dbf, and eddchmsp.dbf.

The dbf files and the EDD table are included in Deepwater\_Horizon\_Ext/Scribe\_to\_QM\_Mapping/FinalReport/Source\_DBF\_files.zip.

Visual Basic for Application (VBA) code, Access tables were created for 231 study-media combinations from the QM station, sample and chemistry dbf files and created EDD table. The code is included in the Excel file, Deepwater\_Horizon\_Ext/Scribe\_to\_QM\_Mapping/FinalReport/Comparing\_QM\_to\_Import\_Files.xlsm. The subroutine, DBF\_Tables, created the 231 tables, which are housed in 24 Access files, DBF1.accdb through DBF24.accdb. The Access tables were divided into groups of 10 tables to avoid the 2 gigabyte (gb) size limit of Access files and to improve processing efficiencies. The 24 Access files that contain the 231 tables can be located at Deepwater\_Horizon\_Ext/Scribe\_to\_QM\_Mapping/FinalReport/Tables\_from\_DBFs.zip.

The 231 CSV files for NODC archiving were also created by the subroutine, DBF\_Tables. These files are contained in Deepwater\_Horizon\_Ext/Scribe\_to\_QM\_Mapping/FinalReport/Final\_Response\_QM\_CSV\_Files.zip.

The subroutine, IMP\_Tables, created 231 tables from the import database files to compare to the 231 tables created from the dbf files. These tables are housed in 24 Access files, IMP1.accdb through Imp24.accdb and are compressed in Deepwater\_Horizon\_Ext/Scribe\_to\_QM\_Mapping/FinalReport/Tables\_from\_Imports.zip.

Queries used to create the sets of tables and perform the following comparisons as well as a Scribe to QM cross-referencing table and DepthConverter module are in Deepwater\_Horizon\_Ext/Scribe\_to\_QM\_Mapping/FinalReport/DBF\_IMP\_Queries.accdb.

# Comparing DBF to Import Files

Differences in the two sets of tables, the 231 tables created from the QM dbf files and the 231 table created from the tables created from the Scribe databases, was evaluated using eight subroutines:

|  |
| --- |
| 1. In\_DBF\_not\_in\_IMP |
| 1. In\_IMP\_not\_in\_DBF |
| 1. Station Difference |
| 1. SedSample\_Difference |
| 1. WatSample\_Difference |
| 1. TissSample\_Difference |
| 1. Chem\_Fields\_Difference |
| 1. Con\_Fields\_Difference. |
|  |

Each subroutine created a table of records with unmatched values. The results are summarized in Table 4. Results of Comparisons of DBF and Import Tables and listed in Deepwater\_Horizon\_Ext/Scribe\_to\_QM\_Mapping/FinalReport/Value\_Differences\_Between\_Import\_and\_DBF.zip

The subroutine, In\_DBF\_not\_in\_IMP, an outer join query to find records in each QM study-media table that cannot be matched in the corresponding import study-media table. The join is between the IMPFILEREC fields in the QM table to the autonumber ID[[1]](#footnote-1) in the import EDD table. The result was all the records in the QM tables were matched to records in the corresponding import tables. 3,880 records were found. All the records were the result of conversions from sieve-size analyses to percent soil type as explained in the QM notes.

The subroutine, In\_IMP\_not\_in\_DBF, an outer join query to find records in each import study-media table that cannot be matched in the corresponding QM study-media table. Again, the join is between the IMPFILEREC fields in the QM table to the autonumber ID in the import EDD table. The result was all the records in the import tables were matched to records in the corresponding QM tables. No records were found.

The results from all the following difference subroutines are explained in Table 4.

The subroutine, Station\_Difference, reports differences in the following three station-level fields:

1. Non-equal values in QM Datum field and the corresponding import file HDatum field,
2. Absolute differences greater than 0.0001 between the QM and import latitude fields
3. Absolute differences greater than 0.0001 between the QM and import longitude fields

The subroutines, SedSample\_Difference, WatSample\_Difference[[2]](#footnote-2), and TissSample\_Difference all report differences in the following three station-level fields:

1. Non-equal values in QM SAMPDATE field and the corresponding import file Sample Collection Date field,
2. Non-equal values in QM SAMPTIME field and the corresponding import file Sample Collection Time field,
3. Non-equal values in QM SAMPTYPE field and the corresponding import file Sample Type Code field.

In addition, the subroutines, SedSample\_Difference, and WatSample\_Difference, report differences in the following two station-level fields:

1. Non-equal values in QM MATRIX field and the corresponding import file Matrix field,
2. Absolute differences greater than 0.02 between the QM WTRDepth and import water depth (m) fields

Finally, the subroutine, SedSample\_Difference, also report differences in the following two station-level fields:

1. Absolute differences greater than 0.02 between the QM UDEPTH and import upper collection depth fields
2. Absolute differences greater than 0.02 between the QM LDEPTH and import lower collection depth fields.

The subroutine, Chem\_Difference reports differences in the following ten non-concentration fields found in the chemistry tables:

1. Non-equal values in QM and the corresponding import file LabID field,
2. Non-equal values in QM CASNUM field and the corresponding import file CAS Number field
3. Non-equal values in QM ANALTYPE field and the corresponding import file Analyte type field
4. Non-equal values in QM DILFACT field and the corresponding import file Dilution Factor field
5. Non-equal values in QM TOTALDISS field and the corresponding import file Total or Dissolved field
6. Non-equal values in QM and the corresponding import file Method field,
7. Non-equal values in QM and the corresponding import file QCBatch field,
8. Non-equal values in QM and the corresponding import file DVLevel field,
9. Non-equal values in QM LABNAME field and the corresponding import file Lab Name field
10. Non-equal values in QM SAMPIDCOC field and the corresponding import file SampleID\_COC field.

The subroutine, ConFields\_Difference reports differences in the following three concentration fields found in the chemistry tables:

1. Concentration (CONC)values in QM and the corresponding import file to within 1 percent,
2. Detection Limit (DL) values in QM and the corresponding import file to within 1 percent,
3. Reporting Limit (RL) values in QM and the corresponding import file to within 1 percent,

Table 1. Deepwater Horizon Scribe Source and Import Files

|  |  |  |  |
| --- | --- | --- | --- |
| Scribe Source File(s) | Version(s) | Import File | Record Cnt |
| ADEM\_DW\_Sampling\_Analytical\_Monitoring.mdb | 14 | ADEM\_to\_QM.accdb | 4295 |
| ALECI\_DW\_Sampling\_Analytical.mdb | 22 | ALECI\_to\_QM-Revised.accdb | 24941 |
| CTEH\_DW\_Sampling\_Analytical.mdb | 34 | CTEH\_to\_QM\_Update.accdb | 594172 |
| DW\_Reporting.mdb | 219 | EPA\_to\_QM-Revised.accdb | 308041 |
| Envstd\_Sampling\_Analytical\_2.mdb | 69 | ESI\_to\_QM\_Revised\_w\_Append\_6-26-12.accdb | 606904 |
| FLDEP\_DW\_Sampling\_Analytical.mdb | 76 | FLDEP\_to\_QM.accdb | 31759 |
| MSDEQ\_DW\_Sampling\_Analytical.mdb | 24 | MSDEQ\_to\_QM.accdb | 12449 |
| NOAADW.mdb/Envstd\_Sampling\_Analytical\_2.mdb | 81/69 | NOAA\_to\_QM-Update.accdb | 613384 |
| NOAADW.mdb | 81 | NOAADW\_OnBoard\_Update\_with\_RyanChouest.accdb | 86155 |
| NPS\_DW\_Sampling\_Analytical.mdb | 4 | NPS\_to\_QM.accdb | 46 |
| USGSDW.mdb | 18 | USGS\_to\_QM-Update.accdb | 35592 |
| BP\_DW\_Sampling\_Analytical.mdb | 6 | NA | 0 |
| CTEH\_DW\_Monitoring.mdb | 15 | NA | 0 |
| TS\_DW\_Monitoring.mdb | 6 | NA | 0 |
| LDEQ\_DW\_Sampling\_Analytical.mdb | 55 | NA | 0 |

Table 2. Fields Mapped - Scribe to Import to QM



Table 3. FoxPro Tables by Media

|  |  |  |
| --- | --- | --- |
| Media | Samples | Chemistry |
| Sediment, Soil | sample.dbf, smpsedsb.dbf | chem.dbf, chemsb.dbf |
| Water | smpwat.dbf | chemwat.dbf |
| Tar, Oil | smptar.dbf | chemtar.dbf |
| Tissue | smptiss.dbf | chemtiss.dbf |

Table 4. Results of Comparisons of DBF and Import Tables

| **Source Table(s)** | **Field(s)** | **Count** | **Comment** | **Status** |
| --- | --- | --- | --- | --- |
| Station | LATITUDE or LONGITUDE | 6 | Roundoff | Acceptable Change |
| 4 | EPA import stations 0022-100815 and 2002-100616-N changed to StationIDs 0032-815 and 2002-616 | Acceptable Change, documented in QM Notes |
| DATUM | 33 | Blank or WGS84 changed to U or WGS84 changed to NAD83 | Acceptable Change, documented in QM Notes |
| Samples | LDEPTH, UDEPTH or WTRDEPTH | 5621 | Change between -9, 0 or blank and a non-zero value or change between two non-zero values | Acceptable Change, documented in QM Notes |
| MATRIX | 4841 | Matrix changed from import files to QM tables due to Ryan Chouest onboard data not normalized and consolidation in QM | Acceptable Change, see Table 5 for count by matrix |
| SAMPDATE | 3 | Changed to next day | Acceptable Change, documented in QM Notes |
| SAMPTIME | 394 | 12 hours added | Acceptable Change, documented in QM Notes |
| 643 | Changed from blank to -9 | Expected Correction |
| 1075 | Form corrected from 'H:MM:' to 'HH:MM' | Expected Correction |
| 3250 | Form corrected from 'H:MM:' to 'HH:MM'and 12 hours added | Expected Correction, documented in QM Notes |
| 36 | Reason for change unknown | Documented in QM Notes |
| Chemistry | DILFACT | 1439 | Roundoff | Acceptable Change |
| DVLEVEL | 35232 | Changed from UNK to VALX; all USGS results | Expected Correction |
| LABID | 35232 | Blank changed to NR; all USGS results | Expected Correction |
| 11666 | LabRep swapped; all EPA records | Acceptable Change |
| LABNAME | 222 | LabRep swapped; all EPA Region 6 records | Acceptable Change |
| QCBATCH | 884 | LabRep swapped | Acceptable Change |
| 10300 | Blank changed to NR | Expected Correction |
| TOTALDISS | 1364 | LabRep swapped | Acceptable Change |
| DL and RL | 5574 | Roundoff | Acceptable Change |
| 442 | Conversion changes | Expected Correction, documented in QM Notes |
| 394 | RL and DL swapped on two EPA samples SRSnd-SD-08252010-Dup and UPMB-SD-20100503 | Acceptable Change, documented in QM Notes |
| CASNUM | 0 | No differences | Accepted |
| ANALTYPE | 0 | No differences | Accepted |
| METHOD | 0 | No differences | Accepted |
| SAMPIDCOC | 0 | No differences | Accepted |
| CONC | 109848 | Concentration changed to DL | Accepted |
| 19631 | Concentration changed to RL | Accepted |
| 4417 | DBF:IMP ratio repeated | Acceptable Change, documented in QM Notes |
| 3574 | Roundoff to within 5 percent | Acceptable Change |
| 2844 | LabRep Swapped | Acceptable Change |
| 1385 | Converted from molar units | Acceptable Change |
| 1327 | -9 concentration changed to value <> RL or DL | Acceptable Change |
| 640 | Metal - change in units; see QM metadata comments | Acceptable Change |
| 250 | Ratio not repeated | Documented in QM Notes |
| 21 | Negative concentration changed to positive value | Documented in QM Notes |

Table 5. Count of Matrix Changes between Import and DBF files

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source Table(s) | Field | Count | Import | DBF |
| Samples | MATRIX | 2 | DS | WH |
| 1 | FLOC | WH |
| 1 | MD | ON |
| 1 | MD | WH |
| 2 | Mousse | OL |
| 2 | Oil | OL |
| 128 | SL | SE |
| 108 | SO | SE |
| 31 | TB | OL |
| 1 | Water | OL |
| 4301 | water | WH |
| 5 | WH | DS |
| 1 | WH | GW |
| 258 | WH | PW |
| 1 | WH | SE |

# 

Appendix A. Data Not in QM

Response Scribe Data were not included in QM for the following reasons:

* The matrix in the Samples table was “air”.
* The data did not have chemistry results, including:
  + Monitoring Data
  + Samples with no associated results
  + Property information not associated with specific samples
  + Locations without associated samples or monitoring results
  + Rototoxicity samples and results.
* Data excluded by the QM Team because they were the following:
  + records for tentatively identified compounds (TICs),
  + sieve-grain analyses that were converted to percent soil types,
  + waste samples,
  + bioassay results,
  + calculated values,
  + repeated data,
  + had negative results
  + non-environmental samples,
  + results of toxicity characteristic leaching procedure (TCLP), or
  + had problems with the units or matrices

The 21 comma delimited (CSVs) files are listed and described as follows and listed in Table1:

1. ADEM - ADEM\_DW\_Sampling\_Analytical\_Monitoring.mdb
   1. 570 monitoring results in ADEM\_Monitoring\_Data. All non-null fields from the Location and Monitoring tables are included in CSV file.
2. ALECI - ALECI\_DW\_Sampling\_Analytical.mdb
   1. One sample without associated LabResults data in ALECI\_Sample\_No\_Data. All non-null fields from the Location and Samples tables are in included in CSV file. Note: SampleTime was null.
3. BP - BP\_DW\_Sampling\_Analytical.mdb
   1. All records imported into QM
4. CTEH - CTEH\_DW\_Monitoring.mdb
   1. 808,302 monitoring results in CTEH\_Monitoring\_Data. All non-null fields from the Location and Monitoring tables are included in CSV file.
5. CTEH - CTEH\_DW\_Sampling\_Analytical.mdb
   1. 26,580 monitoring results in CTEH\_S&A\_Monitoring\_Data. All non-null fields from the Location and Monitoring tables are included in CSV file.
   2. 300,426 air sample/results records in CTEH\_Air\_Data. Samples with a matrix of air were selected. All non-null fields from the Location, Samples, LabResults, SamplesAir, SampleTags and COC tables are included in CSV file.
   3. 2,307 samples without results in CTEH\_Samples\_without\_LabResults. All non-null fields from the Location and Samples tables are included in CSV file.
6. EPA - DW\_Reporting.mdb
   1. 102,828 monitoring results in EPA\_Monitoring\_Data. All non-null fields from the Location and Monitoring tables are in included in CSV file.
   2. 1,091 Property Information records in EPA\_PropertyInfo. All non-null fields from the PropertyInfo table are included in the CSV.
   3. 164,306 air sample/results records in EPA\_Air\_Data. Samples with a matrix of air were selected. All non-null fields from the Location, Samples, LabResults, SamplesAir, SampleTags and COC tables are included in CSV file.
   4. 34 samples without results in EPA\_Samples\_without\_LabResults. All non-null fields from the Location and Samples tables are included in CSV file.
   5. Seven locations without Samples or Monitoring in EPA\_Locations\_without\_Samples. All non-null fields from the Location table are included in CSV file.
   6. 242 toxicity sample/results records in EPA\_Tox\_Data. Results with units of “% survival” and results with units of “%” and null results were selected. All non-null fields from the Location, Samples, LabResults, SampleTags and COC tables are included in CSV file.
7. FLDEP - FLDEP\_DW\_Sampling\_Analytical.mdb
   1. All records imported into QM
8. LDEQ - LDEQ\_DW\_Sampling\_Analytical.mdb
   1. 17,570 air sample/results records in LDEQ\_Air\_Data. Samples with a matrix of air were selected. All non-null fields from the Location, Samples, LabResults, and SamplesAir tables are included in CSV file.
9. MSDEQ - MSDEQ\_DW\_Sampling\_Analytical.mdb
   1. Five locations without Samples or Monitoring in MSDEQ\_Locations\_without\_Samples. All non-null fields from the Location table are included in CSV file.
10. NPS - NPS\_DW\_Sampling
    1. All records imported into QM
11. TS - TS\_DW\_Monitoring.mdb
    1. 296,686 monitoring results in TSDW\_Monitoring\_Data. All non-null fields from the Location and Monitoring tables are included in CSV file.
12. USGS – USGSDW.mdb
    1. One sample without results in USGS\_Sample\_without\_LabResults. All non-null fields from the Location and Samples tables are included in CSV file.
13. ESI - EnvStds\_DW\_Sampling\_Analytical.mdb
    1. 618 air sample/results records in ESI\_Air\_Data. Samples with a matrix of air were selected. All non-null fields from the Location, Samples, and LabResults tables are included in CSV file.
    2. 110 samples without results in ESI\_Samples\_without\_LabResults. All non-null fields from the Location and Samples tables are included in CSV file.
    3. 4,300 toxicity sample/results records in CTEH\_Tox\_Data. Results with “tox-“ within the analyte name were selected. All non-null fields from the Location, Samples, and LabResults tables and the COC and description fields from the SampleTags table are included in CSV file.
14. NOAADW - NOAA\_DW\_Sampling.mdb
    1. 2,284 samples containing Rototox data in NOAADW\_RotoTox\_Data. All non-null fields from the Location, Samples and SamplesRototox tables are included in the CSV file.
    2. 18,693 samples without results in NOAA\_Samples\_without\_LabResults. All non-null fields from the Location and Samples tables are included in CSV file. These are samples without LabResults in the ESI database or samples without onboard lab results.
    3. 821 locations without samples in NOAA\_Locations\_without\_Samples. All non-null fields from the Location and table are included in CSV file.
15. Excluded from multiple databases by QM team. The reasons for exclusion are listed in Table 2.

Table 1. List and Explanations of CSV Files

| Provider | Count | Type | CSV File | Comments |
| --- | --- | --- | --- | --- |
| ADEM | 570 | monitoring results | ADEM\_Monitoring\_Data | All non-null fields from the Location and Monitoring tables |
| ADEM | 1 | sample without associated result | ALECI\_Sample\_No\_Data | All non-null fields from the Location and Samples tables are in included in CSV file. Note: SampleTime was null. |
| CTEH | 808,302 | monitoring results | CTEH\_Monitoring\_Data | All non-null fields from the Location and Monitoring tables are included in CSV file. |
| CTEH | 26,580 | monitoring results | CTEH\_S&A\_Monitoring\_Data | All non-null fields from the Location and Monitoring tables |
| CTEH | 300,426 | air sample/results | CTEH\_Air\_Data | Samples with a matrix of air were selected. All non-null fields from the Location, Samples, LabResults, SamplesAir, SampleTags and COC tables are included in CSV file. |
| CTEH | 2,307 | samples without results | CTEH\_Samples\_without\_LabResults | All non-null fields from the Location and Samples tables are included in CSV file. |
| EPA | 102,828 | monitoring results | EPA\_Monitoring\_Data | All non-null fields from the Location and Monitoring tables are in included in CSV file. |
| EPA | 1,091 | property information | EPA\_PropertyInfo | All non-null fields from the PropertyInfo table are included in the CSV |
| EPA | 164,306 | air sample/results | EPA\_Air\_Data | Samples with a matrix of air were selected. Non-null fields from Location, Samples, LabResults, SamplesAir, SampleTags and COC tables |
| EPA | 34 | samples without results | EPA\_Samples\_without\_LabResults | Non-null fields from Location and Samples tables |
| EPA | 7 | locations without samples or monitoring | EPA\_Locations\_without\_Samples | Non-null fields from Location table |
| EPA | 242 | toxicity sample/results | EPA\_Tox\_Data | Non-null fields from Location, Samples, LabResults, SampleTags and COC tables |
| LDEQ | 17,570 | air sample/results | LDEQ\_Air\_Data | Samples with air matrix. Non-null fields from Location, Samples, LabResults, and SamplesAir tables |
| MSDEQ | 5 | locations without samples or monitoring | MSDEQ\_Locations\_without\_Samples | Non-null fields from Location table |
| Total Safety | 296,686 | monitoring results | TSDW\_Monitoring\_Data | Non-null fields from Location and Monitoring tables |
| USGS | 1 | sample without results | USGS\_Sample\_without\_LabResults | Non-null fields from Location and Samples tables |
| ESI | 618 | air sample/results | ESI\_Air\_Data | Samples with air matrix. Non-null fields from Location, Samples, and LabResults tables |
| ESI | 110 | samples without results | ESI\_Samples\_without\_LabResults | Non-null fields from Location and Samples tables |
| ESI | 4,300 | toxicity sample/results | ESI\_Tox\_Data | Non-null fields from Location, Samples, and LabResults tables and COC and description fields from the SampleTags table |
| NOAA | 2,284 | samples containing Rototox data | NOAADW\_RotoTox\_Data | Non-null fields from Location, Samples and SamplesRototox tables |
| NOAA | 18,693 | samples without results | NOAA\_Samples\_without\_LabResults | Non-null fields from Location and Samples tables. Samples without LabResults in the ESI database or onboard lab results. |
| NOAA | 821 | locations without samples | NOAA\_Locations\_without\_Samples | Non-null fields from Location table |
| Multiple | 110,871 | mixed | Excluded\_by\_QM\_Team | Mostly waste samples, tentatively identified compounds or sieve sizes converted to percent soil types |

Table 2. Excluded by QM Team

|  |  |
| --- | --- |
| **Reason for No Import** | **Count** |
| Bioassay | 686 |
| Calculated value | 17 |
| Problem with Matrix or Units | 2437 |
| Repeated | 497 |
| Negative Value | 168 |
| Non-environmental | 284 |
| Sieve converted to % soil type | 11506 |
| TCLP | 118 |
| TIC | 18219 |
| Waste | 76939 |
| Total | 110871 |

Appendix B. Creating Import Files from Scribe Source Files

The import files to QM were created from the Scribe source files using the procedures described in this appendix. Two interim and up to seven final tables were produced from each source database. The final tables were:

1. Stations – mandatory table.
2. Samples – mandatory table.
3. Chem – mandatory table.
4. ChemQC – mandatory table.
5. EDD – mandatory table.
6. SampleTags – produced only if Scribe SampleTags table was populated. EPA and CTEH were the only providers that used this table.
7. COC – produced only if Scribe COC table was populated. EPA was the only provider that used this table.

The produced tables were created using queries and five linked Scribe tables from the most recent version of each provider Scribe database. The five linked Scribe tables were:

1. Location
2. Samples
3. LabResults
4. COC, and
5. SampleTags

In addition, five tables and two modules were created to map information from Scribe types and names to QM types and names, convert units, handle null values, and format dates and times. These five tables were:

1. Map\_Analyte\_Chemical
2. Map\_Matrices
3. Map\_QCTypes
4. Map\_SampleTypes
5. Conversions

and two modules were:

1. All\_Export\_Functions
2. LabRep.

The provider information was appended to copies of templates. The seven templates were:

1. Template\_Location\_to\_Stationwill be copied toStations
2. Template\_Samples\_to\_Samples
3. Template\_LabResults\_to\_Chem
4. Template\_LabResults\_to\_ChemQC
5. Template\_LabResults\_to\_EDD
6. Template\_SamplesTags
7. Template\_COC

Once the five tables have been linked and the seven copies of the templates have been produced, queries were run to make the two interim tables and append to the seven final tables.

The queries were run in the following order:

1. qry1\_Samples\_Plus: makes a version of the samples table with COCs appended.
2. qry2\_All\_Export: makes a table with all the fields needed to populate the mandatory five final tables.
3. qry3\_Station: appends data to the Stations table.
4. qry4\_\_Samples: appends data to Sample table
5. qry5\_EDD: appends data to EDD table
6. qry6\_Chem: appends data to Chem table
7. qry7\_ChemQC: appends data to ChemQC table
8. qry8\_SampleTags: makes a table with information from the Scribe SampleTags table.

The template, mapping and conversion tables, queries and modules are documented in the NOAA Deepwater Horizon sftp site at Deepwater\_Horizon\_Ext/Scribe\_to\_QM\_Mapping/FinalReport/Scribe\_to\_QM\_Import\_Template.accdb.

Notes on conversions from specific Scribe sources to import database files are listed as follows:

Provider: Alabama Department of Environmental Management (ADEM) - No QC, Sample Tag or CoC information was entered in this database.

Provider: Alabama EnviroChem Inc. (ALECI) – This database was created under the direction of ADEM. No QC, Sample Tag or CoC information was entered in this database. The field length of the Sample field was increased from 30 to 41 to accommodate 14 Sample IDs with field length greater than 30.

Provider: Center for Toxicology and Environmental Health, L.L.C. (CTEH) - LabReps were set as follows:

1. “R” and “X” qualifiers are set to be the last Lab Rep.
2. Sort by units so that mass per mass units, e.g., mg/kg, precedes similar mass per volume units, e.g., mg/l.
3. The Lab Reps were sorted by least to greatest MDLs after accounting for units. Records with no MDL were set to greatest Lab Reps. Note: accounting for units is important because 5,162 of the 41,650 duplicates have different units.
4. The Lab Reps were sorted by least to greatest reporting limits after accounting for units. Records with no reporting limit were set to greatest Lab Reps
5. If all the other conditions were the same, then the lower Lab Rep had the first Lab Sample ID sorted alphabetically.
6. If units, reporting limits and the Lab Sample ID were all the same, the Lab Reps were determined randomly.

“-9” was inserted for null MDLs and reporting limits.

Provider: United Stations Environmental Protection Agency (EPA) – for the Location table:

1. For LocationDescription, use LocationDescription + LocationZone
2. For LocationComment, use LocationComment + Location\_Image\_Path

For the Samples Table:

1. For Notes use, Remarks + Activity + “SubLocation: “ & SubLocation
2. Do not list Activity by itself because it’s only used for 14 samples
3. Use EventID for Project Name

For the LabResults table:

1. DVLevel is “VALX” because QAFlag and changes between lab qualifiers and final qualifiers indicate validation; however the level of validation is not known.
2. Bioassay results were removed by excluding all LabResults with the following analytes:  
   Neanthes arenaceodentata, Leptocheirus plumulosus, Americamysis bahia, Menidia beryllina, and Mysidopsis

Additional Notes:

1. SamplesWater table contains field measurements and observation for some water samples.
2. Because some samples have multiple CoCs and the analytes cannot be related directly to CoCs, COC numbers were removed from the Samples and EDD tables but can be linked to samples in the SampleTags and COC tables.

Provider: United States Geological Survey (USGS) – for Location table:

1. Used LocationDescription for LocationDescription
2. For LocationComment combined GPS\_Comment, LocationComment and Type
3. For Project Name, used Type

For Samples table:

1. For Notes combine Remarks and SampleDepthComment
2. Use Activity but not EventID (all EventIDs are “sampling”)

For LabResults table:

1. DVLevel – The QA\_Flag indicates that the analyses were not QAed. A value of UNK or NV may be accurate. UNK was selected.
2. Total or Dissolved was derived from an added field called Fraction, which reported values of Dissolved, Total, Recoverable, and Bed Sediment with some blank entries. In the exported field Total is entered as “tot”, Dissolved is entered as “dis”; all other entries are reported as blank.

Other Notes:

1. USGS created a separate table called QC. This table has 483 QC samples with 15,802 results. None of samples match samples in the Samples table. The data would require considerable reconfiguring to fit into the ChemQC table format including parsing the Analyte names, transforming qualifiers and results based on the qualifiers. This table if reconfigured would not affect the Chem table because all the QC data are in this table, not in the Samples or LabResults tables. If included it would affect the EDD and ChemQC tables.
2. USGS added a table called QualifierDesc as shown below:

| **Site\_No** | **AutoNumber** | **Qualifier** | **Qualifier\_Desc** | **Analysis** |
| --- | --- | --- | --- | --- |
| ALL | 1 | J | Estimated | All |
| ALL | 2 | U | Not detected | All |
| ALL | 3 | UJ | Not detected above the quantitation limit | All |
| ALL | 4 | R | Rejected | All |
| ALL | 5 | N | Tentatively identified | Organics |
| ALL | 6 | NJ | Tentatively identified; concentration is estimated | Organics |
| ALL | 7 | J+ | Estimated quantity; result may be biased high | Inorganics |
| ALL | 8 | J- | Estimated quantity; result may be biased low | Inorganics |

However, only U and J qualifiers are listed in the Result\_Qualifiers field and the Lab\_ Result\_Qualifiers field is blank.

1. The Lab Reps were set as followed:

* ug/kg is set before ug/l for same sample and chemical.
* The Lab Reps were sorted by least to greatest MDLs. Records with no MDL were set to greatest Lab Reps
* The Lab Reps were sorted by least to greatest reporting limits. Records with no reporting limit were set to greatest Lab Reps
* If all the other conditions were the same, then the lower Lab Rep had the first Lab Sample ID sorted alphabetically.
* If units, reporting limits and the Lab Sample ID were all the same, the Lab Reps were determined randomly.

1. Oxygen Chemical/analyte name was changed to DO when units were mg/l and changed to DOSAT when units were % saturation.

Provider: Environmental Standards Inc (ESI) - The Sample field length was increased from 30 to 40 to accommodate 267 long Sample IDs.

1. Provider: National Oceanic and Atmospheric Administration (NOAA) - The NOAADW database has 5215 unique sample IDs. Some match to more than one ESI sample ID because ESI split their samples among labs and sometime names them differently depending on the analyses performed. Therefore, the 5215 NOAA IDs matched to 5,758 ESI IDs.
2. The Sample and ESI sample field lengths were increased from 30 to 40 to accommodate the 914 ESI Sample IDs and 42 NOAA Sample IDs with length greater than 30 characters.

Provider: ESI/NOAA – Steps to combine the Location and Samples tables from NOAADW with the LabResults table in Env\_Stds\_2 were as follows:

1. Import Extend\_Ref with 3702 matched samples from JAG report analysis
2. Import ESI Location, Samples and LabResults tables and rename E\_Loc, E\_Sam and E\_LR.
3. Import NOAADW Location and Samples tables and rename N\_Loc, N\_Sam.
4. Import “Within\_Specs “that lists matching Samp\_Nos for water samples at depth less than 200m deep
5. Add fields that are in Extend\_Ref to Within\_Specs and save as table Shallow\_Water\_Samples
6. Combine Shallow\_Water\_Samples and Extend\_Ref with Matrix field added and call All\_Water\_Samples.
7. Remove 37 duplicate NOAA Samp\_No and name resulting table, All\_Water\_Samples\_37Dup\_NOAAIDs\_Removed.
8. Create table of possibly matching ESI 1,120 sediment samples by filtering with matrix of sediment, date range of 9/19 to 10/27/10 and EventIDs of DSS, OPANEX, and GYRE. Call resulting table E\_Sed\_Sam
9. Create table of possibly matching NOAA 3,187 sediment samples by filtering with matrix of sediment, ProjectIDs of Ocean Veritas, Ryan Chouest and Gyre, and Samp\_No starting with “SE” Call resulting table N\_Sed\_Sam
10. Add core information from PastSedData table to N\_Sed\_Sam table
11. Link E\_Sed\_Sam Samp\_Nos to N\_Sed\_Sam ESI\_Samp\_Nos to match 943 of 964 ESI\_Samp\_Nos and 943 of 1152 E\_Sed\_Samp Samp\_Nos. Name the resulting table Sed\_First\_Match
12. Link the remaining 177 unmatched ESI sediment samples with the NOAA sediment samples with the vessel, sample date and latitude and longitude to three decimal places. Select the best Samp\_No matches for the 72 ESI sediment samples that met these match criteria. The remaining 105 ESI sediment samples will not be matched.
13. Combine ESI Samp\_No list from tables All\_Water\_Samples\_37Dup\_NOAAIDs\_Removed and Sed\_Final\_Match to create the list of the 5,758 ESI samples matched to NOAADW data. This table is called Matched\_ESI\_Samples.
14. Create the table of unmatched ESI Samp\_NOs called Unmatched\_ESI\_Samples with 9,049 Samples of which the Location of 233 samples is “Not\_Provided” and 8 samples have an air matrix. These samples will not be carried forward.
15. Create new database called NOAADW\_ESI\_Matched Use NOAADW Location and Samples tables and Sample ID cross reference table to create a table called Location with 782 distinct locations. The table includes fields for Vessel and CruiseID.
16. Use ESI and NOAADW Samples table, table with additional sediment data (PastSedData) and and Sample ID cross reference table to create a Samples\_Input and Samples\_Input table table. The Sam\_Depth and SampleTime defaulted to the NOAADW values. If NOAADW did not have a value the ESI value was used. Use ESI LabResults and cross reference table to create the LabResults table with 460651 records.
17. Create new database called NOAADW\_ESI\_UnMatched.
18. Use ESI Location and Samples tables and Sample ID cross reference table to create a table called Location with 3,738 distinct locations with the following locations without coordinates not listed: “Not\_Provided", "Not\_Recorded", "Equip\_Blank", "Storage\_Blank", and “Test Sample".

Provider: The Commonwealth Scientific and Industrial Research Organisation (CSIRO) - Notes on formatting the Ryan Chouest onboard data into Scribe format follows:

Onboard GCMS and monitoring information was found in OsbLogs\_Data.mdb.

Method detection limits and reporting limits for some analytes were found in BP\_Data\_Validation\_Project\_Report\_08\_07\_2011\_final\_27\_Jul\_2011.pdf, APPENDIX E: Method Limit of Reporting (LOR) Calculation.

The qualifiers used:

* U - Compound Not Detected
* V - Overloaded and possibly not a water extract
* Z - Detected but not Quantified (too small)

The CSIRO tables have no clear equivalent for Scribe Location field. The CSIRO values were mapped to the Location fields as follows:

* For those samples collected at a cast, the Location is the Cast\_ID
* For those samples not collected at a cast, the Location is the SampleBottleID.

The query and table are qry\_LocationMaker and LocationMaker, respectively.

1. The autonumber ID is converted to text format prior to joining with IMPFILEREC. [↑](#footnote-ref-1)
2. WatSample\_Difference includes samples with water, tar, and oil matrices. [↑](#footnote-ref-2)