

EPA Region 2 Superset Conversion Tool

The purpose of this document is to provide instruction and guidance on how to process EDD submittals that have been validated by the CLP group. The output from the CLP DAT Editor is in a single file EDD called a superset spreadsheet. The superset files combine the EPAR2SMP, EPAR2TRSQC and EPAR2BAT sections of the EPA Region 2 format into one file or one EDD superset.

The superset files are hard to manually manipulate into the EPA Region 2 three files format. A conversion tool has been added to the EPA Region 2 format which will load a superset single file EDD into the EPA Region 2 three files format in EQuIS Data Processor (EDP). All files run through the DAT Editor after October 2009 will be in an Excel spreadsheet which can be translated using the procedure described below. If you have files prior to this date, please review the section on historical DAT Editor files.

The EQuIS Data Processor (EDP) can be downloaded from this link:

<http://www.earthsoft.com/wordpress/products/edp/edp-format-for-epar2/>

This tool has limitations which are explained in this document. Many of the limitations can only be resolved by the data provider. The tool does not fix or change data. This document outlines where there may be issues with the data that will need to be reviewed by the data provider.

Data Provider:

When Data Providers receive a Superset EDD file from the CLP lab it should be an Excel spreadsheet. Open the Excel spreadsheet, the number of rows in the spreadsheet should equal the number of rows in the EPAR2TRSQC section once the EDD has been loaded into EDP (minus the header row). Data Provider has to put a “#” sign to indicate the first row is the header row.

Before reviewing the EDD data, please modify the file name by adding “clp” in front of the file extension. For example, if the EDD file has the file name of “G3784.EDD02041.B3F76.xls”, data provider has to add “clp” in front of the file extension as part of the file name, so the file name will be changed to “G3784.EDD02041.B3F76.clp.xls”.

EQuIS provided a video instruction on converting CLP EDD files into EPAR2 format, this video can be viewed from this link:

<http://www.screencast.com/t/NzI10WRhNzAt>

Recommended Review

Scroll down through the EDD and review the values in the sys_sample_code column. Make a note of the following:

- Some laboratory QC samples may not have a populated sys_sample_code
- Laboratory Control Samples (LCS) do not have a unique sys_sample_code

- The sys_loc_code column is null.

The EPA Region 2 format requires a sys_sample_code and it must be unique.

Populating Null or Blank sys_sample_codes

The CLP conversion tool will populate a null or blank sys_sample_code by concatenating the values in the sample_name, sample_date and sample_matrix_code columns. For the majority of the laboratory QC rows these 3 fields are enough to create a unique sys_sample_code. However, it is important to review these concatenated values.

Creating Unique sys_sample_codes

The CLP laboratory deliverable does not contain a unique value for Laboratory Control Samples (LCS). Typically the sys_sample_code will be populated with the value, "LCS". The conversion tool attempts to make this value unique by concatenating the same fields as with the null or blank sys_sample_code (sample_name, sample_date and sample_matrix_code).

The values that were originally submitted in the Station_ID column are mapped to the sys_sample_code column in the Superset EDD.

1. Data Provider:

The Data Provider has to provide the name and the code of the company that requested the laboratory analysis and is responsible for reporting the results to EPAR2.

- 2. Test_batch_ID:** For CLP data, the Data Provider has to report the "Test Batch ID". The "Test Batch ID" is a unique identifier for all lab batches. Please provide this identifier for any "Test Batch Type" including "Prep", "Analysis" and "Leach".

3. Interpreted_qualifiers: Data Provider has to check the following three conditions:

1. If both the "Lab_qualifiers" and "Validator_qualifiers" is populated, the "Validator_qualifiers" should be listed as the final Interpreted_qualifier and should be copied to the "Interpreted_qualifier" column
2. If only the "Validator_qualifiers" is populated, use "Validator_qualifiers" as the value in the interpreted_qualifier column.
3. If only the "Lab_qualifier" is populated, the Data Provider has to look at the description in the "Lab_qualifier", and compare the EPA Region 2 description of the same qualifier. If the description does not match, they need to find the qualifier description and code that closely matches their description of their "Lab_qualifiers" to be used in the "Interpreted_qualifiers".

Example for #1:

Lab_qualifiers	Validator_qualifiers	Interpreted_qualifiers
T	E	E
O-	J-	J-

Example for #2:

Lab_qualifiers	Validator_qualifiers	Interpreted_qualifiers
	E	E
	J-	J-

Example for #3:

Lab_qualifiers	Validator_qualifiers	Interpreted_qualifiers
T		E
O-		J-

Description of Example #3:

In this example, the Data Provider has to check the description of T, and if the description of T is closely matched with R2 description in the qualifier column, the Data Provider has to use the code from R2 that has the same or closely match description. In this case, we put “E” to the “Interpreted_qualifier” column because it has the same description as R2. Similar to “Lab_qualifiers” = “O-”, the description from the lab for “O-” is estimated on the low side, which is the same as the description in R2 but the standard code for R2 is “J-”, so we put “J-” to the “Interpreted_qualifiers” column.

4. Cas Number (Cas_rn):

Data Provider has to use the chemical name and cas number that are in the Valid Values Reference Table, if any cas number or Chemical Name are not in the Table, please provide the valid values and make a request to EPA Region 2 that they add the values to the valid values list.

5. Check EDD

For the CLP Superset data, the data provider is responsible to know the EPA Region 2 EDD format, and review the valid values reference table and documentation in order to understand how to prepare the dataset. Each EDD has to go through EQuIS EDP to check for errors (EQuIS EDP can be downloaded from EPA Region 2 EDD web site). The following are an example of the errors that the Data Provider has to modify:

- 1) data_provider: please provide the code of the data provider, codes can be found in the Valid Values Reference Table A-23.
- 2) sample_matrix_code: Please use the codes in the Valid Values Reference Table A-1 Matrix

- 3) **sys_loc_code:** Please provide the location data, including the location code for each normal environmental sample. (Please check the “Comprehensive EDD Specification Manual” in order to prepare the Data Provider, Site, and Location files.)
- 4) **chain_of_custody:** Please modify the chain of custody column if it exceeds the field length. This field allows 40 characters. (For definition, please check the format documentation)
- 5) **sampling_company_code:** This field is required in the EPA Region 2 format and is not provided in the CLP DAT EDD’s. Please populate this column, provide a code that does not exceed 10 characters.
- 6) **lab_matrix_code:** Please use the codes in the Valid Values Reference Table A-1 Matrix
- 7) **Preservative:** Please check the Valid Values Reference Table A-27
- 8) **Interpreted_qualifiers:** please check the Valid Values Reference Table A-10 for valid values, and description of the Interpreted Qualifiers from the above.
- 9) **Test_type:** please check the Valid Values Reference Table A-25
- 10) **Test_batch_id:** Please read the description of Test_Batch_ID from the above.

6. Sign and Submit

After data files have been loaded and all of the errors have been resolved, the Sign and Submit tool can be used to facilitate submittal of data to EQUIS Enterprise EDP. The Sign and Submit tool packages the data files with the correct naming convention which allows easy submittal of data packages. This feature requires a user name and password that may be obtained from the EQUIS Database Administrator. The instructions for using the Sign and Submit can be summarized in the following steps (Detailed instructions can be found on the EPA Region2 EDD web page):

1. Select Sign and Submit from the Application Menu
2. Enter your User Name and Password as well as the Program code that applies to the data package being submitted. The Registry ID will be automatically populated based on the selected Program Code.
3. Click the Submit button

4. You will be prompted for the location where you would like to save the file. The Sign and Submit feature will name the file correctly and save to a zipped file with the current date, a period, Registry ID, a period, and the Format file name used to create the EDDs.

(Example file name: ‘ 20100205.RegID123.EPARegion2EDD.zip’)

5. Select Save

6. Click ‘OK’ when the “EDD Signed and Saved” Verification Window comes up.

7. For each file, except the base map file, the name of the files are named after the section that was populated and have the extension of “csv” if the file was saved as a comma-delimited file or “txt” if the file was saved as a tab-delimited. For example if we made a package of the 3 sections populated by this CLP translation the zip would contain 4 files, EPAR2SMP_v2.csv, EPAR2TRSQC_v2.csv and EPAR2BAT_v2.csv. An additional user file is included (*.usr) as the fourth file. The Sign and Submit process packages or zips the files, the zipped package must be named using the file name as the site name, followed by the submittal date of the EDD with the format for the date being YYYYMMDD. The second part of the file name is the 12-character alphanumeric EPA ID for the facility under investigation. The third part of the file name refers to the EDD file format for the file being submitted. For more information on the user file refer to the Comprehensive manual.

Historical DAT Editor Files

If you have files that were created by the DAT Editor prior to October 2009 you can still process these files with the translation tool in the EPA Region 2 format with some minor modifications.

Review the file you have received. If it is not an Excel file, load it into Excel and save as an Excel spreadsheet. Review the column order and headers. It is important to use the same order and headers as defined below in Table A. Once the order and headers have been modified, follow the instructions above to process the file.

If there are any questions regarding the EDD, please send email to EPA Region 2
Region2_equisedd@epa.gov

Table A

Assumes the valid values and field definitions are consistent with the EPA Region 2 format unless noted.
 X = indicates fields where no data will be provided in CLP DAT file, field will remain null.

Field Order	Field Name	Inorganic not provided OR NULL	Organic not provided OR NULL
1	DATA_PROVIDER	X	X
2	SYS_SAMPLE_CODE	For all field samples = Station Location For Lab Matrix Spike and Lab Matrix Spike Duplicate samples = Lab Sample ID	For all field samples = Station Location For Lab Matrix Spike and Lab Matrix Spike Duplicate samples = Lab Sample ID
3	SAMPLE_NAME	EPA Sample Number	EPA Sample Number
4	SAMPLE_MATRIX_CODE	Limited to Water and Soil in electronic deliverable (ASF) submitted by the labs	Limited to Soil, Sediment, Water in electronic deliverable (EDD) submitted by the labs
5	SAMPLE_TYPE_CODE		SD = Lab Matrix Spike Duplicate considered as separate from spike BS = LCS (Lab Control Sample)
6	SAMPLE_SOURCE		
7	PARENT_SAMPLE_CODE	For Lab Matrix Spike, Post-Digestion Spike and Lab Duplicate samples = Station Location of Parent Sample	For Lab Matrix Spike, Post-Digestion Spike and Lab Duplicate samples = Station Location of Parent Sample
8	SAMPLE_DEL_GROUP		
9	SAMPLE_DATE	Sampling Time will not be reported in this field because of the existing DAT Editor settings.	Sampling Time will not be reported in this field because of the existing DAT Editor settings.
10	SYS_LOC_CODE	X	X
11	START_DEPTH	X	X
12	END_DEPTH	X	X
13	DEPTH_UNIT	X	X
14	CHAIN_OF_CUSTODY		
15	SENT_TO_LAB_DATE		
16	SAMPLE_RECEIPT_DATE		
17	SAMPLER		
18	SAMPLING_COMPANY_CODE		
19	SAMPLING_REASON	X	X
20	SAMPLING_TECHNIQUE	X	X
21	TASK_CODE	X	X
22	COLLECTION_QUARTER	X	X
23	COMPOSITE_YN	Current default value = N	Current default value = N
24	COMPOSITE_DESC	X	X
25	SAMPLE_CLASS	X	X
26	CUSTOM_FIELD_1	X	X
27	CUSTOM_FIELD_2	X	X
28	CUSTOM_FIELD_3	X	X
29	COMMENT	X	X
30	BREAK_1	X	X

31	SYS_SAMPLE_CODE	Same as above	Same as above
32	LAB_ANL_METHOD_NAME		
33	ANALYSIS_DATE		
34	TOTAL_OR DISSOLVED	X	X
35	COLUMN_NUMBER	X	X
36	TEST_TYPE	X	Defaulted to 'ANALYSIS'
37	LAB_MATRIX_CODE	Same values as in SAMPLE_MATRIX_CODE field	Same values as in SAMPLE_MATRIX_CODE field
38	ANALYSIS_LOCATION		
39	BASIS		
40	CONTAINER_ID	X	X
41	DILUTION_FACTOR		
42	PREP_METHOD		
43	PREP_DATE		
44	LEACHATE_METHOD	X	X
45	LEACHATE_DATE	X	X
46	LAB_NAME_CODE	CLP Laboratory Code	CLP Laboratory Code
47	QC_LEVEL		
48	LAB_SAMPLE_ID		
49	PERCENT_MOISTURE		
50	SUBSAMPLE_AMOUNT	Soil Sample Aliquot Weight and Water Sample Aliquot Initial Volume	Sample Aliquot Amount
51	SUBSAMPLE_AMOUNT_UNIT	Sample Aliquot Weight/Volume Units	Sample Aliquot Amount Units
52	ANALYST_NAME	X	
53	INSTRUMENT_ID		
54	COMMENT	Will be populated with column_number value. The name of the column should remain as "COMMENT"	Will be populated with column_number value. The name of the column should remain as "COMMENT"
55	PRESERVATIVE		
56	FINAL_VOLUME	Sample Aliquot Final Volume	Sample Aliquot Final Volume
57	FINAL_VOLUME_UNIT	Sample Final Volume Units	Sample Final Volume Units
58	CAS_RN		
59	CHEMICAL_NAME		
60	RESULT_VALUE	Validated Result	Validated Result
61	RESULT_ERROR_DELTA	X	X
62	RESULT_TYPE_CODE		
63	REPORTABLE_RESULT		
64	DETECT_FLAG	Current default values: N if Lab Qualifier = U Y if Lab Qualifier not equal to U	Current default values: N if Lab Qualifier = U Y if Lab Qualifier not equal to U
65	LAB_QUALIFIERS		
66	VALIDATOR_QUALIFIERS		
67	INTERPRETED_QUALIFIERS	X	X
68	ORGANIC_YN	Current default value = N	Current default value = Y
69	METHOD_DETECTION_LIMIT		
70	REPORTING_DETECTION_LIMIT	Adjusted Contract Required Quantitation Limit (CRQL)	Adjusted Contract Required Quantitation Limit (CRQL)

71	QUANTITATION_LIMIT	Unadjusted Contract Required Quantitation Limit (CRQL)	Unadjusted Contract Required Quantitation Limit (CRQL)
72	RESULT_UNIT		
73	DETECTION_LIMIT_UNIT		
74	TIC_RETENTION_TIME	X	
75	RESULT_COMMENT	X	X
76	QC_ORIGINAL_CONC		
77	QC_SPIKE_ADDED		
78	QC_SPIKE_MEASURED		
79	QC_SPIKE_RECOVERY		
80	QC_DUP_ORIGINAL_CONC	X	
81	QC_DUP_SPIKE_ADDED	X	
82	QC_DUP_SPIKE_MEASURED	X	
83	QC_DUP_SPIKE_RECOVERY	X	
84	QC_RPD		
85	QC_SPIKE_LCL		
86	QC_SPIKE_UCL		
87	QC_RPD_CL		
88	QC_SPIKE_STATUS		X
89	QC_DUP_SPIKE_STATUS	X	X
90	QC_RPD_STATUS	Lab Duplicate RPD Qualifier	X
91	VALIDATION_LEVEL	X	X
92	BREAK_2	X	X
93	SYS_SAMPLE_CODE	Same as above	Same as above
94	LAB_ANL_METHOD_NAME	Same as above	Same as above
94	ANALYSIS_DATE	Same as above	Same as above
95	TOTAL_OR DISSOLVED	X	X
96	COLUMN_NUMBER	X	X
97	TEST_TYPE	X	Defaulted to 'ANALYSIS'
98	TEST_BATCH_TYPE	X	X
99	TEST_BATCH_ID	X	X
100	CASE	Case Number (Field required for DAT Editor)	Case Number (Field required for DAT Editor)
101	SAMPLE_TIME	Sampling Time (Field required for DAT Editor)	Sampling Time (Field required for DAT Editor)
102	FRACTION	Not Applicable. Field will remain blank. This is not the same at the EQUIS fraction field	Not Applicable. Field will remain blank. This is not the same at the EQUIS fraction field
103	PH	Info not available. Field will remain blank.	Info not available. Field will remain blank.