Standalone EQuIS Data Processor (EDP)
User Guide

2008-04-25

Contents

1.0 Introduction

The EQuIS Data Processor, or EDP, is today's answer to the many data quality issues that plague data managers. The latest in EarthSoft's family of data checking software, EDP sets a new standard for simplicity in data quality management while adding a host of new features and capabilities that allow the user the utmost in data checking flexibility.

There are two desktop modes for using the EQuIS Data Processor; Standalone EDP , Professional EDP. Standalone EDP is used primarily by data providers such as laboratories or field staff to check data qualitiy proir to data submittal. Professional EDP is linked directly to the EQuIS 5 database and can be used by EQuIS power users to check data quality and then load the data into the EQuIS database. This manual only address the use of Standalone EDP

EDP checks all of the following data quality issues:

- Required Fields
- Field Lengths
- Data Types
- Valid Dates
- Reference Values
- Duplicate Rows
- Range Checking
- Orphan Rows

2.0 Standalone EDP

The Standalone EQuIS Data Processor, or Standalone EDP, is used to check data without having access to the EQuIS 5 database. All of the functionality found in Standalone EDP is also found in Professional EDP. The following Standalone EDP sections will outline the features found in both Standalone EDP and Professional EDP.

All modes of EDP support formats for many fields of environmental data observation and acquisition. From the familiar "4-File" analytical laboratory format to field activities such as drilling and sampling to automated surface water sampling, EDP can check data files quickly and easily.

You need to know the EQuIS format or EDD format that your data resides in, what reference values you will use to check the data, and if there is an enumeration file associated with your format file.

There are several EQuIS formats included with the EQuIS Data Processor software. The appropriate format you will need to use is determined by the type of data to be imported. After determining the EDD format, you are ready to begin. For this exercise, you will be using the "4-file" format (e.g. EFWEDD format).

Start Standalone EDP by selecting Programs > EarthSoft > EQuIS Data Processor from the Start menu. If you are running the EQuIS Data Processor in evaluation mode, click the OK button on the EQuIS Data Processor Evaluation screen. The EQuIS Data Processor will open to the main window.

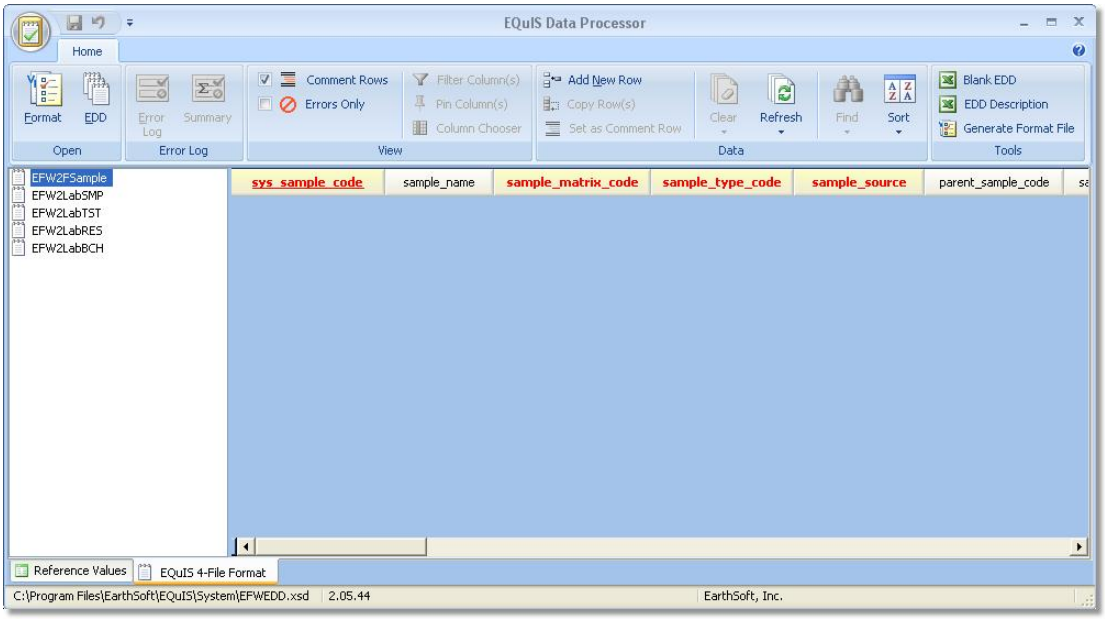


Figure 1: Standalone EDP with the EQuIS 4-File Format open

This tutorial uses example data files that are located in the C:\Program Files\EarthSoft\EQuIS\Tutor\EDD Tutorial\EFWEDD_WithErrors folder.

2.1 Format File

The Format File is the essence of data checking in EDP. The Format File contains the definitions for each individual section belonging to the EDD format. For example, the EFWEDD format file contains the laboratory "4-File" sections EFW2FSample, EFW2LabSMP, EFW2LabTST, EFW2LabRES, and EFW2LabBCH.

There are three or four files that make a format file; the format definition file (xsd) file, custom handler (vb) file, the enumeration (enum) file and the reference values (rvf) file. Together these files determine how your data will be checked.

The XSD file contains the mapping and definition for the sections of the format file, such as; the mapping of the fields in the format to the data table and fields in the EQuIS database, required fields, primary keys, field length, field description and field type (numeric, text, etc.)

The vb file contains the custom handling and business rules that apply to the format such as the analysis date cannot proceed sample date or reportable_result cannot be ‘Yes’ where the lab_qualifier is E, G, P, or R. Another example of a business rule is if a certain sample type is encountered then another field must be populated. For example if you have a sample_type of ‘FD’, short for field duplicate, then the parent sample field must be populated to avoid an orphan quality assurance/quality control sample.

The enum file is an optional file that allows EDP to enforce a set of lookup values similar to a reference table (rt.) lookup. This allows EDP to control the values for specific fields that are not linked to a reference table in the EQuIS database. The rvf is created as an export from EQuIS Professional, while the enumeration file is created and maintained manually using any text editor. Unlike the .rvf, the enumeration file is used in EQuIS Professional EDD as well as standalone EDP. It should also be noted that the enum file may also be used to narrow a list of reference values to a smaller list. For example, typically a unit field such as elevation units would be linked to the rt_unit table which may contain 100 or more units. However, it may make sense to limit the elevation units to “meters” and “feet” since those are the only valid units for this type of data.

The rvf is created as an export from EQuIS Professional reference tables. This file is only needed when running standalone EDP and it allows EDP to check reference values remotely against values established in EQuIS Professional.

Format Files may be encrypted or unencrypted, and typically include both the format definition file (.xsd or .xse) and a custom handler file (.vb or .vbe). The unencrypted format files may be modified to meet clients specific needs, however these modifications are not supported under the EarthSoft Maintenance agreement. If modifications are made to these formats, it is highly recommended that clients rename the formats to include their company name. The formats are written in XML

and may be viewed using a text editor. Many free software applications are available online for viewing and editing XML files. Formats that are unencrypted do not require a license key.

A list of the Standard Format Files that are included with all installations of EDP can be found in Appendix A. A list of the Alternate Format Files that may downloaded with any installation of EDP can be found in Appendix B.

2.2 Opening a Format File

The first time EDP is launched, a format file must be selected before a data file can be loaded.

To open a Format File:

1. Click on the Format button  (located in the Open group on the Home tab).

   NOTE: The default location for system formats is ~\Program Files \EarthSoft\EQuIS\System. The default location for custom formats is ~\Program Files \EarthSoft\EQuIS\Formats.

2. Select the EFWEDD.xsd format file and click "Open".

Once the format file has been selected and successfully opened, the format's sections will be listed in the left hand portion of the EDP window. If prompted to open the reference file, select "refvals.rvf".

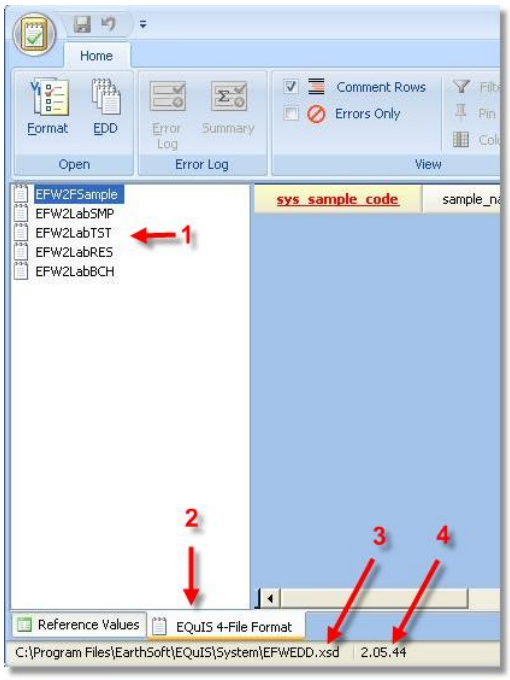


Figure 2: EDP Standalone Legend

(1) – The sections of the format
(2) – The Format Tab
(3) – The directory path to the current format
(4) – The version of the format

Users can identify what format file and the version of the format file has been loaded by selecting the Format Tab. The path to the format file and the version are displayed in the lower left corner. It is recommended that all EDP users check this path and version prior to each use of EDP.

Now that a Reference Value File and Format File have been successfully opened, EDP is ready to load and check a data file.

2.3 Understanding the Format

Standard EarthSoft formats were designed with tools that are available to assist users in understanding the format's requirements. These tools include; tool tips, color coded column headers, and drop down boxes that enforce reference values.

In each format file provided by EarthSoft, there are tool tips in the column headers. To access these tool tips, simply hold your mouse over the column headers name. These tool tips may be used to assist the user in determining what values should be used to populate each field as well as determining which reference value table is used to populate the field.

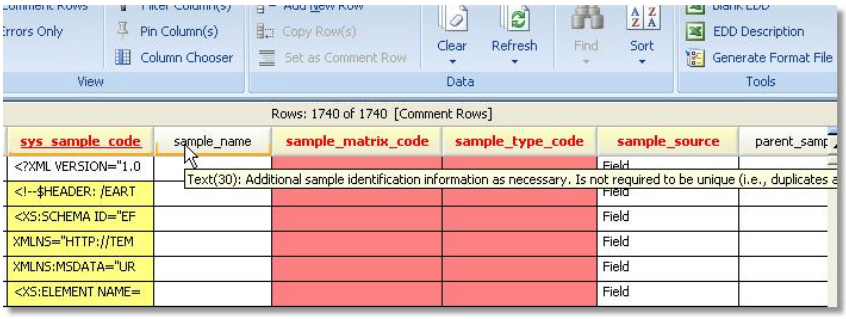


Figure 3: EDP Tooltip Example

Another aid in understanding a format is the color coding that is used to indicate what each column represents. These defaults are:

- **Red** – Column headers with red font (default color) are required fields.
- **(Underlined) Red** – Required primary key fields are identified by underlined red font column header. Recall that primary key fields are required and must be unique.
- **Blue** – Reference value fields are identified using a default blue font. It should further be noted that required fields may also be reference value fields, when this occurs the column header will appear in a red font.

The default font colors may be modified in the Options Tool in EDP. For details on making these modifications see the Options section of this manual.

An additional tool that has been provided to assist users are drop down lists that appear when the curser is placed over any Reference Value or Enumeration Value field. These drop-down lists include the approved reference values from the reference table indicated in the tool tip explained above or from the enum file. If an error indicating a ‘Missing Reference Value’ exists, one of the values from the drop down may be selected from the lists to resolve errors.

2.4 Reference Value File

The Reference Value file specifies what reference values, or valid values, may be allowed in the data file. It is simply a lookup list that says, in effect, these are the values which you may choose from for a particular field.

When EDP starts and a Reference Values File is selected, the "Reference Values" tab shows the various tables and values defined within those tables. Values displayed are not editable in the Reference Values Tab

In order to assist clients in loading the appropriate set of reference values for each format, users do not have to manually open the reference value file. Rather, each time a user opens a format file EDP will look in the same folder as the format file for a file with the extension .RVF (the reference value file). If a file is present, it will be

opened and used. If an *.RVF file is not located or if there are more than one *.rvf in the same directory as the Format File, then the user will be prompted to browse to the appropriate *.RVF file.

For example, if a user loads the 4-file format (C:\Program Files\EarthSoft\EQuIS\System\EFWEDD.XSD), EDP will check the C:\Program Files\EarthSoft\EQuIS\System\ folder for an *.RVF file. If the file is found, it will be loaded. If not, then the user will be prompted to open another *.RVF file of their choice.

If multiple *.RVF files exist in the \\EarthSoft\EQuIS\System\ folder, the user will be prompted to select which *.RVF file they would like to use. This is useful when laboratories and other data providers are supporting multiple clients with different reference values files.

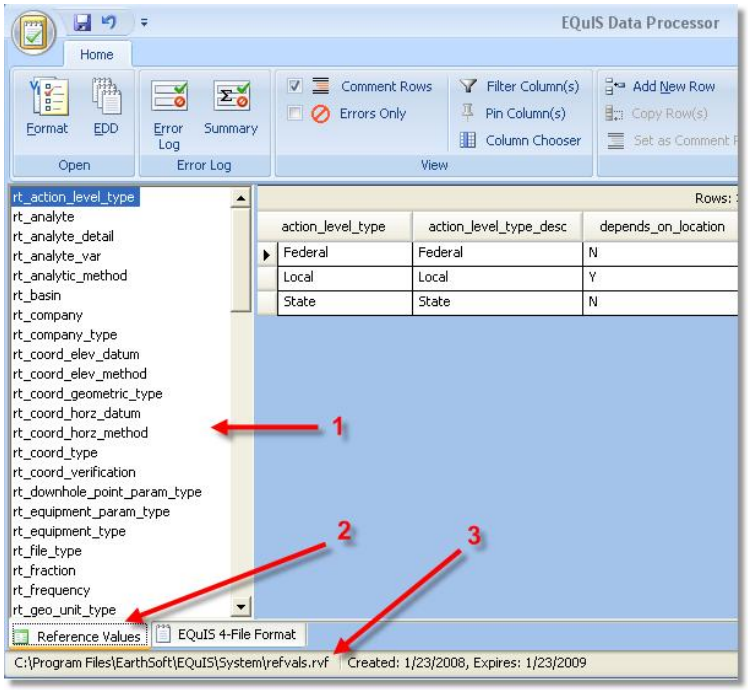


Figure 4: Reference Values window showing a list of reference values tables.

(1) – Reference Tables
(2) – Reference Values Tab
(3) – Directory path to current format

Therefore, when using EDP, it is important to verify that the correct Reference Value File is being used:

The tutorial data files that you will use will be checked using the 4-file format (EFWEDD.xsd), which is located in the C:\Program Files\EarthSoft\EQuIS\Tutor\EDD Tutorial\EFWEDD_WithErrors folder.

- Open Windows Explorer and browse to the C:\Program Files\EarthSoft\EQuIS\System\ folder. Locate the associated *.RVF file that will be loaded when the EFWEDD.xsd format file is loaded into EDP.

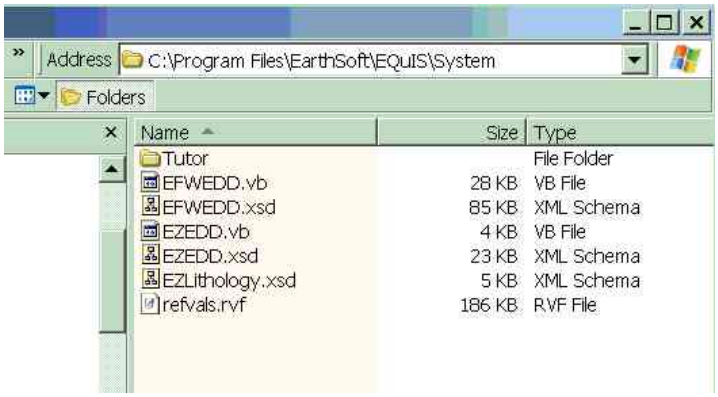


Figure 5: The System directory contains one *.RVF file associated with multiple EDD formats.

Users can identify the reference value file that has been loaded by selecting the Reference Values tab. The path to the reference value file is displayed in the lower left corner. It is recommended that Standalone EDP users check this path prior to each use if multiple *.RVF files exist.

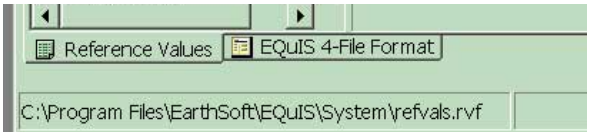


Figure 6: Directory path for reference values, recommended that Standalone EDP users check this path prior to each use.

Now that a Reference Values File and Format File have been loaded, EDP is ready to check data files.

2.5 Data Files

The data files, often referred to as Electronic Data Deliverables (EDDs), are the files that contain the data to be checked. The EDDs must be in an electronic, tabular format that adheres to the layout and rules specified by the Format File.

The data file may be a tab-delimited (.txt) or comma -separated (.csv) file, Excel files (.xls), Access databases (.mdb) as well as zipped EDD files (.zip) may also be loaded. Some formats maybe set up to accept an XML formatted dataset.

2.5.1 Loading Individual Data Files

Individual Data files can be loaded into the EQuIS Data Processor by highlighting the section to load the data file into and then using one of the following methods:

- Click Open > Data File from the Application menu (the  button)
- Right-click a section and select “Load Data File”.

When multiple sections exist within a given Format File, the data file selected is loaded into the currently selected section. By default when EDP is opened, the first section of the format is highlighted. If the users selects File > Open immediately after opening a format file, the data will be loaded into that section of the format regardless of the file’s content. For example, in the EFWEDD (4-File) Format File, the section EFW2FSample is the first section listed. A data file opened through the File menu will be loaded into this section regardless of whether it is appropriate for this format or not if another section was not selected.

To open a data file:

1. Select the EFW2FSample section.
2. Select the Application menu > Open > Data File.

   An alternate method of opening a data file is to right-click on the desired section and select "Load Data File" from the popup menu.

3. Browse to the C:\Program Files\EarthSoft\EQuIS\Tutor\EDD Tutorial\EFWEDD_WithErrors folder and select the EFW2FSample.SDG.txt file.
4. Click "OK".
5. Load the remaining data files in the C:\Program Files\EarthSoft\EQuIS \Tutor\EDDTutorial\EFWEDD_WithErrors folder to the appropriate section using the previous instructions.

The data check occurs in the background while the data file is being loaded. After all of the data files have been loaded, you will review any errors that EDP has detected.

To ensure that all of the business rules are properly processed by EDP it is recommended that the Data Files are loaded in the order that the sections appear in the format. This order represents the Parent/Child Relationship between the data in the data files. For example, EFW2FSample should be loaded before EFW2LabBCH. The EFW2LabBCH data file should be loaded before the EFW2LabTST data file and the EFW2LabRES data file should be loaded before the EFW2LabRES data file. If the files are not opened in order, it is recommended that the user select the refresh button  frequently while correcting data errors.

2.5.2 Load and Append Multiple Data Files

When loading data files, it is also possible to load multiple data files into the same section of the format file in the same “EDP session”. You can check multiple data files simultaneously in EDP by using the “Load and Append” utility. For example, if you receive multiple data files that include Field Sample data, you may load all of

these files into the EFW2FSample section of the format. Note that this functionality is only enabled once a data file has already been loaded. To do this:

Right-click on the EFW2FSample format name

1. Select “Load and Append Data File”.
2. Browse to the C:\Program Files\EarthSoft\EQuIS\Tutor\EDD Tutorial\EFWEDD_WithErrors folder
3. Select the EFW2FSample.append.txt file.

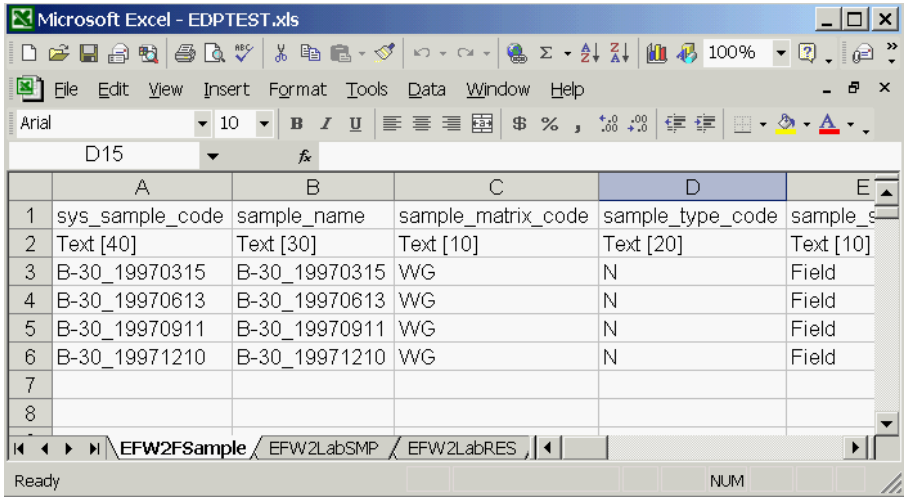
Using the “Load and Append” utility appends the new data file to the end of the original data file.

2.5.3 Clear Tables

If you accidentally load a data file into the wrong section, you can clear all data rows by right-clicking on the section name and selecting "Clear Table". Alternately, if you would like to clear all of the data loaded into all of the sections of the format, you may select Clear> Clear EDD from the datagroup.

2.5.4 Loading Excel Data Files

When using Excel files with EDP, the worksheet must be named with the EDD section’s name. For example, the Excel workbook shown below contains data in the 4-file format. The worksheet shown contains data in the EFW2FSample format, so the worksheet is named accordingly. Note the additional worksheets are also named for their respective EDD section (EFW2LabSMP, EFW2LabRES, etc.).



	A	B	C	D	E
1	sys_sample_code	sample_name	sample_matrix_code	sample_type_code	sample_s
2	Text [40]	Text [30]	Text [10]	Text [20]	Text [10]
3	B-30_19970315	B-30_19970315	WG	N	Field
4	B-30_19970613	B-30_19970613	WG	N	Field
5	B-30_19970911	B-30_19970911	WG	N	Field
6	B-30_19971210	B-30_19971210	WG	N	Field
7					
8					

Figure 7: Excel Workbook with EFWEDD Sections named properly for opening all sections directly into EDP.

The Excel file (.xls) may use any nomenclature. EarthSoft recommends including the format name at the end of the file name. However, to successfully load all of the

sections of the EDD at once the worksheets must be named according to the EDD section name. To do this:

1. Select EDD from the Open group on the Home tab.
2. Browse to the C:\Program Files\EarthSoft\EQuIS\Tutor\EDD Tutorial\EFWEDD_WithErrors folder and select the EFWEDD.SDG.xls file.
3. Click Open.

All of the sections of the EDD are loaded into the appropriate sections of the format.

2.5.5 Loading Access Database Files

Access database files can be opened with EDP in a similar method to the Excel files. When using Access database files with EDP, the Access tables must be named for the EDD section name the tables represent. For example, the Access database shown below contains tables which use the 4-file format's section names as the table names.

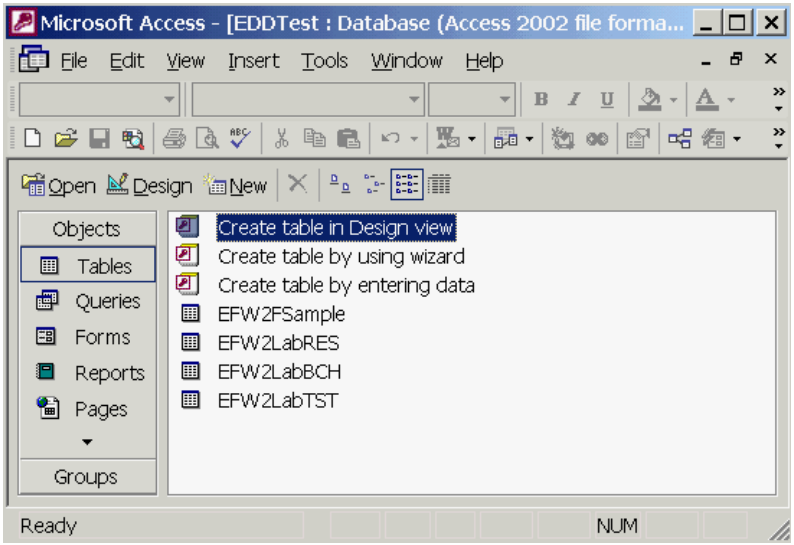


Figure 8: Microsoft Access database with EFWEDD format's section names as table names.

The Access database file (.mdb) may use any nomenclature. To load all of the sections of the EDD at once the data tables must be named according to the EDD format's section names.

2.5.6 Loading Zipped Data Files

Zipped data files can also be used with EDP. Any of the standard types of files (e.g., .txt, .csv, .xls, or .mdb) can be zipped and opened at the same time with EDP. When using Zipped data files with EDP, the data files (in the *.zip file) must be named with the EDD format's section name. For example, the zipped file shown below contains data files which use the 4-file format name as file name.

Additional information may be included in the data file's name, such as the sample delivery group number, as long as that information is separated from the section name by a period (.). For example 'EFW2FSample.SDG.txt' is an acceptable name for loading data.

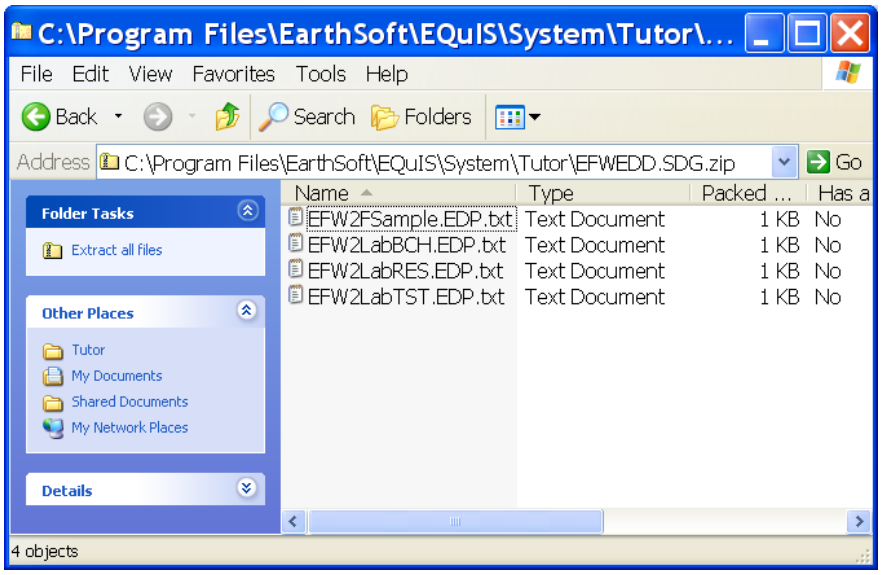


Figure 9: Zipped file's with EFWEDD format's section names as individual flat file names.

Regardless of whether the user is opening Excel, Access or Zipped files, notice that the file name (*.xls, *.mdb, *.zip) may use any nomenclature, but the sheet name (for Excel), the table name (for Access) and the individual file name (for zipped files) must use the EDD format's section names.

2.6 Finding and Resolving Errors

Once a data file has been loaded into EDP, any cell in the data grid that does not satisfy the checking requirements identified in the Format File are highlighted in colors that indicate the type of error. Users may obtain additional details about what each error is by holding their curser over the highlighted cell in the data grid. For additional information on changing these color codes see the Options section of this manual.

In addition to denoting the cell and/or row in which the error(s) occur in the grid, the line number along the left side of the data grid may also signify any line in which an error has occurred by appearing in bold red text.

Also, you can identify which sections of the formats contain at least one error because they will be highlighted in red. Notice that all data files that were checked contain errors as the format names are highlighted in red.

2.7 Error Log Reports

You can view a summary report of all of the errors in the data files by reviewing the Error Log or Error Summary reports.

The Error Log Report is a report that can be created to show what line in each section of the format contains exactly what error. The Error Log Report maintains EDP's error color coding to make error resolution easier.

The Error Summary Report may also be created to show a summary of all of the errors in the data files. Each row of the Error Summary Report contains the number of rows where the specified error occurred in each section of the format. The Summary Report does not maintain the color coding.

The Error Log and Error Summary Reports are saved as an HTML files, and open automatically in your default browser. Note also that the EDD Format File name and Version information are also reported in both Error Logs. In addition, information related to the EDD Data Files, User Name, and Reference Values File that were used during the error checking process are also listed in both of these logs to facilitate error resolution.

Users may find it useful that Error Logs and Error Summary Reports may also be opened in Microsoft Excel to allow for additional sorting. To do this, browse to the saved Error log in Windows Explorer, right click on the file name and select 'Open With' then select Microsoft Excel.

To view the Error Log, follow these steps:

1. Select the Error Log option from the Error Log group
2. Enter a file name (or use the default file name) and click Save.

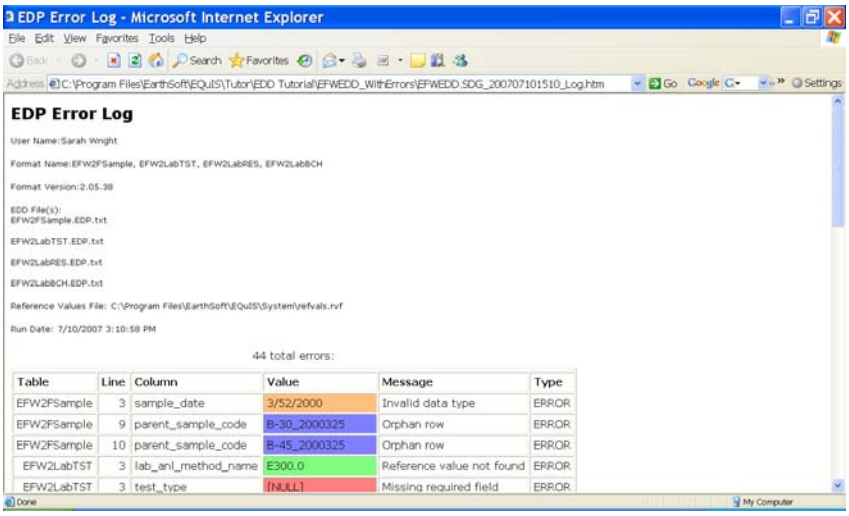


Figure 10: EDP Error Log Report

To view the Error Summary, follow these steps:

1. Select Summary from the Error Log group
2. Enter a file name (or use the default file name) and click Save.

2.8 Error Resolution

As you now know, errors that are found in EDP are highlighted with a specified color to signify what there error is. These data errors can be corrected within EDP if the Auditing feature was enabled during installation of EDP. These modifications may be made to any cell within the data file after it is loaded by simply highlighting the cell and typing the modification.

Before beginning it is important to understand that edits made to data files within EDP are **not** automatically saved. If users are modifying data files using the edit feature available in EDP then users should save their data files frequently during their EDP Session. Modifications made to data in EDP may be saved by selecting Save > Data file from the Application menu.

2.8.1 Column Headers Errors

You may have noticed that many of the errors in the data files loaded in these exercise are attributed to the two header rows that are contained in each of the data files. These two rows contain column header information that is useful to ensure that the correct data populates each column and to check that the appropriate data file was loaded into the correct section of the format.

You can instruct EDP to ignore any header rows by selecting the header rows in the data grid, right-click while the rows are highlighted and select "Set as Comment Row". You will need to complete this step for each section of the format. To do this:

1. Select the EFW2FSample format.
2. Highlight the two header rows, right-click and select "Set as Comment Row".

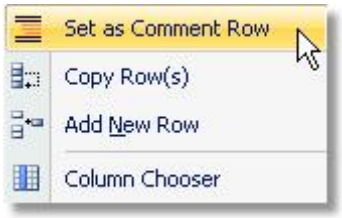
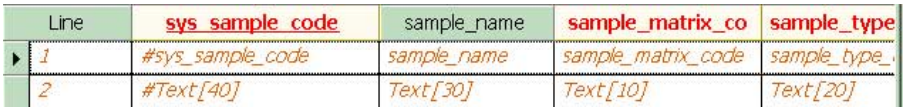


Figure 11: Set as Comment Row

The text in the first two rows becomes italicized and includes a pound sign (#) directly before the first letter of the column header. The colors of the cells in these rows also change indicating that they are no longer problematic. Repeat this step for the remaining loaded files.

The step of setting a row as a column header may be skipped if the data file's comment rows include this pound sign (#) prior to being loaded into EDP.



Line	sys_sample_code	sample_name	sample_matrix_co	sample_type
1	#sys_sample_code	sample_name	sample_matrix_code	sample_type_
2	#Text[40]	Text[30]	Text[10]	Text[20]

Figure 12: Example of Pound Sign (#) Comment Indicator

After you identify "Comment Rows" in the remainder of the data files, you can choose to not view the Comment Rows. On the View group, there is an option for Comment Rows. Note the check next to the word "Comment Rows". If you click on the "Comment Row" option, the check is removed and the Comment Rows will not be displayed. You can always choose to view the Comment Rows by making the same selection again.

After identifying the "Comment Rows", all errors should be removed from the EFW2LabTST and EFW2LabBCH files (the section names are highlighted in green). Use the Refresh All from the Data group option to refresh the sections if the changes are not reflected.

2.9 Data Errors

Let's review the other formats to identify the remaining problems in the data files.

1. Select the EFW2FSample section. Select Errors Only from the View group so that only the problematic data are showing.
2. Row 3 contains an error (the number "3" is highlighted in red). Use the scroll bar to scroll to the right to look for the error. The highlighted cell shows a sample_date = 3/52/2000 (whereas the remainder of the values show the sample_date = 3/25/2000). Click in the highlighted cell and change the sample_date to 3/25/2000. After you make this change, click on another cell. The change will be accepted and row 3 will no longer appear on the screen since there are no other errors in that row. To view row number 3, turn off the Errors Only option.
3. Rows 9 and 10 also contain errors. The parent_sample_code field is highlighted as problematic. Place your cursor over the cell to identify the error. The error "Orphan Row" indicates that the parent_sample_code has not been defined. This error indicates that the parent_sample_code is not related to a sys_sample_code of an 'N' or Normal Environmental Sample from the sys_sample_code field. By showing all records in the EDD (Select Errors Only from the view group), a close inspection of this entry compared to the entries in the sys_sample_code field reveals that the parent_sample_code has been incorrectly entered as "B-45_2000325" The correct sample code is "B-45_20000325". Change the parent_sample_code to "B-45_20000325".

   Additional information may be included in the worksheet's name, such as the sample delivery group number, as long as that information is separated from

the section name by a period (.). For example ‘EFW2FSample.SDG.txt’ is an acceptable name for loading data.

With these changes the EFW2FSample data file is corrected and the EFW2FSample section name is highlighted in green.

NOTE: If all errors have been corrected and the ‘Errors Only’ feature has been selected, no data will appear. You can review all the data by unselecting the “Errors Only” feature.

4. Save changes to the EFW2FSample.SDG.txt data file using the Save > EDD option on the Application menu.
5. Select the EFW2LabRES section of the format.
6. To quickly identify the problem records, make sure that Errors Only option is selected. Note that Row 9 and 18 are highlighted. Scroll to the right to the review highlighted cells. If you place your cursor in the highlighted cell, a pop-up window indicates the cause of the error = "Valid Value Not Found". Note in the chemical_name field that this is a result for Sodium. You may need to check the reference values.

   TIP: If you do not know which reference value table to use, hold your mouse over the column header in the format and the tool tip will tell you which reference values table to refer to for additional troubleshooting.

7. Click on the Reference Values tab. Select the rt_analyte table. Click on the chemical_name header to sort the data in the table by chemical_name. Scroll down to locate the entry for Sodium. Note that the CAS number for Sodium = 7440-23-5.
8. Click back to the EQuIS 4-file format tab and select the EFW2LabRES format. Locate the two problem CAS numbers and replace the incorrect values with the correct values from the preceding step. Use the Refresh Table option from the Data group to ensure that the changes are accepted.
9. Save changes to the data file using the Save > EDD option in the Application menu.

2.10 Sign and Submit

After using the tools outlined above to resolve all of the issues in a set of Data Files the data is ready to be submitted for loading into the EQuIS 5 database. The Sign and Submit tool was designed to facilitate submittal of data to EQuIS Enterprise EDP. Sign and Submit option packages the data files with the correct naming convention which allows easy submittal of data packages. Use of the Sign and Submit feature requires a user name and password that may be obtained from the EQuIS Database Administrator.

To use the Sign and Submit feature, after data files have been loaded and all of the errors have been resolved,

1. Select Sign and Submit from the Application Menu. This will open the Sign and Submit window.

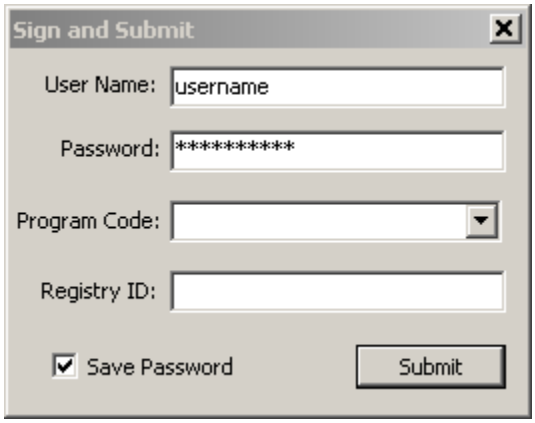


Figure 13: Sign and Submit Window

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. For this exercise use your name as the User Name and Password, and 'NY0123456' for the Program Code. This information will be used to create a user certificate file that EQuIS Enterprise uses to ensure a valid user is submitting data for the appropriate Program.

   NOTE: The Program Code and Registry ID are extremely important for Enterprise Data Submittals; however, the data entered into these sections of the Sign and Submit screen may be modified by the user as needed.

3. Click the Submit button.
4. Users will be prompted to provide a filename and location where you would like to save the file.

   The Sign and Submit feature will save an archived ("zipped") file named with the current date, a period, the Program Code, a period, the Registry ID, a period and the Format File name used to create the EDDs. (Example file name: '20080424. NY0123456.RegID123.EPARegion2EDD.zip'). The contents of the Zipped file includes text files named for the sections of the format used to create them.

5. Select Save.

   Once the zipped EDD Package has been saved the following screen will appear.

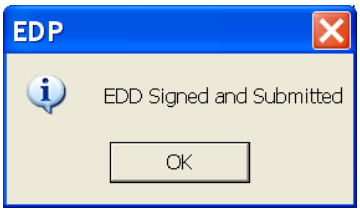


Figure 14: Sign and Submit Verification Window

6. Select OK

After the zipped file has been created the EDD Package is ready to be be submitted to your regulator for loading into EQuIS Professional EDP or EQuIS Enterprise EDP.

2.11 EDP Tools

There are several tools available in EDP to assist users in resolving errors and reviewing data.

2.11.1 Sort

Use the Sorting button  to sort the data in ascending or descending order. For example, if you want to sort the result data based on chemical name, you can highlight the chemical name column and click on the sorting button.

2.11.2 Filter

Use the Filter Column(s) button  to narrow the data set for easier viewing. By clicking on the filter icon, each column is enabled with a pull-down menu which allows you to filter on the available values for that field.

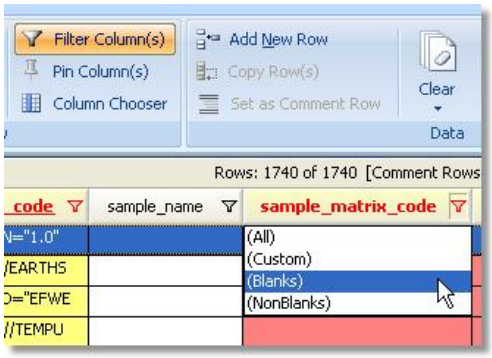


Figure 15: Filtering columns.

For example, if you are interested in reviewing the results for sample B-31_20000325, click on the Filter icon and using the pull-down menu enabled for the sys_sample_code field, select the value B-31_20000325. Now, you can review all of the results gathered for the sample in question.

2.11.3 Pin

Use the Pin button to assist with viewing large data sets. By clicking on the Pin icon, a “pin” is enabled adjacent to each field. The Pin icon allows certain fields to remain stationary while viewing the associated related fields for each record (this is analogous to freeze panes in Excel). For example, if you want to look at the test data for the matrix spike duplicate sample (B-45_20000325_MSD), click on the Pin icon. You will notice that a small “pin” appears adjacent to each field name. To more easily review the test data for B-45_20000325_MSD, click on the pin in the sys_sample_code field. As you scroll to the right, the sys_sample_code field remains stationary so you can review the data associated with each sample.

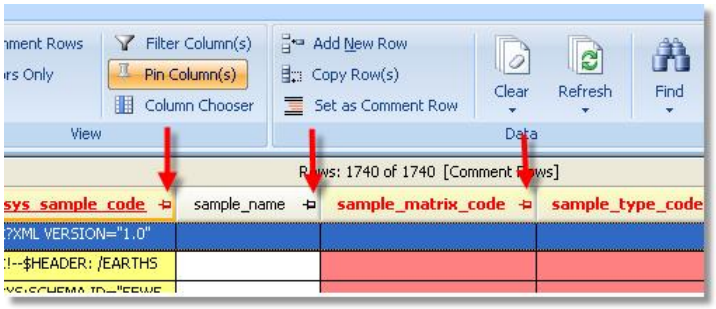


Figure 16: Pinning Columns.

2.11.4 Find and Replace

Use the Find button to find or replace multiple fields with new data. To use this, select any field within the column you would like to search. Then select the Find button, and the **Find and Replace** window opens. This will allow you to enter the text to be found (and the text to replace the original text with if desired).

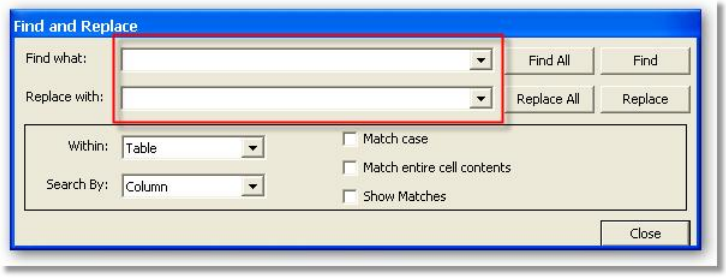


Figure 17: Find and Replace Window

TIP: Recent selections are available for reuse under the drop-down arrow(s).

The Find and Replace tool can be focused on a selection of the available rows in a displayed data file by selecting these rows prior to the opening the Find and Replace option and choosing “Selection” from the drop-down menu in the “Within:” option.

Additional choices are “Table” for the entire data file screen displayed, and “All**”** for all EDD tables open.

- **Search by:** provides options to search by rows or columns.

Other options include **Match Case** and **Match entire cell contents**, which should be used when an exact match is desired. These options also eliminate the inclusion of partial matches, which could create new errors. For example, when changing a Method from E300.0 to E310.1, selecting **Match entire cell contents** will prevent the modification of a Batch ID or Lab QC Sample ID that may also include the text "E300.1". With this setting, only the method cells containing E300.1 as the sole content will be selected.

- **Show matches** will open a detail box that will list the matches found
- **Replace** will replace the items one at a time.
- **Replace All** will replace all items found.

2.12 Export Blank EDD Template (xls)

Many users have their data providers create EDDs using Excel Workbook templates. Built into EDP is an easy tool for creating these templates. To do this:

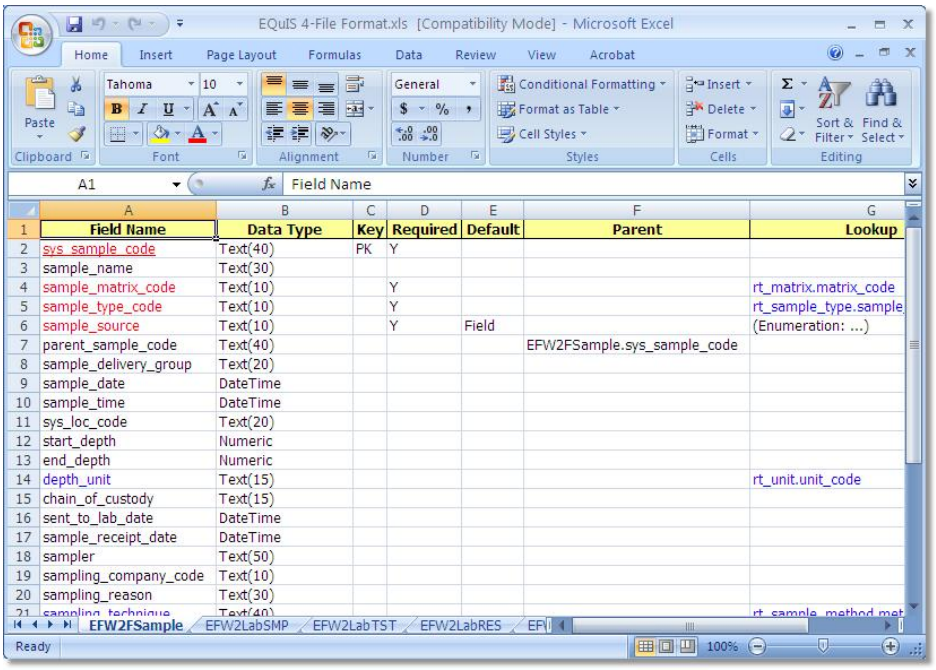
1. From the Tools group select Blank EDD
2. Excel will open and an EDD template will be created. The template includes the column header color coding and tool tips described in a previous section.

Users who create their own custom formats find this tool extremely useful. However, it should be noted that EarthSoft does not support custom formats, therefore custom formats will not be covered in this document.

2.13 Export EDD Description

The EDD Description option allows users to export detailed description of a format. This detail includes data types, key fields, required fields, field description and the database mapping.

To review the descriptions in Excel, select EDD Description from the Tools group. Excel will open and the descriptions will be displayed. If the format contains multiple sections such as the 4-file format, workbooks will be created for each section.



	A	B	C	D	E	F	G
1	Field Name	Data Type	Key	Required	Default	Parent	Lookup
2	sys_sample_code	Text(40)	PK	Y			
3	sample_name	Text(30)					
4	sample_matrix_code	Text(10)		Y			rt_matrix.matrix_code
5	sample_type_code	Text(10)		Y			rt_sample_type.sample
6	sample_source	Text(10)		Y	Field		(Enumeration: ...)
7	parent_sample_code	Text(40)				EFW2FSample.sys_sample_code	
8	sample_delivery_group	Text(20)					
9	sample_date	DateTime					
10	sample_time	DateTime					
11	sys_loc_code	Text(20)					
12	start_depth	Numeric					
13	end_depth	Numeric					
14	depth_unit	Text(15)					rt_unit.unit_code
15	chain_of_custody	Text(15)					
16	sent_to_lab_date	DateTime					
17	sample_receipt_date	DateTime					
18	sampler	Text(50)					
19	sampling_company_code	Text(10)					
20	sampling_reason	Text(30)					
21	sampling_technique	Text(40)					rt_sample_method.met

Figure 18: EDD Description in Excel.

2.14 Generate Format Files

The Generate Format Files option allows user to select an existing Access database that has been set up as a format file. This option will generate an *.xsd. As described previously in this document the *.xsd contains the field specifications and definitions. The *.xsd will also use the primary key fields defined in the Access database to create the primary key fields for the format. It does not contain the database mapping. This option will not create the reference values file (*.rvf) and enumeration file (*.xsd) or the conditional checks file (*.vb). This option can be used to create a skeleton of a format. This is a good starting point for generating user specific formats.

To use this option a user would create an Access database with at least one table which contains the fields, field lengths, types, descriptions as well as a primary key. If more than one table is included in the database this option will create multiple sections for the format. Once the database has been created the user would open EDP and select Generate Format File from the Tools group. First the user will be prompted to Select and Access Database, once selected click Open. The next screen will prompted the user for a new format file name. Once the new format file name has been entered the user selects the Save button and the format will be created. To view the newly created format, select the Format option from the Open group.

Please note that the format will not import data into the EQuIS database until the database mapping has been added to the newly created format.

2.15 Auditing

Auditing is a feature that enables users to save a log of what data is modify in the EDP data grid. This option may be enabled during EDP installation and allows the user to select the location for the Audit logs to be saved. The default location for the 'Log Editing Changes to Directory' is to the 'C:\Program Files\EarthSoft\EQuIS\Audit' directory.

The audit logs will be stored in subfolders by computer and user name, with file names containing the date and time the xml log file was saved. The audit logs contain information about changes made to data files in the EDP data grid.

The Audit option was added to accommodate laboratories who need to log any changes made to an EDD it does not track the EDD only edits made to an EDD. This provides labs with a audit log that tracks any edits made to a dataset so that if laboratory has to regenerate an EDD from their LIMS they will know what edits were made to that data before it was sent to their client.

NOTE: The audit folder is where the audit logs are stored - anyone who has permission to edit data needs WRITE access (but not READ access) to that folder. This is so EDP can write the audit logs to that folder. However, not everyone that can perform edits has the permission to see/review everyone else's edits (which are also stored in the same folder). Therefore, only "QA Personnel" need READ (and optionally WRITE) access to that folder, so they can see all of the audit logs from all of the users making edits.

2.16 Options

To assist the user in resolving issues effectively, EDP was designed to allow the user to modify the visual and functional settings in both Standalone EDP and Professional EDP.

2.16.1 General

The default functionality of EDP may be modified by the user so that EDP; will not recall the last Format File used, use a different comment indicator code, and increase the number of errors that EDP will check for when Data Files are initially loaded into EDP.

If you will not use the same Format File repeatedly, there is a preference you can set in the "Options" menu so the last Format File used is not always opened when EDP starts.

To set this Option:

1. Select the Application menu then the Options button on the bottom right, beside the Exit button.

2. On the Options dialog click on "General".

The option RecallLastUsedFormatFile will be displayed under the Application section. The default setting is ‘Yes’ which indicates that the last format file that was used in EDP will be opened during the following EDP session. This setting may be changed to ‘No’ to have EDP open to a blank screen each time EDP is launched.

If RecallLastUsedFormatFile is set to "Yes" and the specified Format File is not available or cannot be found, a warning message is given and you have the opportunity to select a different Format File

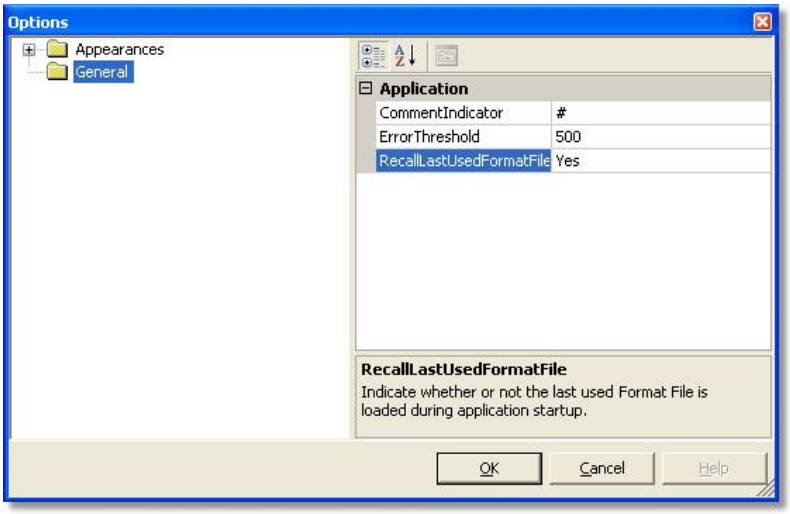


Figure 19: Option to recall last format opened in EDP.

Another modification users may want to make is to the comment row indicator. The default comment row indicator is a pound sign (#). Many users have found that if data is being exported from a Microsoft Access database the pound sign is not a good symbol to use. Therefore, the comment indicator may be changed. To do this:

1. Select the Application menu then the Options button
2. Click the General folder
3. Enter the desired comment indicator such as a dollar sign ($).
4. Select OK to save this setting.

2.16.2 Appearance

The default appearance of Column Headers, Errors, and Informational Messages may be modified by users to enhance the user experience. This functionality is available through the Application menu > Options button. To modify any of these items’ appearance:

1. Select the Application menu then the Options button
2. Expand the node (+) to the left of Appearances

3. Select an attribute you would like to modify. Such as the ‘Reference Value Cell Errors’ under Errors > Cell Errors in the Appearance tree located in the left of the Options screen.
4. Click in the cell that appears in the right hand portion of the Options window next to ‘BackColor’.
5. Click on the Down Arrow to display the colors selection options for cell background colors.
6. Select a color
7. Select OK

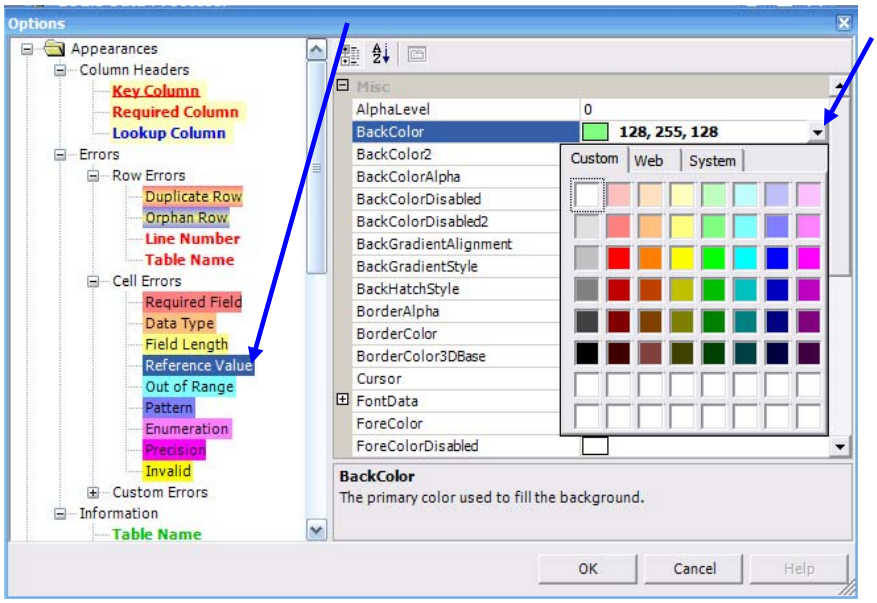


Figure 20: Example of Modifying Appearance in the Options Window.

This process may be repeated for any Appearance setting that are applied to cells in EDP.

3.0 Appendix A – Standard Format Files

There are several formats that are installed by default with the any type of EDP. By default, these standard formats are located in the /Program Files/EarthSoft/EQuIS/System directory and are free of charge.

Format Name	Format Files	Section Names
Action Levels	ActionLevels.xsd	ActionLevel ActionLevelParameter
EFWEDD (4-File Format)	EFWEDD.xsd EFWEDD.vb	EFW2FSample EFW2LabSMP EFW2LabTST EFW2LabRES EFW2LabBCH
EQEDD	EQEDD.xsd EQEDD.vb	Subfacility_v1 Task_v1 Location_v1 LocationParameter_v1 DrillActivity_v1 Lithology_v1 Well_v1 WellConstruction_v1 GeotechnicalSample_v1 WaterLevel_v1 WaterTable_v1 DownholePoint_v1 ExtrationInjectionWells_v1 Purge_v1 BasicResults_v1 FieldSample_v1 SampleParameter_v1 LabSample_v1 TestResults_v1 TestResults_QC_v1 TestBatch_v1 HistoricalLocation_v1 HistoricalWaterLevel_v1 HistoricalChemistry_v1 HistoricalGeology_v1
ESBasic	ESBasic.xsd ESBasic.vb	ESBasic
EZEDD	EZEDD.xsd EZEDD.vb	EZEDD
EZLithology	EZLithology.xsd	EZLithology

Geology EDD	GeologyEDD.xsd GeologyEDD.vb	Site Location Well SiteLoc LocationParameter AlternatePosition DownholePoint DrillActivity Drill Parameter GeologySample GeoSampleParameter Atterberg StaticProps Lithology WaterLevel WaterTable WellConstruction WellDatum
Reference Values	refvals.xsd	All Reference Tables in EQuIS 5 Database
Water Level	waterlevel.xsd	WaterLevel

4.0 Appendix B – Additional Format Files

Additional formats may be installed with any type of EDP. These formats are listed below and by default are installed in the C:/Program Files/EarthSoft/EQuIS/Formats directory. These formats may be licensed by contacting the organization that designed the formats, or by contacting EarthSoft.

Format Name	Format Files	Section Names
Audomated Data Review (ADR)	ADR.xse ADR.vbe	eQAPP A3 A1
Delaware Natural Resources and Environmental Control (DNREC)	DNREC.xse DNREC.vbe DNREC-enum.xsd	Facililty_v1 Task_v1 Subfacility_v1 Location_v1 LocationParameter_v1 DrillActivity_v1 Lithology_v1 Well_v1 WellConstruction_v1 GeoSample_v1 WaterLevel_v1 WaterTable_v1 DownholePoint_v1 ExtrationInjectionWells_v1 Purge_v1 FieldSample_v1 SampleParameter_v1 LabSample_v1 TestResult_v1 TestResultQC_v1 TestBatch_v1 HistoricalLocation_v1 HistoricalWaterLevel_v1 HistoricalChemistry_v1 HistoricalGeology_v1

Format Name	Format Files	Section Names
EPAR2	EPAR2.xse EPAR2.vbe EPAR2.rvf EPAR2-enum.xsd	EPAR2DPR_v1 EPAR2SITE_v1 EPAR2LOC_v1 EPAR2DRA_v1 EPAR2LTH_v1 EPAR2WEL_v1 EPAR2WSG_v1 EPAR2GSMP_v1 EPAR2TBL_v1 EPAR2DHP_v1 EPAR2EIW_v1 EPAR2SGS_v1 EPAR2FRES_v1 EPAR2SMP_v1 EPAR2TRS_v1 EPAR2TRSQC_v1 EPAR2BAT_v1 EPAR2H_BasicLOC_v1 EPAR2_BaiscWTR_v1 EPAR2_BasicCHEM_v1 EPAR2_BasicGEO_v1
EPAR4	EPAR4.xse EPAR4.vbe EPAR4-enum.xsd	EPAR4_Site_v1 EPAR4_Location_v1 EPAR4_Well_v1 EPAR4_SiteLoc_v1 EPAR4_LocationParemeter_v1 EPAR4_AlternatePosition_v1 EPAR4_DownholePoint_v1 EPAR4_DrillActivity_v1 EPAR4_GeologicSample_v1 EPAR4_GeoSampleParameter_v1 EPAR4_Atterberg_v1 EPAR4_StaticProps_v1 EPAR4_Lithology_v1 EPAR4_WaterLevel_v1 EPAR4_WaterTable_v1 EPAR4_WellConstruction_v1 EPAR4_WellDatum_v1 EPAR4_FieldResults_v1 EPAR4_FSample_v1 EPAR4_TST_v1 EPAR4_RES_v1

Format Name	Format Files	Section Names
EPAR5	EPAR5.xse EPAR5.vbe EPAR5-enum.xsd	EPAR5SITE_v2 EPAR5LOC_v2 EPAR5DRA_v2 EPAR5LTH_v2 EPAR5WEL_v2 EPAR5WSG_v2 EPAR5GSMP_v2 EPAR5TBL_v2 EPAR5DHP_v2 EPAR5EIW_v2 EPAR5SoilGas_v2 EPAR5SMP_v2 EPAR5TRS_v2 EPAR5TRSQC_v2 EPAR5BAT_v2 EPAR5HistLOC_v2 EPAR5HistWTR_v2 EPAR5HistCHEM_v2 EPAR5HistGEO_v2
Geotracker	GeotrackerEDF.xse GeotrackerEDF.vbe	EDFSAMP EDFTEST EDFRES EDFQC EDFCL GEO_XY GEO_Z GEO_WELL WELL_CONSTRUCTION EDFFlat EDFCLFlat
gINT	gINT.xse gINT_vbe	PROJECT POINT CPT LITHOLOGY REMARKS SAMPLE TESTS WELL

Format Name	Format Files	Section Names
gINT_Lab	gINT_Lab.xse gINT_Lab.vb	PROJECT POINT CPT LITHOLOGY REMARKS SAMPLE ATTERBERG ATTB READINGS COMPACTION COMP READINGS CONSOL READINGS CONSOLIDATION DIRECT SHEAR DSHR READINGS FALL HEAD K FHK READINGS FINE SG FINE SG READINGS HYD READINGS HYDROMETER LAB SPECIMEN SIEVE SV READINGS UNC READINGS UNCONF COMPR WC DENSITY WELL
Multiple Facility EDD	Facility.vb Facility.xsd	dt_facility dt_coordinate dt_location rt_group rt_group_member
RockWorks	RockWorks.xse RockWorks.vbe	
RUSS	RUSS.xse RUSS.vbe	

Format Name	Format Files	Section Names
West Virginia Department of Enviromental Protection (WVDEP)	WVDEP.vbe WVDEQ.xse	WV_DATAPROVIDER_V2 VW_SITE_V2 WV_SUBSITE_V1 WV_LOCATION_V2 WV_WELL_V1 WV_DRILLACTIVITY_V1 WV_LITHOLOGY_V1 WV_WELLCONSTRUCTION_V1 WV_GEOLOGICSAMPLE_V1 WV_WATERTABLE_V1 WV_DOWNHOLEPOINT_V1 WV_FIELDRESULTS_V2 WV_WATERLEVEL_V2 WV_SAMPLECOLLPROC_V2 WV_SAMPLEINFO_V2 WV_TESTRESULTS_V2

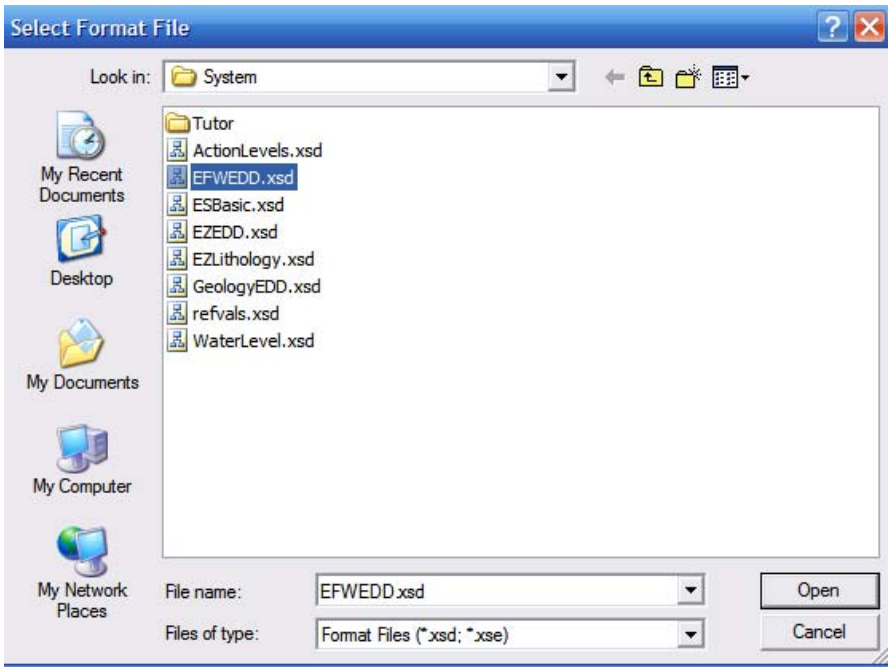
5.0 Appendix C – Exercises

5.1 Exercise 1 – Changing the appearance of errors

Question: How do you change the appearance of errors in EDP Standalone or Professional EDP?

Answer:

1. Launch EDP Standalone from Start > Program Files > EarthSoft > EQuIS Data Processor.
2. Open the 4 file format (EFWEDD) File > Open > Format File
3. Browse to C:\Program Files\EarthSoft\EQuIS\System
4. Double Click on EFWEDD.xsd:



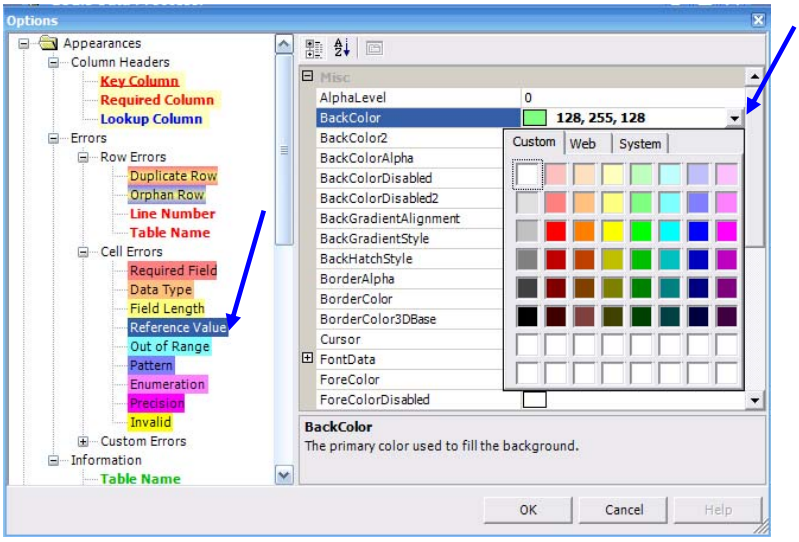
5. Open EDD Data Files by clicking the EDD button  .

   Errors in the Data Files are indicated by color coded cells. When the mouse is placed over an error, an error description will appear.

The appearance of the errors within EDP may be customized by doing the following:

1. Click the Options button in the EDP Application Menu:

2. Select an attribute such as the ‘Reference Value Cell Errors’ under ‘Cell Errors’ in the Appearance item tree on the left
3. Select a BackColor in the Format Section on the right.
4. Click on the Down Arrow to display the default colors selections.



5. Select OK to return to EDP.

5.2 Exercise 2 – Reporting missing reference values

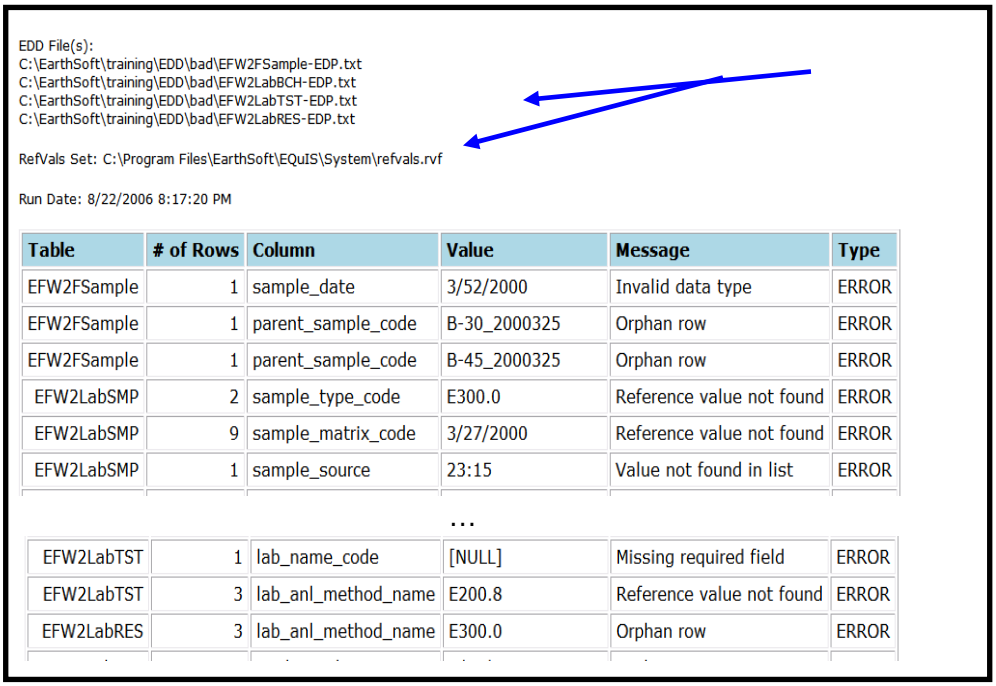
Question: In Standalone EDP, how do you provide a list of reference values that are missing to database managers?

Answer:

Provide the database manager with an Error Summary. The Error Summary will provide a list of missing reference value.

To create an Error Summary:

1. Click the Error Summary button  .
2. Note the location of both the EDD and the reference values files are provided.



EDD File(s):
C:\EarthSoft\training\EDD\bad\EFW2FSample-EDP.txt
C:\EarthSoft\training\EDD\bad\EFW2LabBCH-EDP.txt
C:\EarthSoft\training\EDD\bad\EFW2LabTST-EDP.txt
C:\EarthSoft\training\EDD\bad\EFW2LabRES-EDP.txt

RefVals Set: C:\Program Files\EarthSoft\EQuIS\System\refvals.rvf

Run Date: 8/22/2006 8:17:20 PM

Table	# of Rows	Column	Value	Message	Type
EFW2FSample	1	sample_date	3/52/2000	Invalid data type	ERROR
EFW2FSample	1	parent_sample_code	B-30_2000325	Orphan row	ERROR
EFW2FSample	1	parent_sample_code	B-45_2000325	Orphan row	ERROR
EFW2LabSMP	2	sample_type_code	E300.0	Reference value not found	ERROR
EFW2LabSMP	9	sample_matrix_code	3/27/2000	Reference value not found	ERROR
EFW2LabSMP	1	sample_source	23:15	Value not found in list	ERROR
…					
EFW2LabTST	1	lab_name_code	[NULL]	Missing required field	ERROR
EFW2LabTST	3	lab_anl_method_name	E200.8	Reference value not found	ERROR
EFW2LabRES	3	lab_anl_method_name	E300.0	Orphan row	ERROR

The Error Summary Report shows that reference values such as ‘E200.8’ will need to be added or addressed by the database administrator.

