Windsor Solutions, Inc.

State and Local Emissions Inventory System (SLEIS)

File Import Specification

Version 1.2-NDEQ

August 2018

WINDSOR
SOLUTIONS, INC.
Environmental + Health
Information Systems

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

Revision History

Date	Author	Changes	Version
10/01/2011	Windsor Solutions	Initial version	1.0
03/01/2014	Windsor Solutions	Added column to Processes.csv and ProcessEmissions.csv file definition of "Process Emissions Page Field (Tab: Field Label)" that specifies the equivalent field within the SLEIS application's Process Emission page. Also added some additional text in various Comments/Notes.	1.1
8/29/018	Windsor Solutions	Added NDEQ-specific business rules to data file specifications.	1.2-NDEQ

Introduction

The purpose of this document is to describe the State and Local Emissions Inventory System (SLEIS) file import process and flat file specification. It includes standards and conventions to use when constructing an import file, describes the physical structure of each import file type, and provides basic instruction on how to create a SLEIS import file.

The contents of this document cover the following file import related topics:

File Import Process	Defines the file import process, including an overview and specifics for each step in the process.
CSV File Format	Describes the general design conventions of the Comma Separated Values (CSV) file format.
Data File Specification	Details the flat file specification for each data file type used in the file import process.
Appendix	Specifies data file reference values, and offers brief instructions on editing CSV file formats.

File Import Process

Overview

File imports will be initiated and completed using the SLEIS user interface. This interface provides functionality for:

1) Selecting the emissions report that will receive the contents of the imported data files

2) Invoking the file import process for the selected emissions report

3) Selecting the specific data files to load for the selected emissions report

4) Executing the file import process, which is composed of the following three steps:

   a. Data file format validation

   b. Business rule validation

   c. Database load processing

5) Viewing the results of the file import process, such as data file format errors and business rule violations

Data File Format Validation

Each SLEIS data file must adhere to a pre-defined comma separated values (CSV) file format, with two data files comprising the full payload for an imported emissions report. Upon initiation of the file import process, the following data file format validation processing will be performed:

1) Validate that header columns in each CSV file are properly named, and that all appropriate columns exist in the file (refer to the *Data File Specification* section).

2) Validate that each CSV file line and field is properly formatted (refer to the *CSV File Format* section).

3) Validate that each CSV file line can be parsed into individual field values (refer to the *Data File Specification* section).

Any data file format validation errors will be displayed to the user within the user interface during the file import process.

Business Rule Validation

Following successful data file format validation, the business rule engine will process the individual records extracted from the data file. Relevant business rules will be checked by the system, and any errors discovered will be displayed to the user within the SLEIS user interface during the file import process.

Database Load Processing

Following successful business rule validation, the data extracted from the imported files will be saved to the SLEIS database. Reference the appropriate *Data File Specification* section of this document for additional details.

STATE AND LOCAL EMISSIONS INVENTORY SYSTEM (SLEIS) DESIGN | **FILE IMPORT SPECIFICATION**

Data insert and overwrite/replacement will occur at the emission unit level, and the file import process can be run multiple times for a report, inserting and replacing data as needed.

For example:

- File import 1 contains emissions data for emission unit A, B, and C

- File import 2 contains emissions data for emission unit C, D, and E

In this case, data for emission unit C would be overwritten by the second file import. Data for emission units A, C, D, and E would be inserted once, and not overwritten/replaced.

CSV File Format

A CSV data file format will be used to import data into SLEIS. This file format provides for the storage of structured data in a relatively simple format that can be easily created through automated (e.g., exported from a facility's information system) or manual (e.g., MS Excel, or even a basic text editor) means.

The CSV file format design conventions used by SLEIS data file include:

1) Each record will be contained on one line of the file, terminated by a line feed, or a carriage return and line feed pair, with each field separated by a comma.

2) The first record in the CSV file will contain a column name header in each of the fields.

3) There is no comma following the final field in the record.

4) In case field values have embedded commas, embedded double-quote characters, intentional leading/trailing space characters, or other reserved characters, fields will always be enclosed within double-quote characters, whether necessary or not.

5) Leading and trailing spaces or tabs adjacent to commas (not within double-quotes) will be trimmed.

6) Any embedded double-quote characters must be represented by a pair of double-quote characters.

   For example, a CSV file containing:

   ```
   "header 1"  , "header 2"  , "header 3" , "header 4" ,  "header 5"
   "value 1", "", "value 3, value 3, value 3" , "  value 4  " ,  "value ""value"" 5"
   ```

   Would be processed into individual rows and columns as:

header 1	header 2	header 3	header 4	header 5
value 1		value 3, value 3, value 3	value 4	value "value" 5

Data File Specification

Since the file import process itself will always be initiated by a user from within SLEIS, where a specific report has already been selected, the data files do not require meta-data for report identification (e.g., reporting year, facility identifier, etc.). This data will already be available at the time of import, and is not needed in the contents of the data file. Specific references to this data are made within the CSV Data File Specification where appropriate.

The emission report import payload will be composed of the following CSV data files, and will be processed in the order given:

1) Processes.CSV

2) ProcessEmissions.CSV

3) ProcessSupplementalParameters.CSV

Processes.CSV

General business rules:
- Will not allow import of process emissions with parent emissions units that have been "shutdown" in current reporting year or earlier
- Will not allow process emissions with unit process last/final inventory year less than current reporting year

(Note, all columns except ones highlighted must be completed)

Column Header	Allowed Values	Example	Business Rules/Notes	Process Emissions Page Field (Tab: Field Label)
EmissionUnitId	Text Value	EU1	Required. The unique emission unit identifier for the facility.	Process: Emission Unit Identifier
ProcessId	Text value	PR1	Required. The unique process identifier for the emission unit for the facility.	Process: Process Identifier
IsReported	Text value of: • TRUE • FALSE	TRUE	Defaults to TRUE if not specified. NOTE: If IsReported value is FALSE, no process data or process emissions data will be imported for this process (if any has been specified).	Process: Process is Reported?
IsConfidential	Text value of: • TRUE	FALSE	Defaults to FALSE if not specified. NOTE: If IsConfidential value is TRUE, throughput	Process: Process is CBI?

Column Header	Allowed Values	Example	Business Rules/Notes	Process Emissions Page Field (Tab: Field Label)
	• FALSE		quantity and process emission factors will be treated as confidential business information (CBI), and not reported to the EPA.	
ThroughputQuantity	Floating point number	7183.175	Annual Throughput/Process Rate **NDEQ Business Rules:** • Required • Precision and scale (P,S) is limited to (25,10)	Process: Annual Throughput
ThroughputUnit	Text value*		*Refer to the valid reference values provided by the agency for this attribute in either the ReferenceThroughputValues.CSV file, or by referring to the equivalent Process Emissions page field within the application. **NDEQ Business Rules:** • Required	Process: Throughput Unit of Measure
ThroughputType	Text value*		*Refer to the valid reference values provided by the agency for this attribute in either the ReferenceThroughputValues.CSV file, or by referring to the equivalent Process Emissions page field within the application. **NDEQ Business Rules:** • Required	Process: Throughput Type
ThroughputMaterial	Text value*		*Refer to the valid reference values provided by the agency for this attribute in either the ReferenceThroughputValues.CSV file, or by referring to the equivalent Process Emissions page field within the application. **NDEQ Business Rules:** • Required	Process: Throughput Material
Comments	Text value up to 4000 characters in length		Optional comments/notes regarding the process emission record.	Process: Comments
OpStartTime	Time, including AM or PM	8:00 AM	**NDEQ Business Rules:** • Optional	Operations: Standard Start Time
OpStopTime	Time, including AM or PM	5:00 PM	**NDEQ Business Rules:** • Optional	Operations: Standard End Time

Column Header	Allowed Values	Example	Business Rules/Notes	Process Emissions Page Field (Tab: Field Label)
AvgHrsPerDay	Floating point number, >=0 and <= 24	22.54	**NDEQ Business Rules:** • Required • Precision and scale (P,S) is limited to (4,2)	Operations: Average Hours/Day
AvgDaysPerWeek	Floating point number, >=0 and <=7	6.17	**NDEQ Business Rules:** • Required • Precision and scale (P,S) is limited to (3,2)	Operations: Average Days/Week
AvgWeeksPerYear	Floating point number, >=0 and <=52	50.53	**NDEQ Business Rules:** • Required • Precision and scale (P,S) is limited to (4,2)	Operations: Average Weeks/Year
ActualDaysPerPeriod	Floating point number, >=0 and <=366	250	**NDEQ Business Rules:** • Not used. Leave blank.	Operations: Actual Days/Year
ActualHrsOperation	Floating point number >=0 and <= 8784	7233	**NDEQ Business Rules:** • Required • Precision and scale (P,S) is limited to (5,1)	Operations: Actual Hours/Year
DecToFebPercent	Floating point number, >= 0 and <= 100	24.55	**NDEQ Business Rules:** • Required • Precision and scale (P,S) is limited to (4,1) • Total seasonal % must be between 99.5 and 100.5 or 0	Operations: Season Operations December-February (%)
MarToMayPercent	Floating point number, >= 0 and <= 100	24.55	**NDEQ Business Rules:** • Required • Precision and scale (P,S) is limited to (4,1) • Total seasonal % must be between 99.5 and 100.5 or 0	Operations: Season Operations March-May (%)
JunToAugPercent	Floating point number, >= 0 and <= 100	25	**NDEQ Business Rules:** • Required • Precision and scale (P,S) is limited to (4,1) • Total seasonal % must be between 99.5 and 100.5 or 0	Operations: Season Operations June-August (%)
SepToNovPercent	Floating point number, >= 0 and <= 100	25	**NDEQ Business Rules:** • Required • Precision and scale (P,S) is limited to (4,1) • Total seasonal % must be between 99.5 and 100.5 or 0	Operations: Season Operations September-November (%)

Column Header	Allowed Values	Example	Business Rules/Notes	Process Emissions Page Field (Tab: Field Label)
TotalOzoneSeasonDays	Integer, 0-153	153	Days operated May-September. **NDEQ Business Rules:** • Not used. Leave blank.	Operations: Total Ozone Season Days (May-September)
TotalSummerSeasonDays	Integer, 0-92	92	Days operated June-August. **NDEQ Business Rules:** • Not used. Leave blank.	Operations: Total Summer Season Days (June-August)
TotalCOSeasonDays	Integer, 0-91	91	Days operated December-February. **NDEQ Business Rules:** • Not used. Leave blank.	Operations: Total CO Season Days (December-February)

ProcessEmissions.CSV

General business rules:
- At least one pollutant must be specified for each process reported in the Processes.csv file
- Pollutants cannot be duplicated (reported twice) for the same emission unit and process

Column Header	Allowed Values	Example	Business Rules/Notes	Process Emissions Page Field (Tab: Field Label)
EmissionUnitId	Text value	EU1	Required. Unique emission unit identifier for the facility. Used as pseudo key along with ProcessId to the corresponding Processes.CSV record.	Process: Emission Unit Identifier
ProcessId	Text value	PR1	Required. Unique process identifier for the facility. Used as pseudo key along with EmissionUnitId to the corresponding Processes.CSV record.	Process: Process Identifier
PollutantCode	Text value*	NOX	Required. *Refer to the valid reference values provided by the agency for this attribute in either the ReferenceDataValues.CSV file (search for *PollutantCode* in the AttributeName column), or by referring to the equivalent Process Emissions page field within the application.	Emissions: Pollutant Code
CalculationMethod	Text value*		Required. *Refer to the valid reference values provided by the agency for this attribute in either the ReferenceDataValues.CSV file (search for *CalculationMethod* in the AttributeName column), or by referring to the equivalent Process Emissions page field within the application.	Emissions: Calculation Method
EmissionFactor	Floating point number	3.3579	In Lbs per EmissionFactorUnit **NDEQ Business Rules:** • Required if CalculationMethod code is NOT '1', '2', '3_0', '5', '6', '7'	Emissions: Emission Factor (Lbs/Unit)

Column Header	Allowed Values	Example	Business Rules/Notes	Process Emissions Page Field (Tab: Field Label)
			• Not allowed (leave blank) if CalculationMethod code is '1', '2', '3_0', '5', '6', '7' • Precision and scale (P,S) is limited to (28,15)	
EmissionFactorUnit	Text value*		*Refer to the valid reference values provided by the agency for this attribute in either the ReferenceDataValues.CSV file, or by referring to the equivalent Process Emissions page field within the application. NOTE: The EmissionFactorUnit value supplies the denominator of the implied "Lbs" per EmissionFactorUnit of the EmissionFactor value. **NDEQ Business Rules:** • Required if CalculationMethod code is NOT '1', '2', '3_0', '5', '6', '7' • Not allowed (leave blank) if CalculationMethod code is '1', '2', '3_0', '5', '6', '7'	Emissions: Emission Factor Unit
EmissionQty	Floating point number	2978.45678	Required. Estimated emissions in Tons. NOTE: If a CalculationMethod is specified that requires an EmissionFactor and EmissionFactorUnit, AND the EmissionFactorUnit matches the Processes.csv Throughput Unit, the EmissionQty will be calculated upon import, and the EmissionQty from the import file will be ignored. If CaclulationMethod is specified that DOES NOT require EmissionFactor, EmissionFactor and EmissionFactorUnit cannot be specified, and EmissionQty must be specified. **NDEQ Business Rules:** • Must be >= 0 • Precision and scale (P,S) is limited to (28,15)	Emissions: Estimated Emissions (Tons)
StackTestDate	Date format		**NDEQ Business Rules:** • Required if CalculationMethod code is '4' or '24'	Emissions: Stack Test Date
Comments	Text value up to 4000 characters in length		Optional comments/notes regarding the process emission record.	Emissions: Comments

ProcessSupplementalParameters.CSV

General business rules:
- This is an optional import file
- Values entered will be validated based on configurations stored in the database for each parameter (variable). For example, fields that are percentages must be a valid number greater than 0 and less than or equal to 100.

Column Header	Allowed Values	Example	Business Rules/Notes	Process Emissions Page Field (Tab: Field Label)
EmissionUnitId	Text value	EU1	Required	Process: Emission Unit Identifier
ProcessId	Text value	PR1	Required	Process: Process Identifier
ParameterCode	Text value	A	Leave as is. This field will be populated after performing a Download Template operation, but is for information purposes only.	Process: Supplemental Calculation Parameters
ParameterName	Text value	% Ash	Leave as is. This field will be populated after performing a Download Template operation, but is for information purposes only.	Process: Supplemental Calculation Parameters
Value	Floating point number	2.6	Required	Process: Supplemental Calculation Parameters

Appendix

Reference Data Values

Reference data used to validate values within the imported data files is agency specific, and is not documented within this specification. The *Download Template* function available on the report home page will generate the following CSV files:

1. Processes.CSV

   Data from the currently selected report, formatted per specification in the *Data File Specification* section for this CSV file type.

2. ProcessEmissions.CSV

   Data from the currently selected report, formatted per specification in the *Data File Specification* section for this CSV file type.

3. ReferenceDataValues.CSV (see below)

   A reference file that contains the list of allowed values that are used in order to pass validation rules per the specification in the *Data File Specification* for any CSV file type.

Column Header	Example	Comments/Notes
AttributeName	AnnThroughputUnit	Corresponding column header/attribute name as documented in the data file specification.
Value	GAL	Code/value to input into the import data file.
Description	GALLONS	Descriptive text of the code/value specified in the value column.
Notes		Additional descriptive text or instructions on the attribute value.

4. ReferenceThroughputValues.CSV (see below)

   A reference file used to select a valid throughput combination for the Processes.CSV file. It contains the list of valid throughput combinations based on Source Classification Code that are used in order to pass validation rules. To use, find the SCC number of the Process, select the correct ThroughputUnit, ThroughputType, and ThroughputMaterial (columns B, C, and D), then copy and paste them into Procedsses.CSV file (columns F, G, and H).

Column Header	Example	Comments/Notes
SCC	10100101	The Source Classification Code value. Search this column for the SCC for which allowed throughput combinations are needed.
ThroughputUnit	TON	Throughput unit of measure code to input into the import data file.
ThroughputType	I	Throughput type code to input into the import data file. I = Input, O = Output, E = Existing
ThroughputMaterial	640	Throughput material code to input into the import data file.
MaterialType	Anthracite	Descriptive text of the throughput material name. It will not be used in the import data file, but will be used as a reference when selecting material code (above).

Editing CSV Files

Ideally, generation of SLEIS-compatible CSV files will be handled programmatically through a facility's existing information system(s), or possibly Microsoft Excel macros/scripting.

For manual editing of the CSV files, the default editor for most users will likely be Microsoft Excel, which recognizes the CSV file format and automatically loads and formats the data. Any text editor may also be used (e.g., Notepad, etc.) to edit CSV files, and there are a variety of dedicated CSV file editors. For example:

- http://csved.sjfrancke.nl/

- http://www.ronsplace.me.uk/Professional/Products/TabularDataEditor/Default.aspx

Note that one significant limitation of editing CSV files with Microsoft Excel is that an undesired data conversion will occur when opening a CSV file with leading 0's (zeros) in alpha-numeric fields. For example, a process name, or code value with all numeric values will be converted to a numeric value with leading zeros removed upon opening the file. This can cause values such as '001' to be converted to the number 1 (no leading zeroes).

To edit a CSV file in MS Excel without having leading zeros removed from fields, perform the following steps:

1. Extract the individual .CSV files from the .ZIP file generated using the Export or Download Template function to your local system (e.g., C:\Users\[username]\Desktop)

2. Open the MS Excel application (Note: <u>NOT</u> by double-clicking the file, just open the application itself)

3. Click on the *Data* tab

4. Click *From Text* on the ribbon bar of the *Data* tab. A file selection dialog will be opened.

5. Navigate to where you extracted the files from the .ZIP file and select one of the file (e.g., Processes.csv), then click *Import*. The Text Import Wizard dialog is opened.

6. Ensure the *Delimited* file type option is selected, then click *Next*.

7. Ensure *Comma* is the only item selected in the *Delimiters* section and click the *Next* button.

8. In the *Data Preview* section, select field that could contain leading zeros (e.g., the EmissionUnitId and ProcessId columns) and select the *Text* option in the *Column data format* section. Repeat for all columns that could contains leading zeros you want to retain in the value.

9. Click *Finish*.

Note: The file can then be edited and saved as an MS Excel document, but you will need to save as .CSV file format before it can be imported back into SLEIS (through File Save As dialog in MS Excel, choose .csv file format).