CSV File Format Specification

Abstract:

This document describes CSV file format specification for information exchange from Energy Suppliers to Ofgem IT Systems.

Document Control Information:

Change Control Reference:	**V0.2**
Date:	**22/04/2015**
Issue:	**0.2**

Approvals:

Author:	Richa Mishra	
	Lead Architect	Date:24/12/2012
Approved by:	Kumar Ramakrishnan	
		Date: 27/12/2012
Authorised by:	David Davies	
		Date: 22/04/2015

Amendment Record

Issue No	Date	By Whom	Pages Altered/Comments
0.1	20/12/2012	Richa Mishra	Initial Draft
0.2	22/04/2015	Dai Davies	Removed header content. Document content unchanged.

Document Distribution

Distribute To	Role	Issue No

Referenced Documents

Ref	Document Reference	Document Name	Issue No

Table of Contents

Introduction

1.1　Background

The comma-separated values (CSV) format is a widely used text file format often used to exchange data between applications. It contains multiple records (one per line), and each field is delimited by a comma.

CSV Files

1.2　CSV File Format

The primary function of CSV file is to separate each field values by comma separated and transport text-based data to one or more target application. A source application is one which creates or appends to a CSV file and a target application is one which reads a CSV file

1.2.1　CSV File Structure

The CSV file structure use following two notations

- FS (Field Separator) i.e. comma separated
- FD (Field Delimiter) i.e. Always use a double-quote.

Each line feed in CSV file represents one record and each line is terminated by any valid NL (New line i.e. Carriage Return (CR) ASCII (13) and Line Feed (LF) ASCII (10)) feed.

Each record contains one or more fields and the fields are separated by the FS character (i.e. Comma)

A field is a string of text characters which will be delimited by the FD character (i.e. double-quote ("))

Any field may be quoted (with double quotes).

Fields containing a line-break, double-quote, and/or commas should be quoted. (If they are not, the file will likely be impossible to process correctly).

The FS character (i.e. comma) may appear in a FD delimited field and in this case it is not treated as the field separator.

If a field's value contains one or more commas, double-quotes, CR or LF characters, then it MUST be delimited by a pair of double-quotes (ASCII 0x22).

DO NOT apply double-quote protection where it is not required as applying double quotes on every field or on empty field would takes more file space

If a field requires Excel protection, its value MUST be prefixed with a single tilde character.
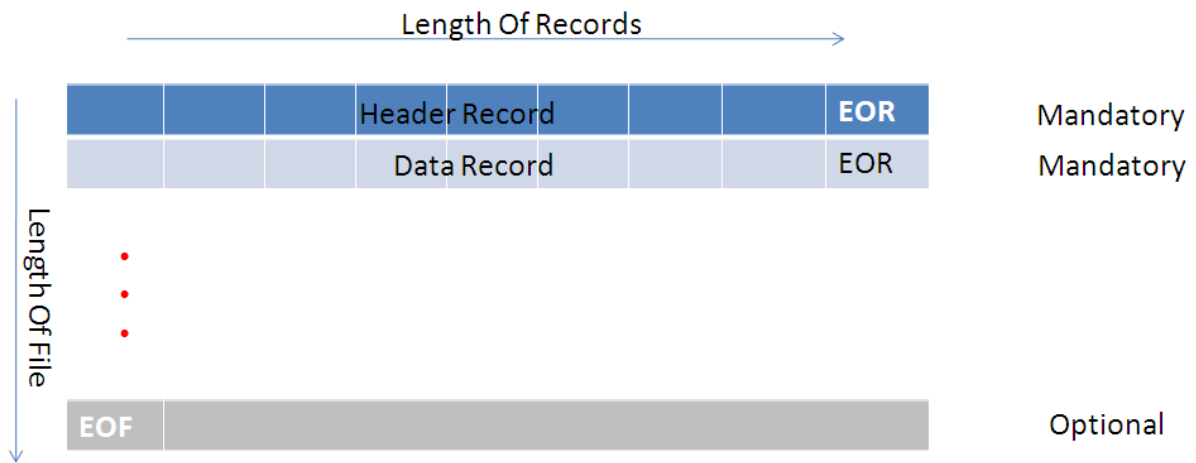
See example below:

FS =,
FD ="

Data Record: Test1,Test2,, "Test3,Test4" ,"Test5 ""Test6"" Test7", " Test8,""", ",Test9"

Indicates the following four fields
Test1   5 characters
Test2   5 characters
                           0 characters
Test3,Test4   11 characters
Test5 "Test6" Test7   20 characters
 Test8,"   8 characters
,Test9   6 characters

1.3   CSV File Rules

- The file type extension MUST be set to .CSV
- The character set used by data contained in the file MUST be an 8-bit (UTF-8).
- No binary data should be transported in CSV file
- A CSV file MUST contain at least one record.
- No limit to the number of data records
- The End of Record must be set to CR+LF (i.e. Carriage Return and Line Feed)
- Do not use whitespaces in the file name
- The EOR marker MUST NOT be taken as being part of the CSV record
- EOF (End of File) character indicates a logical EOF (SUB - ASCII 0x1A) and not the physical end.
- A logical EOF marker cannot be double-quote protected.
- Any record appears after the EOF will be ignored

1.4   File Size

Maximum csv file size should be 30 MB.

CSV Records

A CSV record consists of two elements, a data record followed by an end-of-record marker (EOR). The EOR is a data record delivery marker and does not form part of the data delivered by the record

1.5 CSV Record Rules

Pls. note this rule applies to every CSV record including the last record in the file.

1.6 CSV Field Column Rules

- Each record within the same CSV file MUST contain the same number of field columns. The header record describes how many fields the application should expect to process.
- Field columns MUST be separated from each other by a single separation character
- A field column MUST NOT have leading or trailing whitespace

1.7 Header Record Rules

A header record allows the Ofgem IT systems to guard against the potential issues such as missing column or additional column that are not in scope

- The header record MUST be the first record in the file.
- A CSV file MUST contain one header record only.
- Header labels MUST NOT be blank.
- Use single word only
- Do not use spaces (Use _ if words needs to be separated)

Acronyms and Abbreviation

ASCII	American Standard Code for Information Interchange
CR	Carriage Return
CSV	Comma Separated Value
EOF	End of File
EOR	End of Record
FD	Field Delimiter
FS	Field Separator
LF	Line Feed
NL	New Line
UTF	Unicode Transformation Format