SDMX csv FORMAT

SPECIFICATIONS

Version 1.0

17/01/2018

Contents

[Introduction 2](#_Toc503968834)

[RFC 4180: A common format for CSV files 2](#_Toc503968835)

[Design principles 2](#_Toc503968836)

[Localised names 4](#_Toc503968837)

[Optional parameters 4](#_Toc503968838)

[Examples 4](#_Toc503968839)

# Introduction

SDMX-CSV Data Message is an SDMX data exchange format based on the [RFC 4180](https://tools.ietf.org/html/rfc4180). CSV is a widely used standardised and simple format to exchange data supported by many tools.

SDMX-CSV integrates with other specifications, i.e.:

* The SDMX API RESTful specification (e.g. content negotiation with mime-type to get SDMX-CSV representations, specific formats for responses, language selection through HTTP content negotiation)
* The [RFC 4180](https://tools.ietf.org/html/rfc4180) specification (determined column number, "comma" separated)

SDMX-CSV is flexible enough in its representation to support the needs of different target audiences:

* A representation optimised for public data dissemination and similar, and for usage in common statistical software
* A representation optimised for creating pivot tables in spreadsheets applications

# RFC 4180: A common format for CSV files

In order to benefit from best practices, SDMX-CSV is based on the rules defined in the [RFC 4180](https://tools.ietf.org/html/rfc4180), which defines a common format and MIME Type for CSV files. It is advised to read the (very short) RFC for a full list of requirements but, in a nutshell, the RFC defines rules such as:

* How the CSV file should be structured (the RFC specifies that all records must have an identical structure, like when using an SDMX "flat" representation for data);
* When double-quotes should be used and how to escape them when needed;
* How spaces should be handled;
* Which separator should be used (comma or locale specific);
* Which mime type should be used;
* What is the default character set, etc.

However, in order to assure the possibility to always clearly identify the data contained in the message, the SDMX specification excludes switching off column headers.

# Design principles

* There is no additional SDMX-specific header. The SDMX-CSV format is designed for the purpose of general public dissemination of statistical data.
* After the mandatory header row, each row contains the information related to one specific observation.
* Columns: There must be one column for the dataflow, one column per dimension, one column for the measure and one column per attribute. All dimensions defined in the related Data Structure Definition (DSD) are to be included. In case the SDMX RESTful 2.1 web service implementation supports a streaming mechanism, columns for all attributes defined in the DSD are present in the output, regardless of whether these attributes are used. Implementers have the possibility to add any other custom columns as required, e.g. serieskey, sender, prepared, etc.
* Column headers (first row):
  + For the first column, the dataflow column, always is the term *DATAFLOW*.
  + For a dimension column, is the dimension's ID or both ID and localised name (see option below).
  + For the measure column, always is the term *OBS\_VALUE*.
  + For an attribute column, is the attribute's ID or both ID and localised name (see option below).
  + For any custom column, is any custom but unique term.
* Column content (all rows after header):
  + For the dataflow column, is the reference to the *dataflow* in the following form: AGENCY:DATAFLOW\_ID(VERSION), or both reference and localised name (see option below).
  + For a dimension column, is the ID or both ID and localised name (see option below) of the observation's code in the corresponding dimension.
  + For the measure column, is the value of the observation.
  + For a coded attribute column, is the ID or both ID and localised name (see option below) of the code in the corresponding attribute. For attributes defined at series, group or dataset level, the codes are replicated for all observations concerned.
  + For an uncoded attribute column, is the value of the corresponding attribute. For attributes defined at series, group or dataset level, the values are replicated for all observations concerned.
  + For any custom column, contains any potentially localised custom content.
* Comma (,) separator for columns is used by default, but it is recommended for implementers to provide the response according to the locale of the client as indicated in the http Accept-Language header (which means that in some cases the semi-colon ‘;’ is acceptable as separator). See the second example below. The separator used in a message can be determined by the ninth character of the message, which is just after the fixed first column header term *DATAFLOW*.
* HTTP content negotiation, see [RFC 2616 - HTTP 1.1 Header Field Definitions](https://www.w3.org/Protocols/rfc2616/rfc2616-sec14.html)
  + Always use this mime-type in the Accept header:

application/vnd.sdmx.data+csv; version=1.0.0

* + The client can indicate preferred languages through the Accept-Language header, e.g.:

fr, en-gb;q=0.8, en;q=0.7

# Localised names

The first best language match according to the user’s preferred language choices in the http Accept-Language header (or if that is not available than according to the system's default language order) is to be used for each localisable name element. The message does however not indicate the returned language per localisable name element. In case that there is no such language match for a particular localisable name element, it is optional to return the element in a system-default language or alternatively to not return the element. **It is recommended to indicate all languages used anywhere in the message for localised name elements through http Content-Language response header (languages of the intended audience).**

# Optional parameters

Optional parameters can be added to the HTTP Accept header. They need to be separated by the character combination "; ".

* labels (id|both; default=id): This parameter applies to all Nameable SDMX Artefacts contained in the header and the body of the message:
  + If the parameter value is id then only the id of the Artefacts is displayed.
  + If the parameter value is both then the concatenated id and localised name of the Artefacts (see the section on [localised names](https://github.com/sdmx-twg/sdmx-csv/blob/develop/data-message/docs/sdmx-csv-field-guide.md#localised-names) on how the message deals with languages) separated by ": " are displayed. Note that the character combination ": " could also be part of the Artefact name and could therefore occur several times within the concatenated string.
* timeFormat (original|normalized; default=original):
  + If the parameter value is original then the *TIME-PERIOD* values are displayed in the SDMX *TIME\_PERIOD* format as originally recorded.
  + If the parameter value is normalized then the *TIME\_PERIOD* values are converted to the most granular [ISO 8601](https://www.iso.org/iso-8601-date-and-time-format.html) representation taking into account the highest frequency of the data in the message and the moment in time when the lower-frequency values were collected (which, e.g. at the ECB, is typically either at the beginning, middle or end of the reporting period). This eases comparisons and business analysis of multi-frequency values, e.g. in pivot tables. As an example, if annual and daily data are available in the message and the annual data were collected at the end of the reporting period, the formatted value for the annual period 2014 becomes 2014-12-31.

Support of above non-default parameters is not required by implementers.

# Examples

**1) HTTP Accept header: application/vnd.sdmx.data+csv; version=1.0.0**

DATAFLOW,DIM\_1,DIM\_2,DIM\_3,OBS\_VALUE,ATTR\_2,ATTR\_3,ATTR\_1,SERIESKEY

ESTAT:NA\_MAIN(1.6),A,B,2014-01,12.4,Y,"Normal, special and other values",N,A.B

ESTAT:NA\_MAIN(1.6),A,B,2014-02,10.8,Y,"Normal, special and other values",Y,A.B

The following default parameter settings are automatically applied:

* labels=id
* timeFormat=original
* *SERIESKEY* is a custom column.

**2) HTTP Accept header: application/vnd.sdmx.data+csv; version=1.0.0; labels=both**

**HTTP Accept-Language header: fr-FR, en;q=0.7**

DATAFLOW;DIM\_1: Dimension 1;DIM\_2: Dimension 2;DIM\_3: Dimension 3;OBS\_VALUE;ATTR\_2: Attribut 2;ATTR\_3: Attribut 3;ATTR\_1: Attribut 1;SERIESKEY

ESTAT:NA\_MAIN(1.6): Principaux agrégats des comptes nationaux;A: Value A;B: Value B;2014-01;12,4;Y: Oui;Normal, special and other values;N: Non;A.B

ESTAT:NA\_MAIN(1.6): Principaux agrégats des comptes nationaux;A: Value A;B: Value B;2014-02;10,8;Y: Oui;Normal, special and other values;Y: Oui;A.B

The following default parameter settings are automatically applied:

* timeFormat=original
* *SERIESKEY* is a custom column.

Note that in this example the client prefers French (fr) language with the France (FR) locale, but will also accept any type of English. Therefore, in the message the French language with the France locale is realized, transforming also the field separator from comma (,) to semicolon (;), and the decimal separator from dot (.) to comma (,).

**3) HTTP Accept header: application/vnd.sdmx.data+csv; version=1.0.0; labels=both; timeFormat=normalized**

DATAFLOW,DIM\_1: Dimension 1,DIM\_2: Dimension 2,DIM\_3: Dimension 3,OBS\_VALUE,ATTR\_2: Attribute 2,ATTR\_3: Attribute 3,ATTR\_1: Attribute 1,SERIESKEY

ESTAT:NA\_MAIN(1.6): National Accounts Main Aggregates,A: Value A,B: Value B,2014-01-01,12.4,Y: Yes,"Normal, special and other values",N: No,A.B

ESTAT:NA\_MAIN(1.6): National Accounts Main Aggregates,A: Value A,B: Value B,2014-02-01,10.8,Y: Yes,"Normal, special and other values",Y: Yes,A.B

The following default parameter settings are automatically applied:

* *SERIESKEY* is a custom column.