Tips & Tricks
Data Entry Tool
How to import files from Excel or Access into the DET

2

This document explains how data prepared in other programs (e.g. Excel, Oracle, Access, …) can be imported to the Data Entry Tool.
It is written based on the experience on the DET v8.1.1 b02 and will be adapted according to the new versions. You might need to contact your IT department in order to help you.

0. General overview

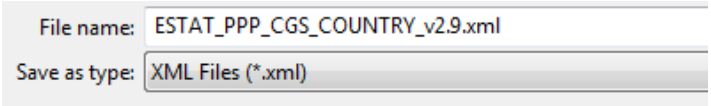
The DET writes and reads all price observations in XML files. XML files are nothing but text files that can be read in a web-browser or text editor. The content of the files is structured according to definitions laid down in a "Key Family" (or "Data Structure Definition"). Thus, a database program needs the Key Family to understand the content of an XML file. The PPP Key Family is compatible with SDMX standards (SDMX v2.0 format)(see www.sdmx.org).

Because of this compatibility, it is possible to convert XML files into CSV files (and back) that can be read by Excel or Access. It is thus possible to prepare price observations in a CSV file and then convert this into a XML file that can be imported into the DET, provided the CSV file follows the Key Family definitions. This document provides you with an overview of how this can be managed.

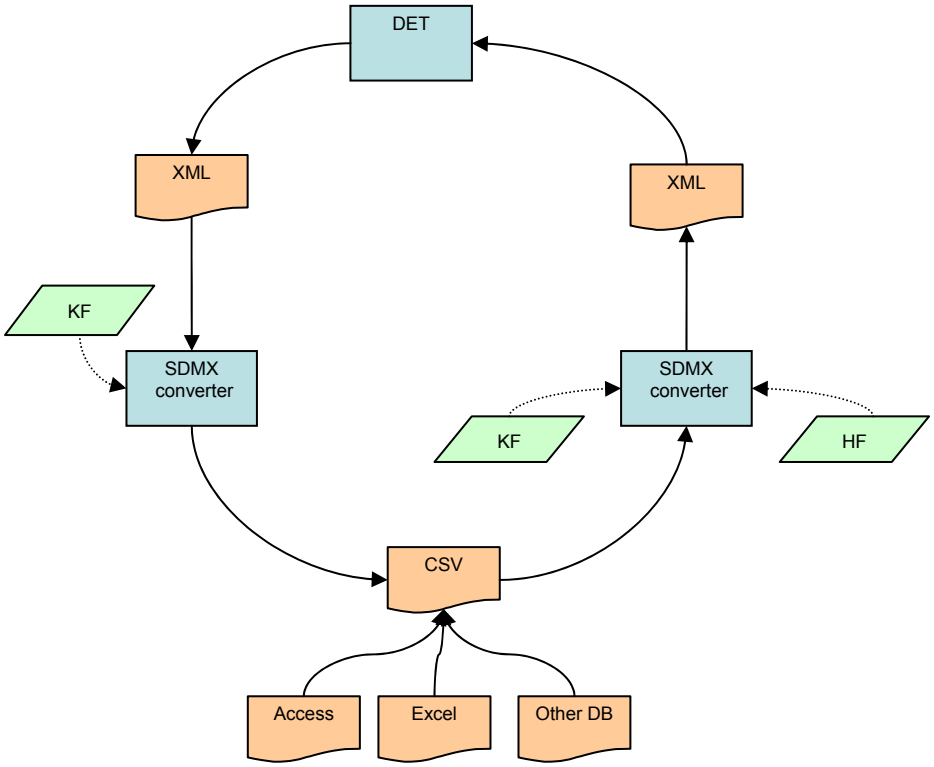
Here below you see the overview of how files can be exchanged between the DET and other databases. A tool named "SDMX converter" can be found on CIRCABC (see annex I) and is used for the conversion of the XML file to the CSV file (and vice versa).

To use this convertor, you also need to save the Data Structure Definitions (DSD or Key Family) from the DET homepage (see annex I). The latest version of the Data Structure Definitions will be available there.

Click on SDMX-ML then save the page as *.xml in your computer.



The diagram on the next page describes the whole cycle of starting and ending at the DET with the CSV file as an in between stage, in which this CSV file acts as the gateway to your database(s). Obviously, once you have created the CSV file and you are familiar with its format, you can skip this step for future uploads.



For the converter to work successfully a special attention has to be paid to the Annex II (Annex A or Annex B) when creating the CSV file. The annex gives the detailed description of the fields included in the file. Additionally you may export an XML out of the Data Entry Tool (DET) or use an XML file from a previous survey and convert it into a CSV to better understand the description.

1. Export out of the Data Entry tool

- Go to the Data Entry tool, import your SUA-file (Survey > Initialise List) and fill in 1 complete observation as an example.
- Go to "Dataset" select "Export" and save your file (e.g. Export.xml). Please notice that the DET adds automatically a prefix to the file name that contains the abbreviation and the year of the survey, e.g. PERAP2012Export.xml.

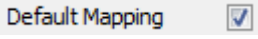
2. Use the SDMX-converter to convert to CSV

Open the SDMX-converter (Annex I) and fill in the following fields:

- Input: select the input file (dataset.xml) and the format (**CROSS SDMX**).
- Output: select the output file (e.g. Export.csv) and the format (**CSV**) (you would need to create an empty CSV file in advance. For this you may change the extension of an empty TXT file into CSV)
- DSD File: This is the file you saved from the DET homepage (see below General Overview step), named "ESTAT_PPP_CGS_COUNTRY_v2.10.xml"
- Use Registry : Select "False"
- Uncheck **Default Mapping** from the CSV/FLR/Excel section (see picture below) and

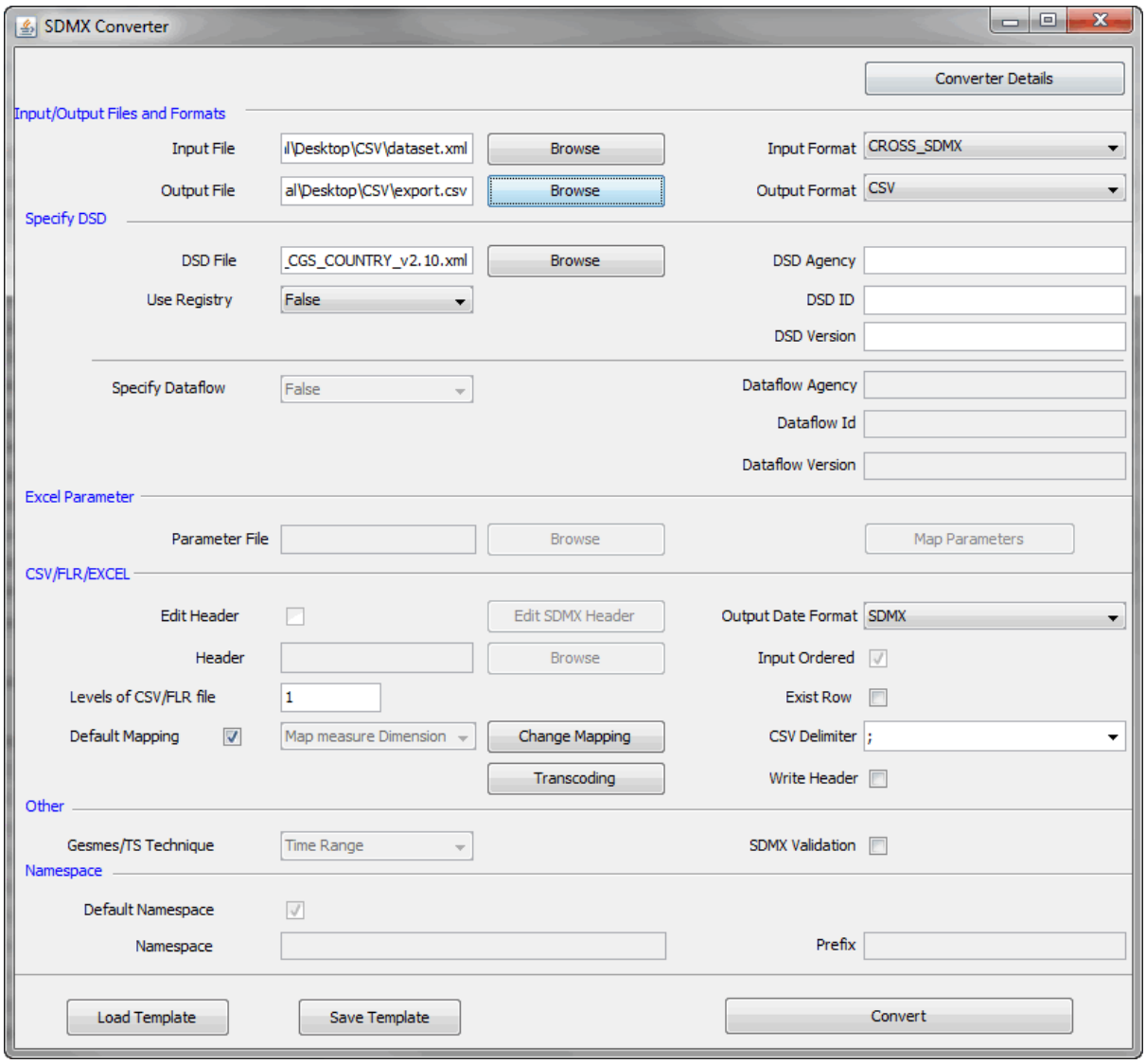
    - Option A. Choose 'Map CrossX measures' from the drop-down menu. Click on **Change mapping**. A pop-up window will appear, click on 'Default mapping' and click OK.
    - Option B. Choose 'Map measure dimension' from the drop-down menu. Click on **Change mapping**. A pop-up window will appear. Click on 'Default mapping' and click OK.

Note: Choosing between option A and option B is the decision of the user. The difference lays in the resulting CSV file, the structure definition will differ upon the choice. Option A corresponds to the description given in Annex A. Option B corresponds to the description given in Annex B.

Once you followed option A or B and complete the steps, the radio button next to **Default mapping** is automatically selected 

, do not uncheck it.

- Click on the "Convert"-button



You'll see the following message when the conversion was successful:

3. Fill the CSV with the data in your own DB's

Once the structure of the CSV is prepared, you can easily import this file in e.g. MS Access and fill it with your own observations. If you use Access or Excel for data entry, then the best way would be to program a macro that exports your data in the format of the CSV file.

An example of complete definition can be found in Annex A and Annex B.

Annex A includes the complete definition to be followed when the mapping 'Map CrossX measures' is used (this mapping is not available for all versions of the SDMX Converter). The definition includes **one row** for each observation.

Annex B includes the complete definition to be followed when the mapping 'Map measure dimension' is used. The definition includes **two rows** for each observation: one row corresponds to the price (value of column E is "P") and the second, to the quantity (value of column E is "Q"). Some of the other fields are completed in correlation with the value "P" or "Q".

Pay attention to the following rules:

- Don't add columns (e.g. don't add a primary key)
- Don't use quotes (' or ") to define text fields
- Each line in your CSV file should end with a semi colon (;)
- Field delimiter = semi colon (;)
- The delimiter for concatenated fields (e.g; Specify) is a pipe (|)
- Each item has a series of mandatory characteristics (e.g. number of hours worked, travel costs) to be reported during the price collection. Therefore column N may differ from item to item, all the parameters have to be included
- Pay attention to column 19 and 31 (VAT rate) which should be defined as decimals (decimals should be between 0 and 1).

Export your CSV-file once you have filled it with the data of your own DB's, e.g. Export_DB.csv

4. Use the SDMX-converter to convert to SDMX

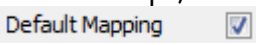
Open the SDMX-converter (Annex I) and fill in the following fields:

- Input : select the input file (Export_DB.csv) and the format (**CSV**)
- Output : select the output file (e.g. SDMXData_db.xml) and the format (**CROSS SDMX**)
- DSD File : The same as in step 2 (ESTAT_PPP_CGS_COUNTRY_v2.10.xml saved from DET homepage)
- Use Registry : Select "False"

Additionally, you need to specify the following parameters:

- Check **Edit Header**, click on **Edit SDMX Header** and fill in the ID (the eDamis dataset code e.g. PPP_PERAP_3). Depending on the version of the SDMX converter that is used, you may have to fill in as minimum other fields (e.g. Prepared, Provide Sender Info). All obligatory fields are marked with *
- Check if the CSV Delimiter is filled in correctly (;)
- Uncheck **Default Mapping** and if the CSV file contains one row for each observation follow steps described below under Option A, if the CSV file contains two rows for each observation follow the steps described below under Option B,

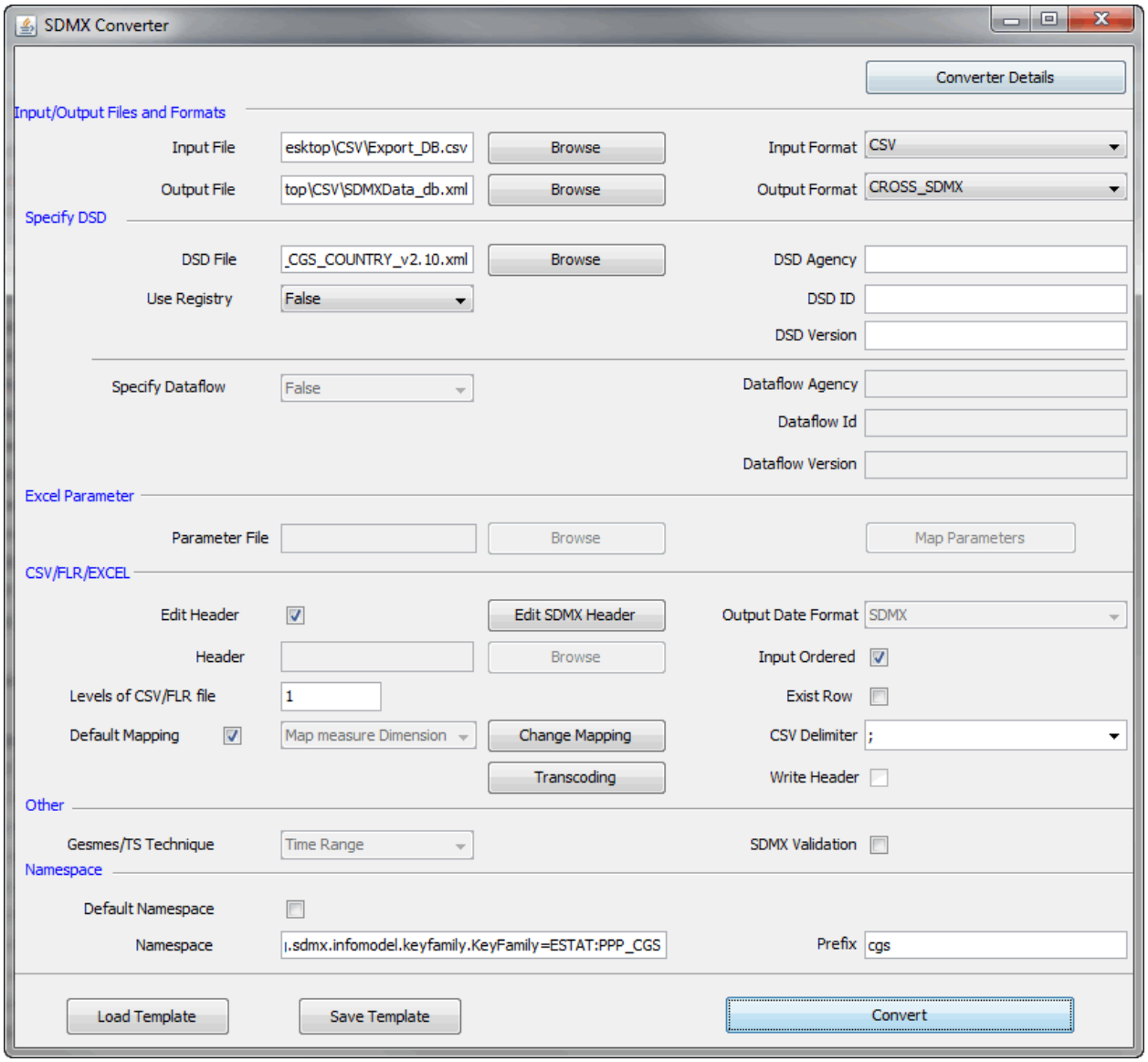
- Option A. Choose 'Map CrossX measures' from the drop-down menu. Click on Change mapping. A pop-up window will appear, click on 'Default mapping' and click OK.
- Option B. Choose 'Map measure dimension' from the drop-down menu. Click on Change mapping. A pop-up window will appear. Click on 'Default mapping' and click OK.

Once you followed option A or B and complete the steps, the radio button next to **Default mapping** will automatically be selected 

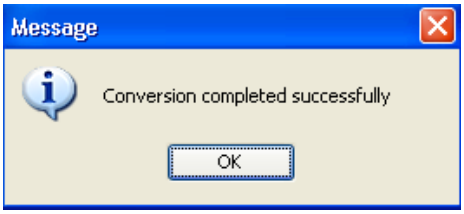
, do not uncheck it.

- In the Namespace section, uncheck the Default Namespace
- Fill in the Namespace:

  urn:sdmx:org.sdmx.infomodel.keyfamily.KeyFamily=ESTAT:PPP_CGS
- Fill in the Prefix of the namespace: **cgs**
- click on "Convert"



Again, you'll see the following message when the conversion was successful:

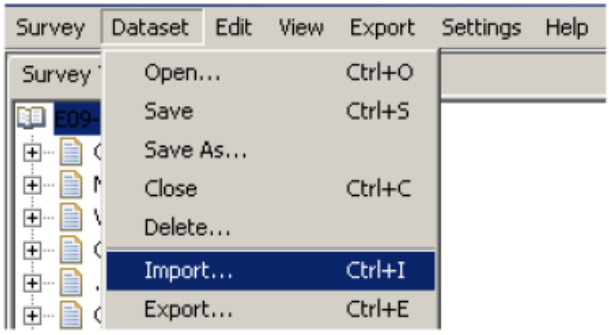


If the conversion is not successful you will get an error message. Check if the message refers to the input file or to the output file. Possible reasons for this could be:

- format errors in the .csv file
- parameters were not filled in
- the .xml file is not filled in correctly
- the ID in the SDMX header is not filled in
- the format of the input/ output files is not choose as it should
- your file may contain special-not supported characters.

5. Re-Import the xml-file into the DET

You can easily import the created SDMX file into the DET. You just need to use the feature Import in the menu Dataset.



6. Good to know

- Don't forget to update your own Survey Details in the DET on clicking "Settings" > "Contact Information"
- In the CSV file, don't add columns
- Don't use quotes (' or ") to define text fields
- Each line in your CSV file should end with a semi colon (;)
- Field delimiter = semi colon (;)
- Use the dot "." as decimal separator
- The delimiter for concatenated fields (e.g; Specify) is a pipe (|)

- Each item has a series of mandatory characteristics (e.g. number of hours worked, travel costs) to be reported during the price collection. Therefore column N may differ from item to item, all the parameters have to be included
- Pay attention to column 19 and 31 (VAT rate) which should be defined as decimals (decimals should be between 0 and 1).

ANNEX I

SDMX-converter:

The SDMX-converter (and documentation) can be found on the following CIRCABC link here below (copy the following link in your browser in case it is not working by the simple click).

It's recommended to use version 3.1.7 Interim SDMX 2.0 (v.3.1.7_2014.06.30) of the converter or ask Eurostat if a newer version is compatible

https://circabc.europa.eu/w/browse/1cdc3796-1b7c-4f62-8523-b041ea1a4eaf

Download the .ZIP file that can be found in the Application Package folder (e.g. SDMX_Converter_Platform-Independent_v.3.1.7_2014.06.30-interim-sdmx20.zip). After saving and unzipping the software, click on Converter.bat and Run the application.

DET homepage:

https://webgate.ec.europa.eu/ppp_tool/DET/index.html

ANNEX II

ANNEX II.A : Description of the CSV-file if using Map CrossX measures

	Column	Description	Allowed Values	Value Pattern (if any)	Important Notes	Example
1	Column A	Survey code	Any valid survey code (e.g. FBETO, FRHEA, HOGAR etc)		The code should be always in capital letters. A reference of the survey codes can be found in the used DSD file.	HOGAR
2	Column B	Country code	Any valid country code (e.g. IS, CY etc)		The code should be always in capital letters. A reference of the survey codes can be found in the used DSD file.	CY
3	Column C	Item code	Any valid item code that belongs to the survey			11.04.31.1.01.ac
4	Column D	N° of observation	Positive integers (e.g. 1, 2 etc)			1
5	Column E	Year	Positive integers (e.g. 2011, 2012 etc)			2012
6	Column F	Price of the observation	Positive decimal		The '.' should always be used as the decimal separator	20.85
7	Column G	Quantity of the observation	Positive decimal		The '.' should always be used as the decimal separator	1.5
8	Column H	Year and month of observation (with hyphen inbetween)		Year-Month	If column 5 = Q then leave blank	2012-6
9	Column I	Currency	Any valid currency code (e.g. EUR)			EUR
10	Column J	Brand	Any text or blank if brand is not available			ABrand
11	Column K	Model	Any text or blank if model is not available			AModel
12	Column L	Shop Type	Any valid shop type codes			5
13	Column M	Shop Identifier	Any text or blank if not available			sID
14	Column N	Specify: name of Specify parameter & "=" & value. Concatenated	Text or blank if not available	Specify_Name_1=Value1|Specify_Name_2=Value_2		NUMBER OF HOURS WORKED=2.5|TRAVEL COSTS=0|PRICE OF MATERIALS=125
15	Column O	Separator used in column 14	|		Always use symbol | as separator	|
16	Column P	Comments at observation level	Any text or blank if not available			Some comments
17	Column Q	Concatenation of OTHER & PRICE COLLECTOR	Text or blank if not available	OTHER=Value_for_other|COLLECTOR=Value_for_collector		OTHER=Any other comments|COLLECTOR=The price collector
18	Column R	Discount Flag	N, Q, R, T			N
19	Column S	VAT at item level	Positive decimal between 0-1		The '.' should always be used as the decimal separator	0.20
20	Column T	Representativity	true, false or blank		All letters should be in lowercase. Blank equals to undefined	true
21	Column U	Seasonality	true, false or blank		All letters should be in lowercase. Blank equals to false	false
22	Column V	Specify if average prices are being reported	true, false or blank		All letters should be in lowercase. Blank equals to false	true
23	Column W	Flag	O (valid observation), E (eliminated			O

			observation)			
24	Column X	Item Comments	Any text or blank if not available			Item comments here
25	Column Y	Item comments for local use	Text or blank if not available	LOCALCOMMENT= Value_Of_Comment		LOCALCOMMENTS=Some comments for local use
26	Column Z	Specify if item is finalised or not	true, false or blank		All letters should be in lowercase. Blank equals to false	true
27	Column AA	Variation Coefficient at observation level	Positive decimal		The '.' should always be used as the decimal separator.	5.64
28	Column AB	Number of observations at observation level	Positive integers			12
29	Column AC	Minimum of prices at observation level	Positive decimal		The '.' should always be used as the decimal separator. If column 22 = false then leave blank.	4.74
30	Column AD	Maximum of prices	Positive decimal		The '.' should always be used as the decimal separator. If column 22 = false then leave blank.	7.56
31	Column AE	VAT at observation level	Positive decimal between 0-1		The '.' should always be used as the decimal separator. If column 22 = false then leave blank.	0.1523

ANNEX II.B : Description of the CSV-file if using Map measure Dimension

	Column	Description	Allowed Values	Value Pattern (if any)	Important Notes	Example
1	Column A	Survey code	Any valid survey code (e.g. FBETO, FRHEA, HOGAR etc)		The code should be always in capital letters. A reference of the survey codes can be found in the used DSD file.	HOGAR
2	Column B	Country code	Any valid country code (e.g. IS, CY etc)		The code should be always in capital letters. A reference of the survey codes can be found in the used DSD file.	CY
3	Column C	Item code	Any valid item code that belongs to the survey			11.04.31.1.01.ac
4	Column D	N° of observation	Positive integers (e.g. 1, 2 etc)			1
5	Column E	P(rice) or Q(uantity) of the observation	**P, Q**			P
6	Column F	Year	Positive integers (e.g. 2011, 2012 etc)			2012
7	Column G	Value of column 5	Positive decimal		The '.' should always be used as the decimal separator	20.85
8	Column H	Year and month of observation (with hyphen inbetween)		Year-Month	If column 5 = Q then leave blank	2012-6
9	Column I	Currency	Any valid currency code (e.g. EUR)			EUR
10	Column J	Brand	Any text or blank if brand is not available		If column 5 = Q then leave blank	ABrand
11	Column K	Model	Any text or blank if model is not available		If column 5 = Q then leave blank	AModel
12	Column L	Shop Type	Any valid shop type codes		If column 5 = Q then leave blank	5
13	Column M	Shop Identifier	Any text or blank if not available		If column 5 = Q then leave blank	sID
14	Column N	Specify: name of Specify parameter & "=" & value. Concatenated	Text or blank if not available	Specify_Name_1=Value1|Specify_Name_2=Value_2	If column 5 = Q then leave blank	NUMBER OF HOURS WORKED=2.5|TRAVEL COSTS=0|PRICE OF MATERIALS=125
15	Column O	Separator used in column 14	|		If column 5 = Q then leave blank. Always use symbol | as separator	|
16	Column P	Comments at observation level	Any text or blank if not available		If column 5 = Q then leave blank	Some comments
17	Column Q	Concatenation of OTHER & PRICE COLLECTOR	Text or blank if not available	OTHER=Value_for_other|COLLECTOR=Value_for_collector	If column 5 = Q then leave blank	OTHER=Any other comments|COLLECTOR=The price collector
18	Column R	Discount Flag	N, Q, R, T		If column 5 = Q then leave blank	N
19	Column S	VAT at item level	Positive decimal between 0-1		The '.' should always be used as the decimal separator	0.20
20	Column T	Representativity	true, false or blank		All letters should be in lowercase. Blank equals to undefined	true
21	Column U	Seasonality	true, false or blank		All letters should be in lowercase. Blank equals to false	false

22	Column V	Specify if average prices are being reported	true, false or blank		All letters should be in lowercase. Blank equals to false	true
23	Column W	Flag	O (valid observation), E (eliminated observation)		If column 5 = Q then leave blank	O
24	Column X	Item Comments	Any text or blank if not available			Item comments here
25	Column Y	Item comments for local use	Text or blank if not available	LOCALCOMMENT= Value_Of_Comment		LOCALCOMMENTS=Some comments for local use
26	Column Z	Specify if item is finalised or not	true, false or blank		All letters should be in lowercase. Blank equals to false	true
27	Column AA	Variation Coefficient at observation level (if column E = Q leave blank)	Positive decimal		The '.' should always be used as the decimal separator. If column 5 = Q then leave blank	5.64
28	Column AB	Number of observations at observation level	Positive integers		If column 5 = Q then leave blank	12
29	Column AC	Minimum of prices at observation level	Positive decimal		The '.' should always be used as the decimal separator. If column 5 = Q then leave blank. If column 22 = false then leave blank.	4.74
30	Column AD	Maximum of prices	Positive decimal		The '.' should always be used as the decimal separator. If column 5 = Q then leave blank. If column 22 = false then leave blank.	7.56
31	Column AE	VAT at observation level	Positive decimal between 0-1		The '.' should always be used as the decimal separator. If column 5 = Q then leave blank. If column 22 = false then leave blank.	0.1523