

Exercise 5

Using SDMX Converter to retrieve data from a mapped Excel file

In this exercise, you will use Eurostat SDMX Converter and the Global SDG Data Structure Definition to retrieve data from an Excel file. This is the same dataset that was used in the previous exercise, but it has been modified to add SDMX-SDG codes and mappings.

1. Open file **Exercise 5.xlsx**.
2. Observe that new columns have been added:
 - Series Code
 - Unit Code
 - Country Code

These columns contain valid series, unit, and reference area codes from the SDG DSD.

3. Observe that columns Age and Sex have been updated to use SDG DSD codes. For example:
 - **25+** became **Y_GE25**, which is the code for age group “25 years old and over”.
 - **Female** became **F**, which is the code for sex “Female”.
 - **Total** sex became **_T** (“Total or no breakdown by sex”)
4. Note that rows where sex was not provided, indicating the data is not broken down by sex, received code **_T** (“Total or no breakdown by sex”). Similarly, rows without age received code **_T** (“Total or no breakdown by age”).
 - Every dimension must receive a valid code, even if it is not used in a particular indicator.
 - By convention, all SDG code lists use value **_T** when there is no breakdown by that dimension, whether or not the dimension is applicable. For example, series “Total official flows for water supply and sanitation, by recipient (6.a.1)” will always have value **_T** for its Age, Sex, Occupation, and other irrelevant dimensions.
 - By contrast, in some indicators a specific code is expected for Age or Sex. E.g. series “Women in managerial positions (5.5.2)” is expected to have code **F** in its Sex dimension, while “series Infant mortality rate (3.2.1)” is expected to have value **Y0** (“under 1 year old”) in its Age dimension.
5. Open the Parameters worksheet and inspect the mappings.

6. Note the dimensions and attributes that have a fixed value. Why?
7. Note that some columns in the spreadsheet, such as Series or Country, were not mapped. Why not?
8. Launch SDMX Converter by clicking on the file Converter.bat in your exercise folder.

9. SDMX Converter opens. On the first screen, you select the input file containing the data (Exercise 5.xlsx), the output file where the SDMX data will be written, and the format of the output file.

The screenshot shows the SDMX Converter application window. The interface is divided into several sections: 'Input', 'Operation', and 'Output'. The 'Operation' section has three radio buttons: 'Convert', 'Validate', and 'Convert and Validate', with the latter selected. The 'Input' section has an 'Input File' field with a 'Select' button. The 'Output' section has an 'Input Format' dropdown menu, an 'Output Format' dropdown menu set to 'GENERIC_SDMX', and a 'Select Path' field with a 'Select' button. At the bottom, there are 'Res' and 'Next' buttons, and a progress bar showing 0%.

1. Select: **Convert and Validate**

2. Locate input file by clicking **Select**

3. Make sure input format is **Excel**

4. Select output format: **STRUCTURE_SPECIFIC_DATA_2_1**

5. Select output file: **Ex5.xml**

6. Click **Next**

10. On this screen, you load the Data Structure Definition. Once the structure file has been located, you click Detect Structures to identify any DSD(s) contained in the file.

The screenshot shows the SDMX Converter application window. The interface is divided into several sections: **StructureType** with radio buttons for **DSD** (selected) and **Dataflow**; **Structure File** with a **File** input field, a **Select** button, and a **Detect Structures** button; and **Structure Identification** with fields for **Agency**, **Artefact ID**, and **Artefact Version**. At the bottom, there are **Reset** and **Next** buttons. Four callout boxes with arrows point to these key elements: **7. Select: DSD** points to the DSD radio button; **8. Click Select and locate the DSD file (SDG_DSD.xml)** points to the File input field; **9. Click Detect Structures** points to the Detect Structures button; and **10. Click Next** points to the Next button.

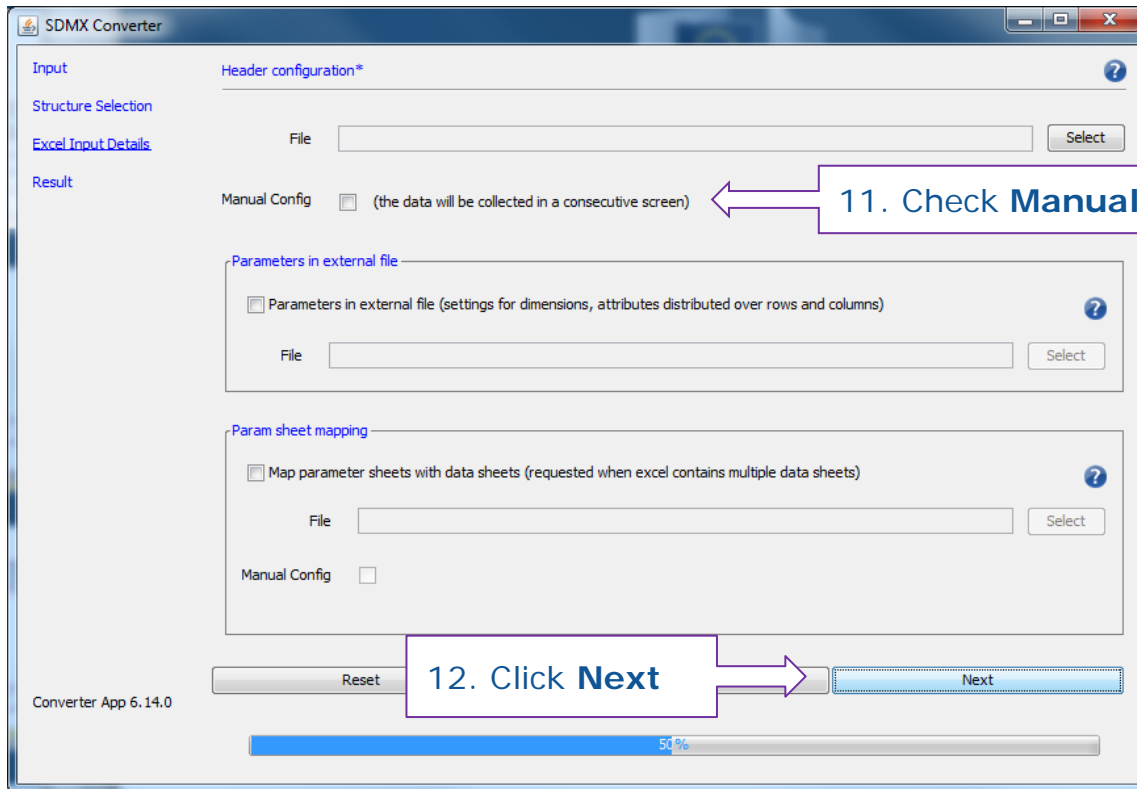
7. Select: **DSD**

8. Click **Select** and locate the DSD file (**SDG_DSD.xml**)

9. Click **Detect Structures**

10. Click **Next**

11. On this screen, you can load an SDMX message header and mapping parameters.
Every SDMX message must have a header in a defined format, but it is not used in this exercise.
As for parameters, they are contained in the Excel file.



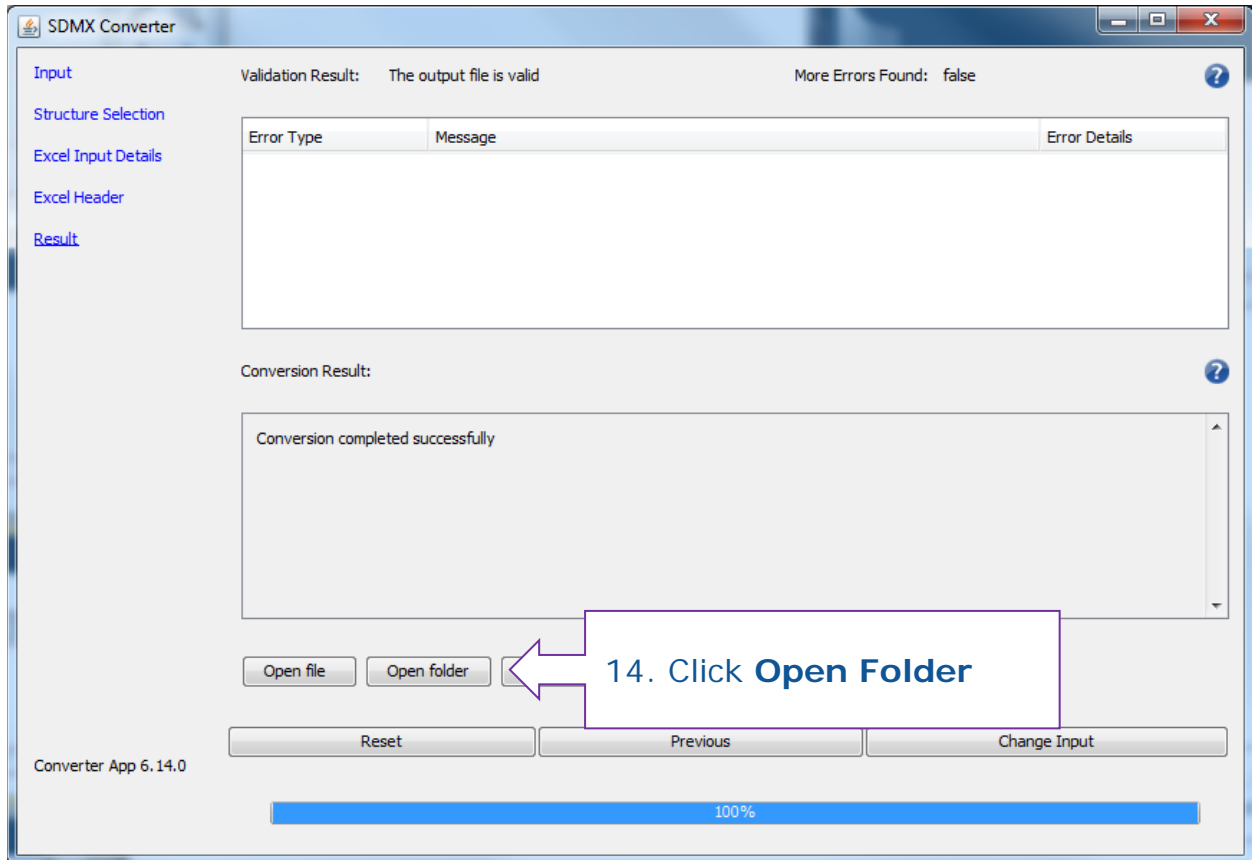
12. On this screen, you can configure the header. You may wish to update the header fields, or simply leave the defaults because we are not using them.

The screenshot shows the SDMX Converter application window. The title bar reads "SDMX Converter". On the left is a sidebar with the following items: "Input", "Structure Selection", "Excel Input Details", "Excel Header", and "Result". The main area is titled "Header" and contains three buttons: "Erase Fields", "Load Default", and "Save Header". Below these are three input fields: "Id*" with the value "JD014", "Test" with an unchecked checkbox, and "Prepared*" with the value "Sep 11, 2019".

Below the "Header" section are two columns of input fields. The left column is titled "Sender" and contains fields for "Id*" (value "BIS"), "Name", "Contact Name", "Contact Dept", "Contact Role", "Contact Phone", and "Contact Fax". The right column is titled "Receiver" and contains fields for "Id", "Name", "Contact Name", "Contact Dept", "Contact Role", "Contact Phone", and "Contact Fax".

At the bottom of the main area are two buttons: "Reset" and "Next". A callout box with a purple border and arrow points to the "Next" button, containing the text "13. Click Next". A progress bar at the very bottom shows "60%" completion.

13. If the steps have been completed correctly, data will be retrieved. Click Open Folder, find and open the SDMX file with an XML viewer such as Notepad++, or a browser such as Internet Explorer or Firefox.



Exercise 6

Mapping an Excel file

In this exercise, you will map an Excel spreadsheet with SDG data to the pilot SDG Data Structure Definition and use SDMX Converter to retrieve the data into an SDMX file.

1. Open file **Exercise 6.xlsx**.
2. Inspect the data in the file. Note that this is the same dataset as that used in the previous exercise, but the format is different. In this file, the data format is record-based, i.e. each row in the file contains one observation. This is generally easier to map than the time-series format.
3. Leave the spreadsheet open and also open the spreadsheet used for the previous exercise, **Exercise 5.xlsx**.
4. Copy the mappings from Exercise 5 to Exercise 6:
 - Right-click on the spreadsheet **Parameters** and select **Move or Copy...**
 - In the **To book:** dropdown, select **Exercise 6.xlsx**.
 - Check **Create a copy**
 - Click **OK**.
 - Return to **Exercise 6.xlsx** and ensure that worksheet **Parameters** has been copied.
5. Although the mappings have been copied to the new spreadsheet, they need to be updated because the data is in a different format. Since this is a record-based format, the concepts map to columns not rows. There must be a column with codes for **SERIES**, **REF_AREA**, **UNIT_MEASURE**, **AGE**, and **SEX** dimension.
 - Open worksheet **Data**.
 - Right-click column **Series**, click **Copy**. Right-click the column next to it, select **Insert Copied Cells**. You should have a copy of the **Series** column now. Name it **Series Code**.
 - Repeat the above step for Country, Age, Sex, and name the new columns **Country Code**, **Age Code**, **Sex Code** respectively.
6. Working with worksheet **Data**, provide valid codes for series:

Return to spreadsheet **Exercise 5.xlsx**. Identify the code for series **Number of deaths and missing persons attributed to disasters (number)**. Copy the code (**VC_DSR_MMHN**).

- Open the spreadsheet **Exercise 6.xlsx**. Select the column **Series Code**. Click Ctrl-F.
- The **Find and Replace** dialog box opens. Click **Replace**.

- In the **Find what:** box, paste the series description: **Number of deaths and missing persons attributed to disasters (number)**. In the **Replace with:** box, paste the series code **VC_DSR_MMHN**. Click **Replace All**.
 - Repeat the above steps for the second series, **Employed population below international poverty line, by sex and age (%)**.
7. Repeat Step 6 for Country Code, Age Code, and Sex Code, replacing descriptions with valid codes.
 - Ensure that each cell in these columns has a valid code, including cells that are currently empty.
 - Tip: when replacing Sex codes, the order should be Female, then Male and Total; otherwise, be sure to match the letter case when you replace.
 8. Now, you need to add units, which are currently not on the spreadsheet. Each of the two series in the spreadsheet uses its own unit.
 - Insert an empty column next to Series Code and name it **Unit Code**.
 - Open spreadsheet **Exercise 5.xlsx**. Find and copy unit code for the first series **Number of victims of intentional homicides (number)**.
 - Paste the code into cells in column **Unit Code** that correspond to the first series.
 - Repeat the steps above for the second series' unit.
 9. Open the spreadsheet **Parameters**. Examine each mapping and update as necessary.
 - Since the format is record-based, there will be no dimensions that map to rows.
 - For column positions, you can use either letters, as in the spreadsheet (A, B, C,...) or numbers (1, 2, 3, ...).
 - Be sure to map the code columns, not descriptions.
 - Each dimension and mandatory attribute must be mapped!
 - The **TIME_DETAIL** attribute should be mapped to the same column as the **TIME_PERIOD** dimension.
 10. In your **Parameters** worksheet, find **DataStart** with the cell that contains the first observation value. The column should be the one that contains the observation, and the row should be 2 (since the headers are in the first row).
 11. Update **NumColumns**. Since we only have one observation per row, the value should be 1.
 12. Save spreadsheet **Exercise 6.xlsx** and leave it open.
 13. Open the manual for Exercise 5. Follow the steps to retrieve the data from the spreadsheet you have mapped.

- On the first screen, choose Input File **Exercise 6.xlsx** and use **Ex6.xml** for the output file. The remaining steps are the same as in Exercise 5.

14. Open the SDMX file you created with Notepad++ and inspect its contents.