# 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.xslx**.
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.xslx**.
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.xslx**.
	* 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 (%)**.
1. 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.
2. 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.
3. 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.
4. 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).
5. Update **NumColumns**. Since we only have one observation per row, the value should be 1.
6. Save spreadsheet **Exercise 6.xslx** and leave it open.
7. 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.xslx** and use **Ex6.xml** for the output file. The remaining steps are the same as in Exercise 5.
8. Open the SDMX file you created with Notepad++ and inspect its contents.