# Exercise 1

## Mapping an Excel file

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

1. You will need to look up codes from the global SDG DSD. Use one of the following methods to view the DSD:
* Open the SDMX Global Registry at <https://registry.sdmx.org> and click **SDMX Global Registry Home**. Expand **Data** and click **Data Structures**. Locate and select the SDG DSD and Click **View Data Structure Definition**.
* Open the SDMX-SDG page at <https://unstats.un.org/sdgs/iaeg-sdgs/sdmx-working-group/>. Locate and download the **SDG DSD Matrix**. This is a spreadsheet that contains the DSD including the Concept Scheme, Code Lists, Dataflows, and cube-region Content Constraints. You can use the code list spreadsheets to look up the codes.
1. Open the SDMX-SDG page at <https://unstats.un.org/sdgs/iaeg-sdgs/sdmx-working-group/> and download the spreadsheet marked SDG Series Content Constraints Matrix. This will help you find allowable dimension and attribute values.
2. Open the spreadsheet file with your national data.
3. Inspect data in the file. Note that 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, where there are several observations per row.
4. Open the **Parameters** worksheet. Note that SDG concept names are there but mappings are missing and need to be added. 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**, **SEX** and **URBANISATION** dimensions.
	* Open the data worksheet.
	* Note that there already is a column with Series and Reference Area codes. They do not need to be mapped.
	* Right-click column **units**, click **Copy**. Right-click the column next to it, select **Insert Copied Cells**. You should have a copy of the **units** column now. Name it **Unit Code**.
	* Repeat the above step for Age, Sex, and Location, and name the new columns **Age Code**, **Sex Code,** and **Urbanisation Code** respectively.
5. Working with worksheet **Data**, provide valid codes for the unit:
	* Open the SDG Content Constraint Matrix. Find and copy unit code for the first series **Proportion of children moderately or severely stunted (%)** – note that it is **PT** (percent).
	* Open the spreadsheet with national data. Select the column **Unit Code**. Click Ctrl-F.
	* The **Find and Replace** dialog box opens. Click **Replace**.
	* In the **Find what:** box, paste the source unit code: **PER\_POP\_U5**. In the **Replace with:** box, paste the SDMX unit code **PT**. Click **Replace All**.
	* ****Repeat the steps above for the second series’ unit.

1. Repeat Step 7 for Age Code, Sex Code, and Urbanisation Code, replacing descriptions with valid codes.
	* Note that where the Content Constraints Matrix indicates **ALL** in a particular dimension, that means any code is allowed. You may need to consult the Global DSD (the matrix or the Global Registry) to find correct codes.
	* Ensure that each cell in these columns has a valid code, including cells that are currently empty.
	* Tip: when replacing Sex codes, replace Female first, then Male and Total; otherwise, be sure to match the letter case when you replace.
2. 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.
	* Be sure to map the code columns, not descriptions.
	* Each dimension and mandatory attribute must be mapped!
3. In your **Parameters** worksheet, find **DataStart** and update it 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).
4. Update **NumColumns**. Since we only have one observation per row, the value should be 1.
5. Save and close the data spreadsheet.
6. Open SDMX Converter at <https://webgate.ec.europa.eu/sdmxconverter>.
7. Convert the spreadsheet to Structure-Specific Format using dataflow **DF\_SDG\_GLC**.
8. Congratulations! You mapped an Excel spreadsheet to the global SDG DSD and converted it to SDMX.