SDMX Converter

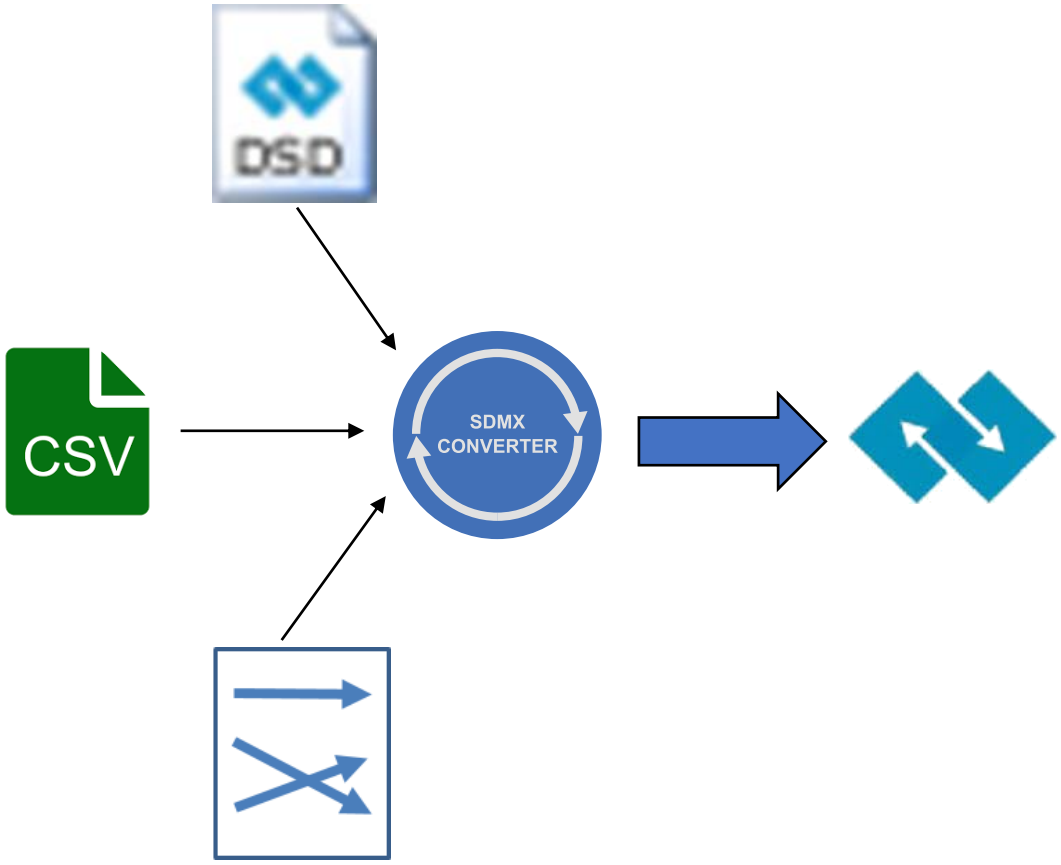
SDMX Converter

- Software developed by Eurostat
- Can be used to convert data from a variety of format into SDMX and vice versa
- Supports CSV, Excel, DSPL, and others
- Can be used to transform….
  - Non-SDMX data to SDMX
  - SDMX to non-SDMX
  - SDMX format to another SDMX format

SDMX Converter: Applications

- SDMX Converter is available as
  - Desktop application with a Graphical User Interface
  - Command-line application
  - Web service
  - Java library
  - **NEW:** Web application
- Download from https://ec.europa.eu/eurostat/web/sdmx-infospace/sdmx-it-tools/sdmx-converter

Converting data to SDMX



- To transform data to SDMX using the SDMX Converter, you need
  - Source data as CSV, DSPL, Excel, etc.
  - A Data Structure Definition (DSD) according to which the SDMX dataset will be structured
  - Mappings that show how the source data maps to the concepts of the Data Structure Definition
  - As always in setting up SDMX exchange, configuring mappings takes the most time and effort

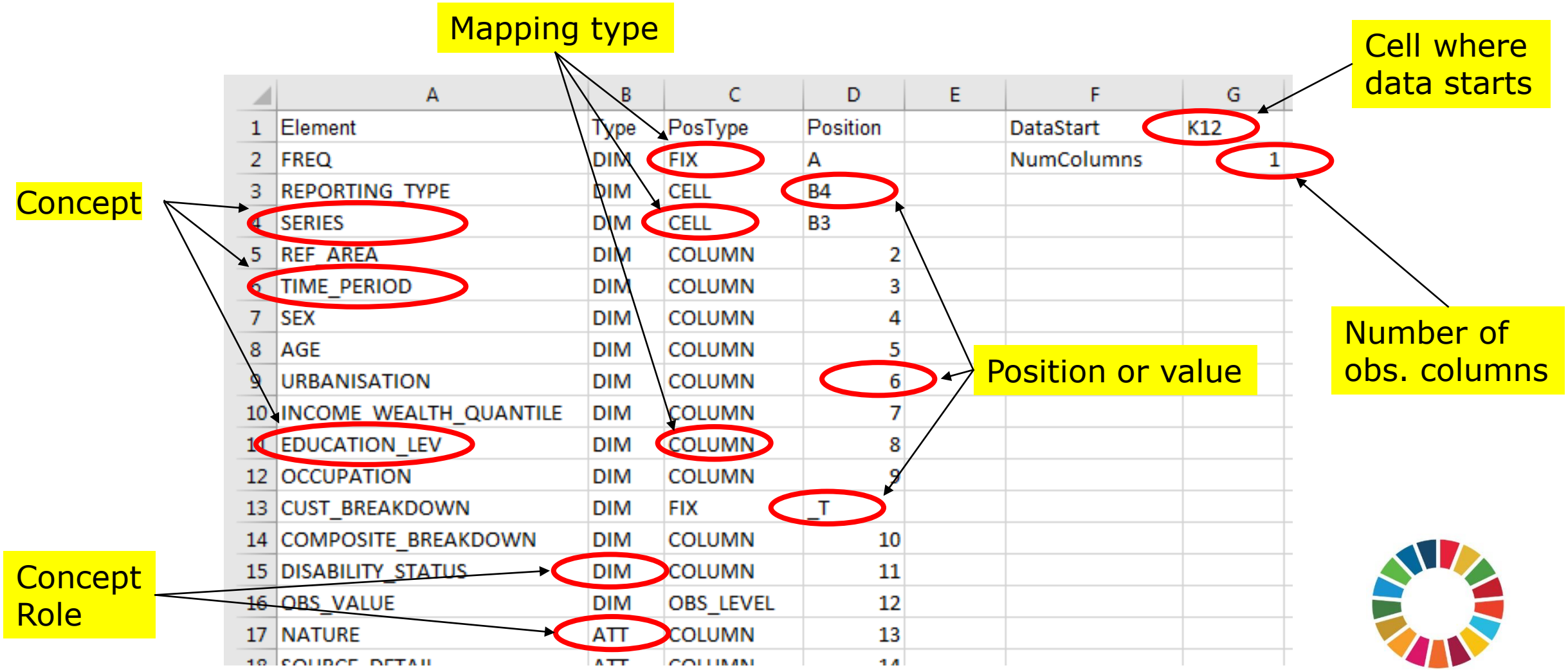
Using SDMX Converter with Excel

- Data and mappings can be placed into the same spreadsheet
- Additional information can be added to facilitate data entry
  - E.g. code lists for validation and display of descriptions

Worksheet names

- Worksheet named **Parameters** contains mappings
  - Only one can be used at a time
  - Shows how cells, rows, and columns map to the DSD dimensions and attributes
- Worksheet names starting with **Val** are ignored
  - Can be used to store code lists or other ancillary information
- All other worksheets are considered to contain data and will be processed

Excel Mappings



	A	B	C	D	E	F	G
1	Element	Type	PosType	Position		DataStart	K12
2	FREQ	DIM	FIX	A		NumColumns	1
3	REPORTING_TYPE	DIM	CELL	B4			
4	SERIES	DIM	CELL	B3			
5	REF_AREA	DIM	COLUMN	2			
6	TIME_PERIOD	DIM	COLUMN	3			
7	SEX	DIM	COLUMN	4			
8	AGE	DIM	COLUMN	5			
9	URBANISATION	DIM	COLUMN	6			
10	INCOME_WEALTH_QUANTILE	DIM	COLUMN	7			
11	EDUCATION_LEV	DIM	COLUMN	8			
12	OCCUPATION	DIM	COLUMN	9			
13	CUST_BREAKDOWN	DIM	FIX	_T			
14	COMPOSITE_BREAKDOWN	DIM	COLUMN	10			
15	DISABILITY_STATUS	DIM	COLUMN	11			
16	OBS_VALUE	DIM	OBS_LEVEL	12			
17	NATURE	ATT	COLUMN	13			

Excel mappings worksheet

- **Element**: name of the DSD concept
- **Type**: role of the concept
  - **DIM:** Dimension
  - **ATT:** attribute
- **DataStart**: the first cell containing an observation
- **NumColumns**: number of observations per row

Exercise 5: Using SDMX Converter

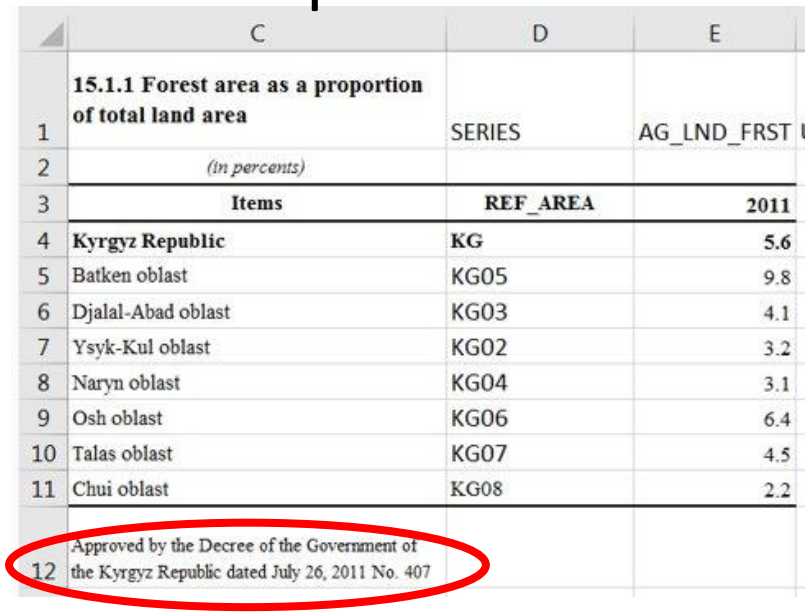
- Use SDMX Converter to retrieve data from a pre-mapped spreadsheet and convert it to SDMX

Column PosType: mapping or position type

- The following mapping/position types are supported:
  - **CELL**
  - **ROW**
  - **COLUMN**
  - **FIX**
  - **OBS_LEVEL**
  - **MIXED**
  - **SKIP**

Mapping type: CELL

- The value for the entire dataset is provided in the cell provided in the column **Position**
- E.g. if the spreadsheet is expected to only contain data for a single country, its code can be provided in a cell.



	C	D	E
1	**15.1.1 Forest area as a proportion of total land area**	SERIES	AG_LND_FRST
2	*(in percents)*		
3	**Items**	**REF_AREA**	**2011**
4	**Kyrgyz Republic**	**KG**	**5.6**
5	Batken oblast	KG05	9.8
6	Djalal-Abad oblast	KG03	4.1
7	Ysyk-Kul oblast	KG02	3.2
8	Naryn oblast	KG04	3.1
9	Osh oblast	KG06	6.4
10	Talas oblast	KG07	4.5
11	Chui oblast	KG08	2.2
12	Approved by the Decree of the Government of the Kyrgyz Republic dated July 26, 2011 No. 407		

Mapping type: ROW

- Values for the concept are stored in the row specified in column **Position**

	C	D	E	F	G	H	I
1	**1.1.1 Proportion of population living below the international poverty line (1.9 USD)**						
2	*(as a per cent to number of population)*						
3		SERIES:	SI_POV_DAY	AGE:	_T	UNIT	PT
4		URBANISATION:	_T	EDU_LEV	_T	UNIT_M	0
5	**Items**	**REF_AREA**	**2007**	**2008**	**2009**	**2010**	**2011**
6	**Kyrgyz Republic**	KG	0.13	0.09	0.29	0.28	**0.01**
7	Batken oblast	KG05	0.00	0.27	0.00	0.21	0.00
8	Djalal-Abad oblast	KG03	0.20	0.00	0.00	0.22	0.00
9	Ysyk-Kul oblast	KG02	0.05	0.34	0.58	0.11	0.00
10	Naryn oblast	KG04	0.13	0.63	0.03	3.46	0.31
11	Osh oblast	KG06	0.33	0.03	0.00	0.09	0.00

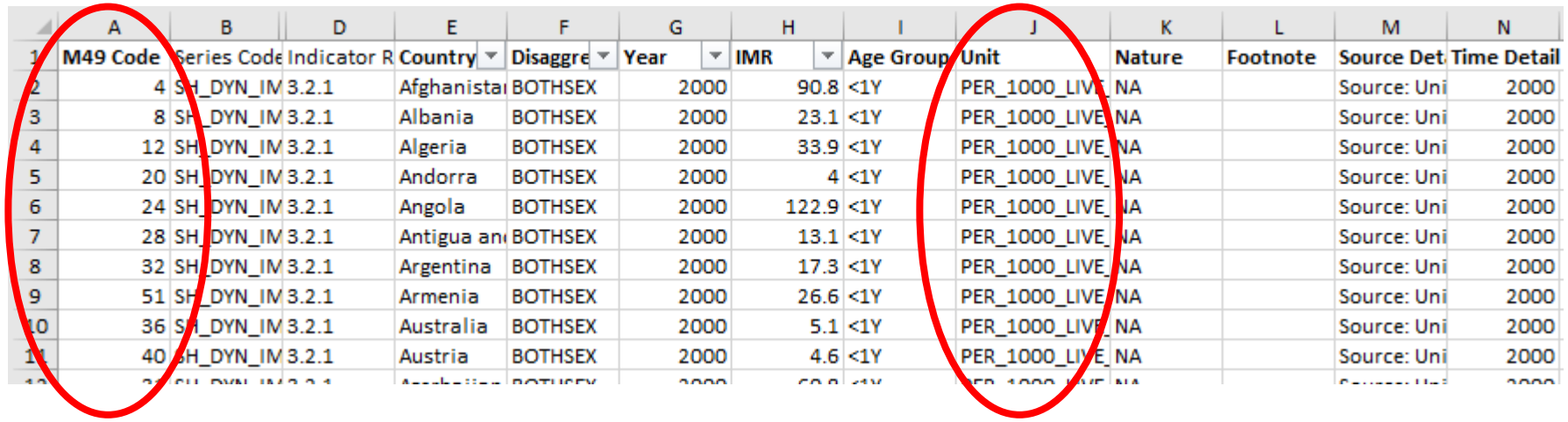
Mappings type: COLUMN

- Values for the concept are stored in the column specified in column **Position**

	C	D	E	F	G	H	I
1	**1.1.1 Proportion of population living below the international poverty line (1.9 USD)**						
2	*(as a per cent to number of population)*						
3		SERIES:	SI_POV_DAY	AGE:	_T	UNIT	PT
4		URBANISATION:	_T	EDU.LEV	_T	UNIT M.	0
5	**Items**	**REF_AREA**	**2007**	**2008**	**2009**	**2010**	**2011**
6	**Kyrgyz Republic**	KG	**0.13**	**0.09**	**0.29**	**0.28**	**0.01**
7	Batken oblast	KG05	0.00	0.27	0.00	0.21	0.00
8	Djalal-Abad oblast	KG03	0.20	0.00	0.00	0.22	0.00
9	Ysyk-Kul oblast	KG02	0.05	0.34	0.58	0.11	0.00
10	Naryn oblast	KG04	0.13	0.63	0.03	3.46	0.31
11	Osh oblast	KG06	0.33	0.03	0.00	0.09	0.00

Mapping type: COLUMN (2)

- Also used with record-based representation (a.k.a. flat file), when each row contains one record or observation



	A	B	D	E	F	G	H	I	J	K	L	M	N
1	M49 Code	Series Code	Indicator R	Country	Disaggre	Year	IMR	Age Group	Unit	Nature	Footnote	Source Det	Time Detail
2	4	SH_DYN_IM	3.2.1	Afghanistar	BOTHSEX	2000	90.8	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
3	8	SH_DYN_IM	3.2.1	Albania	BOTHSEX	2000	23.1	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
4	12	SH_DYN_IM	3.2.1	Algeria	BOTHSEX	2000	33.9	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
5	20	SH_DYN_IM	3.2.1	Andorra	BOTHSEX	2000	4	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
6	24	SH_DYN_IM	3.2.1	Angola	BOTHSEX	2000	122.9	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
7	28	SH_DYN_IM	3.2.1	Antigua an	BOTHSEX	2000	13.1	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
8	32	SH_DYN_IM	3.2.1	Argentina	BOTHSEX	2000	17.3	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
9	51	SH_DYN_IM	3.2.1	Armenia	BOTHSEX	2000	26.6	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
10	36	SH_DYN_IM	3.2.1	Australia	BOTHSEX	2000	5.1	<1Y	PER_1000_LIVE	NA		Source: Uni	2000
11	40	SH_DYN_IM	3.2.1	Austria	BOTHSEX	2000	4.6	<1Y	PER_1000_LIVE	NA		Source: Uni	2000

Mapping type: FIX

- Fixed value for the entire dataset is stored in the column **Position** and does not appear in the data spreadsheet
  - E.g. if the data is always expected to be annual, frequency can be coded for the entire spreadsheet

	A	B	C	D
1	Element	Type	PosType	Position
2	FREQ	DIM	FIX	A

Mapping type: MIXED

- The concept value is conditional
- Can be used to provide a default value

	A	B	C	D	E	F	G
1	Element	Type	PosType	Position		DataStart	K12
2	FREQ	DIM	FIX	A		NumColumns	1
3	REPORTING_TYPE	DIM	CELL	B4			
4	SERIES	DIM	COLUMN	2			
5	REF_AREA	DIM	MIXED	CELL	B3	FIX	TH
6	TIME_PERIOD	DIM	COLUMN	4			

- “Use cell B3 for concept REF_AREA. If the cell is empty, use fixed value TH”

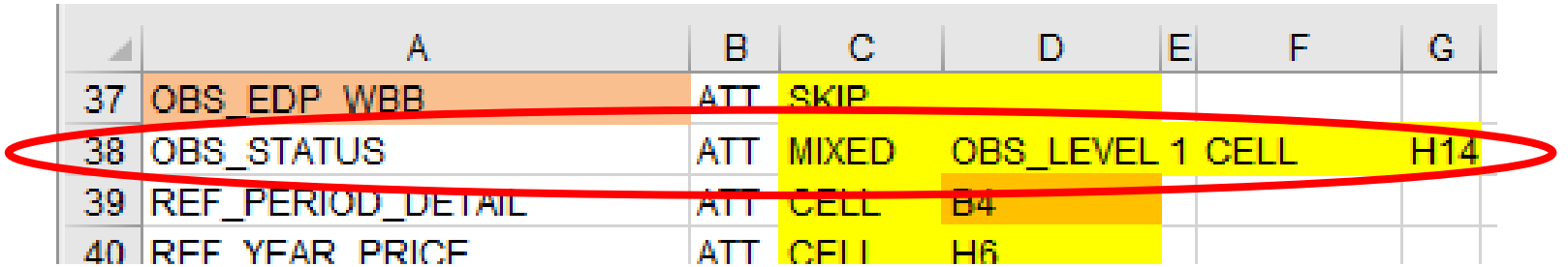
Mapping type: SKIP

- The concept value is not mapped
- Can only be used with optional attributes

18	NATURE	ATT	FIX	C
19	SOURCE_DETAIL	ATT	SKIP	
20	COMMENT_OBS	ATT	SKIP	

Mapping type: OBS_LEVEL

- Can be used in to specify attributes attached at the observation level relative to the cell containing the observation.
  - E.g. when each row has multiple observations *and their attributes*.



	A	B	C	D	E	F	G
37	OBS_EDP_WBB	ATT	SKIP				
38	OBS_STATUS	ATT	MIXED	OBS_LEVEL 1 CELL			H14
39	REF_PERIOD_DETAIL	ATT	CELL	B4			
40	REF_YEAR_PRICE	ATT	CELL	H6			

- “For attribute OBS_STATUS, use cell that is 1 column to the right of the cell containing the observation value. If that cell is empty, use the value in cell H14.”

Transcoding

- Transcoding refers to code mapping, when internal codes are different from DSD codes.

Internal Age Group		SDG DSD Age Group
15-19	→	Y15T19
20-24	→	Y20T24
25-29	→	Y25T29

- For conversion to work, internal codes need to be replaced with DSD codes.
- Transcoding can also be stored in external files. This is very useful when multiple sheets need to be mapped and reused.

Exercise 6: Mapping an Excel file

- Map a spreadsheet to the SDG DSD and use SDMX Converter to retrieve and convert the data

Exercise 7: Mapping country indicators

- Map your global indicators produced by your country to the global SDG DSD.

THANK YOU!