SDMX CENTRAL

HELP GUIDE FOR CONVERTING DATA FILES

Last updated: October 17, 2019

IMF SDMX CENTRAL

INTRODUCTION

This user guide aims to help IMF member countries to convert data in excel files into machine-readable (SDMX) format using SDMX Central.

One application of such conversion takes place when IMF member countries that participate in SDDS Plus, SDDS, and e-GDDS populate their National Summary Data Pages (NSDP) with SDMX files that allow for the automatic exchange and sharing of statistical data and metadata.

Before using this guide to disseminate SDMX data on the NSDP, country agencies are required to structure their data files according to a standard Data Structure Definition (DSD) such as the EcoFin DSD.

For questions about this, please contact SDMXCentral@imf.org.

SDMX CENTRAL

The IMF provides an online tool to help countries convert data into the SDMX format required for their NSDP.

Using the convert features on [SDMX Central](#) you can convert the validated Excel files into the required SDMX format.

This document describes the steps to follow to validate and convert your Excel data files into SDMX format.

VALIDATING AND CONVERTING DATA USING SDMX CENTRAL

TO VALIDATE AND CONVERT THE EXCEL DATA FILES TO SMDX:

Go to https://sdmxcentral.imf.org/ to open the web-based converter tool.

➢ Click **Convert** from the home page.

Note: You can also initiate the conversion process by click **Load Data** under the option *Data on the left-hand side of the screen*.

➢ The "Load Data" page will appear on the screen.
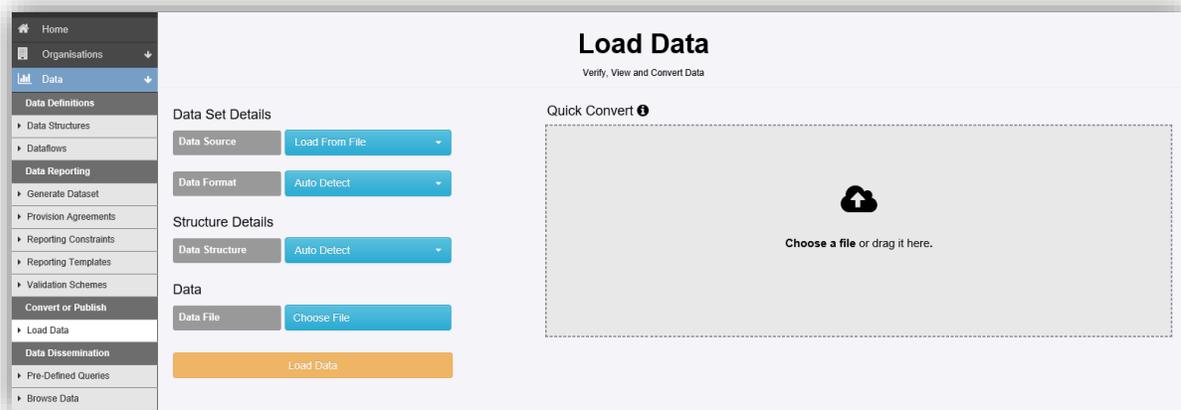
➢ You now have the following conversion options:

OPTION 1: QUICK CONVERT: DRAG-AND-DROP

Load Data

Verify, View and Convert Data

Quick Convert ⓘ

Choose a file or drag it here.
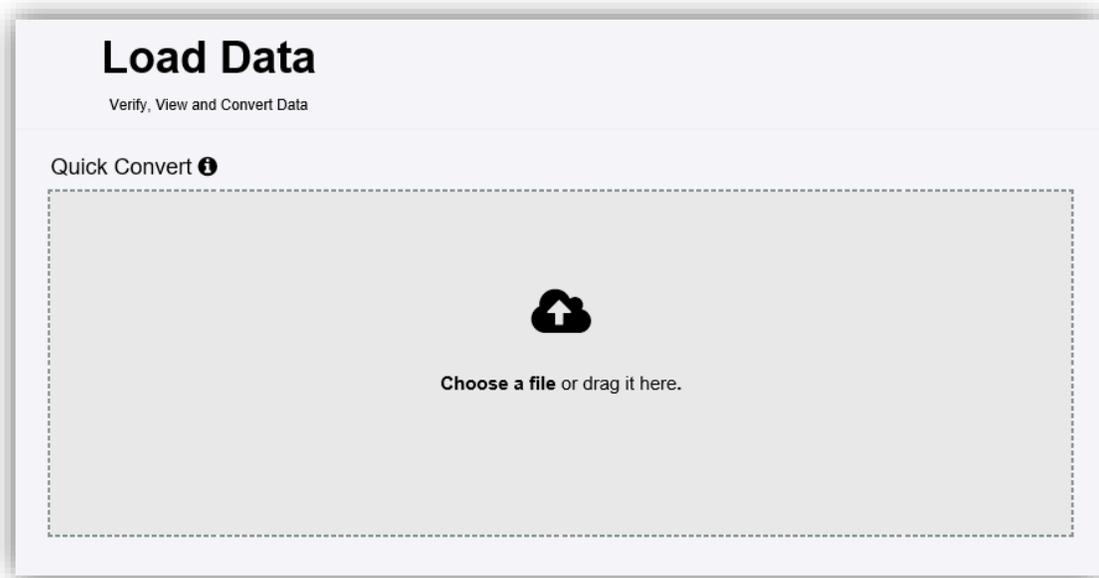
➢ Drag and drop an excel file from your local drive into the drop zone (the grey box shown in the image above). You can do this by having opened your computer browser to the folder where your file is stored.

The source file will be analyzed for conversion. If the data file contains information about the source Data Structure Definition (DSD) or Dataflow, SDMX Central will automatically convert the data into SDMX format and the SDMX file will be automatically be generated. A pop-up box will help you save the SDMX file on your local drive.

If the data file does not contain information about the DSD, or if it references an unknown DSD, additional pop-up boxes will ask for this information to be provided, as shown in the image below.

Quick Convert

Link Data Structure

Data file does not reference a Data Structure, please select one to use for validation

IMF:ECOFIN_DSD[1.0] ECOFIN Data Struct... ⌄

Cancel   Convert
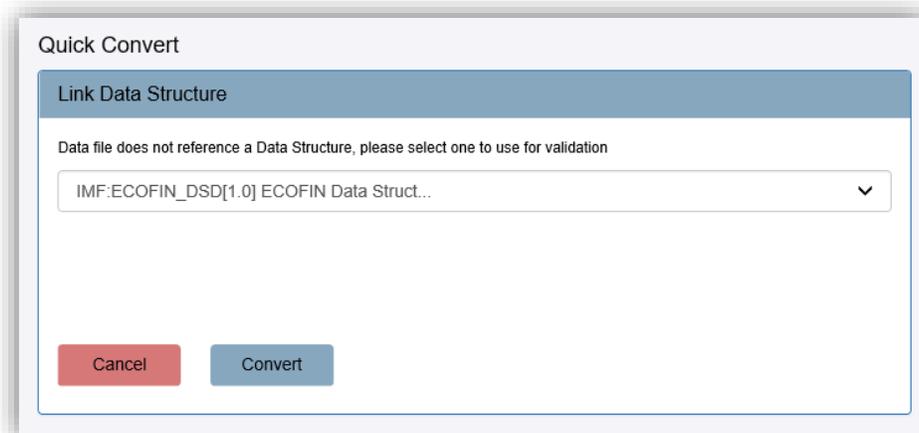
A summary pop-up box will surface, which highlights the outcomes of the conversion (see image below). In some cases, additional warnings and instructions will alert users of potential issues in the conversion.

> **NOTE:** An excel file can fail conversion and an SDMX file will not be generated if certain issues exists – see section on Conversion Categories.

SDMX Central provides a detailed report on the checks that were performed on the dataset. To review these checks in detail, please click on 'View Conversion Report', which will show a breakdown of test results as described in the Dataset Conversion Report section.
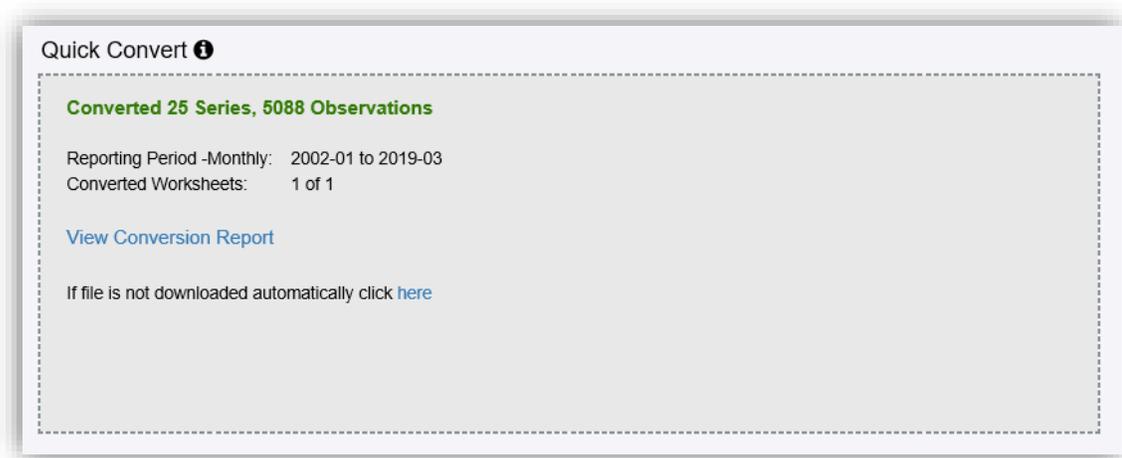
Quick Convert ⓘ

Converted 25 Series, 5088 Observations

Reporting Period -Monthly: 2002-01 to 2019-03
Converted Worksheets: 1 of 1

View Conversion Report

If file is not downloaded automatically click here

OPTION 2: LOAD FROM FILE

> ➤ Click on Choose File to browse and select the Excel data file that needs to be converted.

> ➤ Click on Load Data to submit the Excel data file for validation against EcoFin DSD.

Similar to the *Drag-and-Drop* feature above, the source file will be analyzed for conversion. If the data file contains information about the source Data Structure Definition (DSD) or Dataflow, SDMX Central will automatically navigate you to the **Conversion Report**. Additional pop-ups may appear if more information is needed.

> **NOTE:** Unlike the *Drag-and-Drop* feature, under this conversion option, an SDMX file is not automatically generated at this time. The option to generate the SDMX file is provided as part of the Conversion Report.

DATASET CONVERSION REPORT

The 'Dataset Details' reports, shown below, lists all relevant information acquired during conversion.

Particular attention should be paid to the test results at the bottom of the page.

➢ If the validation is successful, the system will display the results as below.

➢ If there is an error, the report, will display the results as below. Errors are limited to the first 60 per category.

NOTE: As long as the **Convert** Data button is solid blue, the dataset can be converted even if it triggers any of the errors listed above. Only specific errors prevent conversion – see Section on Conversion Categories.

➢ Click on **Convert Data** to generate the SDMX file.

➢ Click on **Download** to generate the SDMX file.

SAMPLE SDMX FILE

The SDMX file generated will look like the image below.

SDMX CENTAL CHECK: CONVERSION CATEGORIES

SDMX Central analyzes the following checks during conversion. As highlighted below, some categories require additional action prior to conversion (in red). For help on permitted excel format and additional tips see Section on SDMX Central Excel Data Formats.

Semantically Compliant: Ensures that the underlying dataset specification is correct. This is the only category which is specific to the data format.

For example, this highlights when dates in the worksheet columns are not in ascending chronological order.

This category does not prevent conversion. IMF SDMX Central will try to fix all semantic failures on conversion. Users are encouraged to fix issues identified in this category.

Duplicate Observations: Ensures the data file does not contain any duplications which contradict each other.

For example, if there are two observations with different values for the same indicator and period.

This category prevents conversion. Users are required to fix issues identified in this category.

Mandatory Attributes: Ensures all Attributes, as defined in the DSD, are reported if they are marked as Mandatory.

For example, datasets using the EcoFin DSD are required to define the scale [UNIT_MULT] in the Excel file.

This category prevents conversion. Users are required to fix issues identified in this category.

Time Period Format: Ensures all reported time periods formats conform to the reported Frequency.

For example, FREQ=A will expect time periods in format YYYY. The following table shows the expected date format for each frequency of data.

Frequency	Format	Example
Annual	YYYY	2010
Daily	YYYY-MM-DD	2010-01-01
Date Time	YYYY-MM-DD-Thh:mm:ss	2010-01T20:22:00
Monthly	YYYY-MM	2010-01
Quarterly	YYYY-Qn	2010-Q1
Semester	YYYY-Sn	2010-S1
Trimester	YYYY-Tn	2010-T1

This category does not prevent conversion. Users are encouraged to fix issues identified in this category.

However, for dataset with multiple frequencies, FREQ should always be the <u>highest</u> reported frequency.

This category prevents conversion. Users are required to fix issues identified in this category.

Missing Observations: Ensures that the series reports an observation value for the time period that lies within the time range for which the series is reporting data.

For example, if the first Observation value for a series is reported for 2001 and the last is reported for 2004, then this test that observations were reported for 2001, 2002, 2003, and 2004

This category does not prevent conversion. Users are encouraged to fix issues identified in this category.

Valid Calculations: If any additional rules exist to perform mathematical calculations to ensure compliance, then these tests are run, and any failures are reported under this category.

This category prevents conversion. Users are required to fix issues identified in this category.

Valid Constraint: A Dimension or Attribute in the dataset may report a value which is valid in the DSD but has been restricted for the reporting entity in the Reporting Constraints. These are additional checks which are performed when a Dataflow and/or Provision Agreement are selected.

For example, a registered data provider for FR (France), will not be allowed to report data for the UK (United Kingdom). Both FR and UK are valid countries in the Codelist but are limited based on their reporting constraints.

This category prevents conversion. Users are required to fix issues identified in this category.

Valid Representation: Ensures the reported values for Dimensions, Attributes, and Observation values exists within the DSD.

For example, if the Excel file reports using an INDICATOR that is not contained in this DSD those instances will be surfaced in this test.

This category does not prevent conversion. Users are encouraged to fix issues identified in this category.

SDMX CENTAL EXCEL DATA FORMAT

The format of a data file in Excel is expected to follow the following conventions:

#1 Dimensions, Attributes and Time Periods appear as header columns

	A	B	C	D	E	F	G	H	I
1	INDICATOR	FREQ	BASE_PER	UNIT_MULT	TIME_FORMAT	2001	2002	2003	2004
2									

Figure 1 demonstrating rule #1 – Dimensions, Attributes, and Time Periods as Column Headers

#2 If any Dimensions or Attributes have fixed values for the whole dataset, these may be placed in a Header section.

Shown below are fixed values for DATA_DOMAIN, REF_AREA and COUNTERPART_AREA

	A	B	C	D	E	F	G	H	I
1	DATA_DOMAIN	BOP6							
2	REF_AREA	JP							
3	COUNTERPART_AREA	W1							
4									
5	INDICATOR	FREQ	BASE_PER	UNIT_MULT	TIME_FORMAT	2001	2002	2003	2004
6									

Figure 2 demonstrating rule #2 providing fixed Dimension and Attribute values for the dataset

#3 Multiple Frequencies are supported in the same worksheet (Annual and Monthly for example). The frequency of the reported values is derived from the date formats.

Note: When using SDMX Central to derive the Frequency from the Time Period column, use the highest frequency reported in the FREQ Dimension.

1															
2	DATASTRUCTURE	IMF:ECOFIN_DSD(1.0)		Datastructure											
3	DATASTRUCTURE_NAME	ECOFIN Data Structure Definition		Datastructure name											
4	DATA_DOMAIN	CPI		Dataset											
5	REF_AREA	SB		Country											
6	COUNTERPART_AREA	_Z		Counterpart area							Annual				
7	UNIT_MULT	0		Scale = Unit						Monthly					
8	FREQ	M													
9	COMMENT	Published		Observation status											
10															
11	Country code	Descriptor		Indicator	Units	2017-09	2017-10	2017-11	2017-12	2018-01	2018-02	2001	2002	2003	2004
12		Consumer Price Index													
13	PCPI_IX	Consumer Price Index, All items		PCPI_IX	Index	589.9	590.3	590.1	594.6	604.9	617.5	568.6	584.1	585.6	608.8
14	PCPI_CP_01_IX	Food and non-alcoholic beverages		PCPI_CP_01_IX	Index	544.7	539.8	535.6	538.8	544.4	572.2	527.4	538.5	540.1	591.7
15	PCPI_CP_02_IX	Alcoholic Beverages, Tobacco, and Narcotics		PCPI_CP_02_IX	Index	952.2	963.5	986.1	1011.8	1134.6	1127.8	920.8	925.5	934.5	936.5
16	PCPI_CP_03_IX	Clothing and footwear		PCPI_CP_03_IX	Index	234.6	234.6	234.6	234.2	234.5	234.5	227.7	227.7	227.7	227.7
17	PCPI_CP_04_IX	Housing, Water, Electricity, Gas and Other Fuels		PCPI_CP_04_IX	Index	933.3	948.7	951.7	952.3	940.6	943.4	877.2	943.4	943.6	948.3
18	PCPI_CP_05_IX	Furnishings, household equipment and routine household	PCPI_CP_05_IX	Index	477	475.4	475.6	477.3	474.3	475.5	481.9	487	487.2	487.2

Figure 3 showing multiple frequencies being reported

When the data is read, the value for the dimension FREQ will be derived from the corresponding date format. The following rules are used:

FREQ	Frequency	Format	Example
A	Annual	YYYY	2010
M	Monthly	YYYY-MM	2010-01
D	Daily	YYYY-MM-DD	2010-01-01
I	Date Time	YYYY-MM-DD-Thh:mm:ss	2010-01T20:22:00
Q	Quarterly	YYYY-Qn	2010-Q1
S	Semester	YYYY-Sn	2010-S1
T	Trimester	YYYY-Tn	2010-T1

Table 1 showing derived FREQ code Id against the format of the date String in the date column

#4 It is permissible to have multiple worksheets with data. This mechanism can be used to report different frequencies of data per worksheet.

The Header section of each worksheet must be consistent in terms of layout, as shown in the image below.

Figure 4 showing worksheet 1 with Quarterly Data

Figure 5 showing worksheet 2 with Annual Data

#5 Observation Attributes may be reported in the header section, which applies this as a default value for all observations.

Figure 6 demonstrating rule #3 applying a default value for an observation attribute

#6 The header section should be separated from the data section by a blank row

Figure 7 a single blank row separates the header and data sections

#7 Reported values appear in the data section

Figure 8 demonstrating reported values for Dimensions, Attributes and Time Periods

#8 Dimension values are mandatory. If a value is not reported, this will result in an error.

Figure 9 omitting Dimension value FREQ will result in error

#9 Extra rows and columns in the spreadsheet may be added to improve the readability for the user. Blank rows are permitted but with certain restrictions. Blank rows may appear before the header section and between the header section and data section. However, a blank row may not exist within the header section. If a blank row is encountered in the header section, then this is assumed to indicate the end of the header section and this may cause your spreadsheet to be read incorrectly.

#11 Blank columns indicate that no further information should be read from that row. IMF SDMX Central will read from the first column of information in a row until it reaches a blank cell (unless it's a data row). The image below shows a spreadsheet where column H is blank and row 9 is blank. This would mean that in the data section only data for 2001 and 2002 are read (columns F and G). Columns I and J will not be read. However, all the 3 series (rows 7,8 and 10) will be read.

	A	B	C	D	E	F	G	H	I	J
1	DATA_DOMAIN	BOP6								
2	REF_AREA	JP								
3	COUNTERPART_AREA	W1								
4	OBS_STATUS	A								
5										
6	INDICATOR	FREQ	BASE_PER	UNIT_MULT	TIME_FORMAT	2001	2002		2003	2004
7	BCA_BP6_XDC	A		0	P1Y	12.2	22.2		32.2	42.2
8	BXCA_BP6_XDC					12.3			32.3	42.3
9										
10	BEFD_BP6_XDC	A		0	P1Y	12.4	22.4		32.4	42.4

Figure 10 demonstrating the effect blank rows and columns have on the processing of data

#12 It is permissible to have columns in the data section that are not dimensions, attributes or data but may contain additional information for the reader of the spreadsheet. The image below shows a spreadsheet where column F is for additional notes for each data row. The presence of this column will not prevent the data (in columns G to J) from being read even though the data rows themselves do not have a value for the row.

	A	B	C	D	E	F	G	H	I	J
1	DATA_DOMAIN	BOP6								
2	REF_AREA	JP								
3	COUNTERPART_AREA	W1								
4	OBS_STATUS	A								
5										
6	INDICATOR	FREQ	BASE_PER	UNIT_MULT	TIME_FORMAT	Additional Notes	2001	2002	2003	2004
7	BCA_BP6_XDC	A		0	P1Y		12.2	22.2	32.2	42.2
8	BXCA_BP6_XDC						12.3		32.3	42.3
9	BEFD_BP6_XDC	A		0	P1Y		12.4	22.4	32.4	42.4

Figure 11 demonstrating a column (column F) that is not data, dimension or attribute but will not prevent data processing

#13 It is permissible to have entire rows in the data section that are there to indicate what the data represents. This has the restriction that the text in these rows must not be in a column that indicates a Dimension, Attribute or Value. The image below shows row 7 being used to explain what rows 8 and 9 represent. The text for row 7 is in column A, which is now used as additional information.

	A	B	C	D	E	F	G	H	I	J	K
1	DATA_DOMAIN	BOP6									
2	REF_AREA	JP									
3	COUNTERPART_AREA	W1									
4	OBS_STATUS	A									
5											
6	Comment	INDICATOR	FREQ	BASE_PER	UNIT_MULT	TIME_FORMAT	Additional Notes	2001	2002	2003	2004
7	*Balance of Payments, Current Account, Total*										
8	Net, National Currency	BCA_BP6_XDC	A		0	P1Y		12.2	22.2	32.2	42.2
9	Credit, National Currency	BXCA_BP6_XDC									
10	*Balance of Payments, Exceptional financing*										
11	Direct investment, National Currency	BEFD_BP6_XDC	A		0	P1Y		12.4	22.4	32.4	42.4
12											
13											

Figure 12 demonstrating rows in the data area to aid readability

#14 It is permissible to have rows that do not report any observations. In the image above, row 10 reports no values (cells H10, I10, J10 and K10 are all blank).