

Statistical Data and Metadata eXchange (SDMX) for SDGs

Abdulla Gozalov (UNSD) & Lucy Gwilliam (ONS)

Abstract

[SDMX](#) is a standard used for the exchange of standardised data and metadata, which is sponsored by several international organisations. To facilitate development of SDMX for SDG indicators, the [Working Group on SDMX for SDG Indicators](#) was set up in April 2016. The working group released the official Global SDG DSD in June 2019 and production data exchange has started. There are two dataflows that are used by countries and SDG Custodian Agencies to submit SDG indicator data to the SDG Global Platform, using the Global SDG DSD. Countries can also create national SDG DSDs and set up National Dataflows with their complete datasets that include national indicators that are not part of the global framework. For example, Kyrgyzstan have set up a National Dataflow in order to disseminate SDG data on their Open SDG platform. There are a number of free tools and platforms available to support countries with the implementation of SDMX.

Contents

1. Introduction to SDMX
2. SDMX-SDGs Working Group
3. SDG Global Platform
4. SDMX dataflows for SDG reporting
5. Global and National DSDs
6. Tools and platforms used for SDMX-SDG reporting
7. Kyrgyzstan case study

1. Introduction to SDMX

SDMX is an initiative sponsored by seven international organizations:

- Bank for International Settlements (BIS)
- European Central Bank (ECB)
- Statistical Office of the European Union (Eurostat)
- International Monetary Fund (IMF)
- Organisation for Economic Co-operation and Development (OECD)
- United Nations (UN)
- World Bank

SDMX aims to standardise data and metadata exchange between international organisations and member countries.

SDMX provides standards for:

- structuring of statistical data
- packaging of statistical data as XML, JSON, CSV, and other formats

- registry of data and metadata

There are many specifications, guidelines, tools and manuals that are freely provided to support the implementation of SDMX

SMDX artefacts

Data Structure Definitions (DSDs) and Metadata Structure Definitions (MSDs)

DSDs and MSDs are SDMX artefacts developed and used to model data for statistical subject-matter domains, such as National Accounts, Labour or SDGs.

DSDs and MSDs describe the structure of related datasets and reference metadata through the use of concepts and codelists.

Data and metadata are structured according to SDMX artefacts such as DSDs or MSDs, and disseminated or exchanged as SDMX.

2. SDMX-SDGs Working Group

The Working Group on SDMX for SDG Indicators was established by the IAEG-SDGs in April 2016 to facilitate the development of SDMX for SDG indicators.

The working group is made up of:

- 12 countries: Brazil, Cape Verde, Cambodia, Cameroon, Colombia, Egypt, France (Chair), Mexico, Russian Federation, United States and two vacancies
- 10 international agencies: AfDB, Eurostat, FAO, ILO, OECD, Paris21, UNEP, UNICEF, World Bank, WHO

UNSD is the secretariat of the working group.

The first working group meeting took place in October 2016, and since then there have been virtual monthly meetings, as well as annual in-person meetings.

Objectives

The SDMX-SDGs working group is responsible for developing global DSD(s) and MSD(s) for SDG indicators. The DSD(s) and MSD(s) are made up of standards for the packaging of SDGs data and metadata including concepts, code lists, and structures.

The working group is also responsible for developing, piloting and establishing data exchange mechanisms for SDG indicators so that SDG data and metadata are disseminated and reported using SDMX at the national, regional and global level.

Progress

- Draft DSD created in February 2018
- Pilot data exchange took place between April and September 2018
- Official DSD was released and production data exchange started in June 2019
- Pilot metadata exchange will take place through October 2020
- Official MSD to be released in December 2020

3. SDG Global Platform

The SDG Global Platform consists of the Global SDG Indicators Database and the SDG Lab.

Global SDG Indicators Database

The Global SDG Indicators Database is maintained and regularly updated by the Statistics Division. It contains data on the global SDG indicators and is available to everyone. It includes both country-level data and regional and global aggregates. The data in the global database is from custodian agencies, compiled using harmonised methodology that makes the figures internationally comparable.

SDG Lab

The SDG Lab provides data exchange and comparison functionality for countries and custodian agencies. It contains SDG indicator data from the Global SDG Indicators Database as well as national figures submitted by countries. It's an online platform established by UNSD as part of the UNSD-DFID Project on SDG Monitoring to:

- I. Allow the exchange of SDG indicator data, and soon metadata, by countries and SDG custodian agencies with UNSD.
- II. Allow the comparison of national data provided by countries against the global data provided by the custodian agencies.
- III. Gradually replace all of other forms of data submission to the UNSD SDG Global Platform.

4. SDMX dataflows for SDG reporting

There are two different SDMX dataflows used for reporting data on SDG Lab:

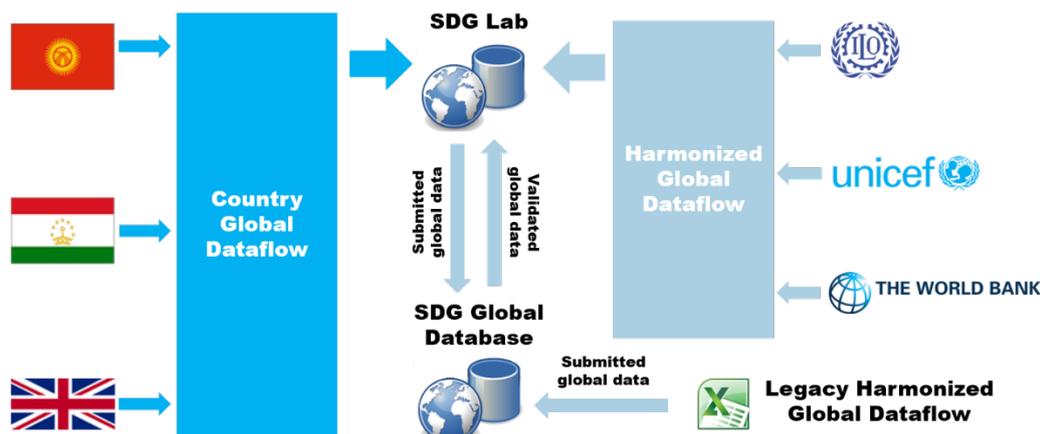
- Harmonized Global Dataflow – this dataflow is used by SDG Custodian Agencies to submit their data in SDMX format to the SDG Lab. Currently only a few Custodian Agencies are using this dataflow. The majority still submit their data in Excel.
- Country Global Dataflow – this dataflow is used by countries to submit their data in SDMX format. When a country submits any data to the SDG Lab, the data is automatically matched with the global dataset for comparison and assessment. Any differences between the observations should be explained in the metadata.

The custodian agencies can also submit data to the SDG Global Database:

- Legacy Harmonized Global Dataflow – this dataflow is used by SDG Custodian Agencies to submit harmonised global data to the SDG Global Database. This dataflow is currently used by the majority of Custodian Agencies to submit their data in Excel files as SDMX exchange is gradually implemented.

There is also a connection between the SDG Lab and SDG Global Database:

- Any global data that is submitted to the SDG Lab by the Custodian Agencies is passed to the SDG Global Database for validation
- Any global data that is submitted to the SDG Global Database and is successfully validated against the Global SDG DSD is passed to the SDG Lab



5. Global and National DSDs

Global SDG DSD:

The Global SDG DSD can be used by both countries and SDG Custodian Agencies to submit their data to the SDG Lab through the Country Global Dataflow and Harmonized Global Dataflow, respectively. This DSD supports the official global SDG indicators and their disaggregation.

National SDG DSDs:

Countries can create national customized DSDs and establish national dataflows, by extending the global DSD with their own indicators, disaggregation, and codes e.g. subnational reference areas or custom age groups, as well as national disaggregation. However, National SDG DSDs cannot be used to report to the SDG Lab. They are used to disseminate the full national dataset in SDMX format, as well as in support of dissemination platforms such as Open SDG.

6. Tools and platforms used for SDMX-SDG reporting

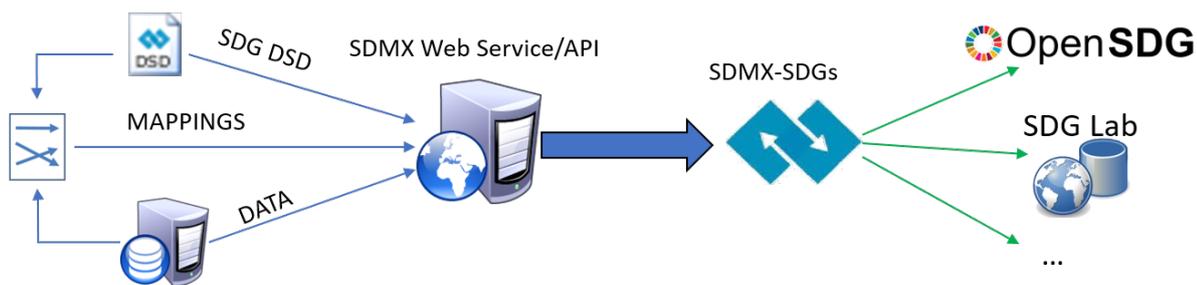
There are a range of [tools that can be used to support SDMX reporting](#) in various infrastructure environments. Some of the commonly used free tools include:

- [Eurostat's SDMX Converter](#) and [ILO SMART](#), which enable mapping data in Excel, CSV, or similar formats to a DSD and its conversion to SDMX
- [Eurostat's SDMX Reference Infrastructure](#), which enables mapping a database to a DSD and subsequently automatically establishing an SDMX API
- [ILO DSD Constructor](#) and [OECD SDMX Matrix Generator](#), which facilitate DSD maintenance

When implementing SDMX, most effort goes into mapping data to DSD. However due to the number and variety of tools available, no software development is required in either of the following scenarios.

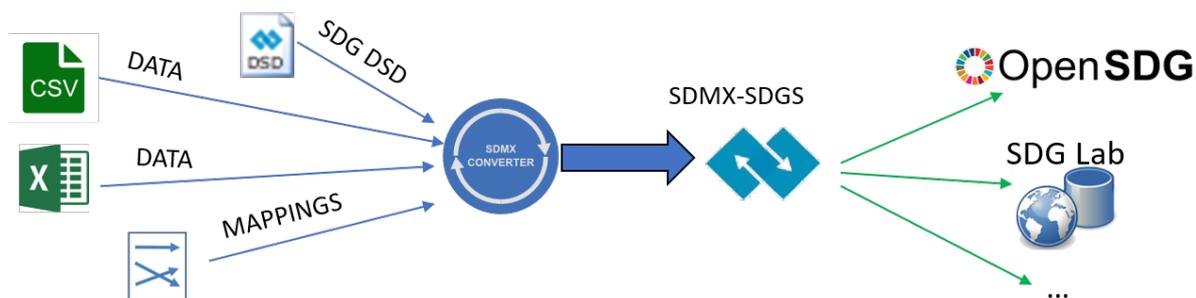
Scenario: existing database

Implementing SDMX is more efficient if an existing database is available. Software such as **SDMX Reference Infrastructure** can be used to map the database to the DSD. Once the mappings have been created, the full SDMX API is available.



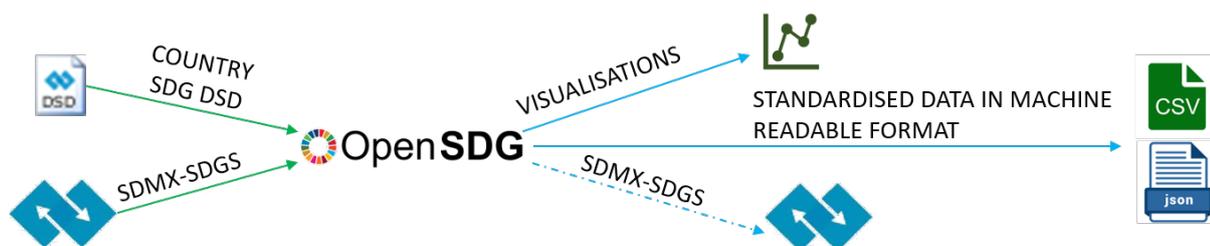
Scenario: no existing database

If there is no existing database, data files need to be mapped to the SDG DSD. Once mapped, tools such as **SDMX Converter** can be used to transform the data to SDMX files.



Open SDG

[Open SDG](#) is an open-source, free-to-use SDG reporting platform, which currently can use SDMX as input for visualization. This SDMX input functionality means that Open SDG users can upload standardised data, and in the future, metadata to their platforms. Functionality which will allow SDMX files to be downloaded will soon be added, which will simplify data exchange and enable users to download SDMX directly from the platform.



7. Kyrgyzstan case study

Kyrgyzstan began implementing SDMX for their SDGs in May 2019. The Kyrgyzstan SDG team have received in-country and remote support for implementing SDMX and using Open SDG.

To aid their implementation of SDMX, the Kyrgyzstan team use:

- OECD's SDMX Matrix Generator to generate their national DSD
- Eurostat's SDMX converter to convert Excel files into SDMX-ML files

They now have two data flows set up:

- National dataflow to report full national dataset, structured in accordance with the national DSD, on [Kyrgyzstan's Open SDG platform](#)

- Country global dataflow to report data that complies with global DSD on SDG Lab

