Implementing Statistical Data and Metadata eXchange (SDMX)

Nazira Kerimalieva & Lucy Gwilliam

Abstract

Kyrgyzstan is one of the pilot countries of the UNSD-DFID Project on SDG Monitoring. The main objectives of the project include building national reporting platforms for SDG indicators in project countries. This is done by supporting implementation of new tools and formats of data and metadata exchange such as Statistical Data and Metadata Exchange (SDMX). Kyrgyzstan decided to implement SDMX for SDG indicators.

Within the project, the capacity in SDMX of Kyrgyzstan's SDG team was been increased. As a result, today Kyrgyzstan has 90 global and 47 national SDG indicators that have been converted into SDMX and disseminated on their Open SDG platform. The Kyrgyzstan team plan to use this new capacity in SDMX for expanding implementation SDMX to other statistical domains throughout the National Statistical Committee of the Kyrgyz Republic (NSC).

The team uses tools such as Eurostat's SDMX Converter and OECD's SDMX Matrix Generator to make maintaining data in SDMX format easier.

Even though there have been some challenges, the team is now confident is making changes to their Data Structure Definitions and creating SDMX files and want to share their experience so others can learn from it.

Contents

1. Background

SDMX is a standard sponsored by seven international organizations. It was endorsed by the United Nations Statistical Commission in 2008 as a preferred standard for data exchange and has been approved as an ISO standard.

The goal of SDMX is to create common standards for the exchange of statistical data and metadata. This allows those in the field of statistical information to gain efficiency and avoid duplication of effort in their work. The Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs) established the SDMX-SDGs Working Group to facilitate the development of SDMX-based data and metadata exchange formats for SDG indicators.

The UNSD-DFID Project on SDG Monitoring is expected to make indicators more accessible through national reporting platforms (NRPs), and supports 20 pilot countries. Part of the Project's National Platforms Strategy Proposal in November 2018 is to support all project countries being able to provide data in SDMX format.

2.   Why Kyrgyzstan decided to use SDMX

Kyrgyzstan is a UNSD-DFID project country so the SDG team needed to implement SDMX as part of the project. However, there are a number of other reasons why the team decided to implement SDMX for their SDG indicators.

Comparable data

Kyrgyzstan now has data for global and national indicators in SDMX format, which follows the global SDG DSD. This means their SDG data is easily comparable with other countries' SDG data, where it also follows the global SDG DSD. When comparing data, each country's metadata should still be taken into account to ensure that the data is comparable.

Reduce reporting burden

Now that the SDG data is in the SDMX format, no further conversion is required. It is uploaded to the Open SDG platform and other platforms. As SDMX supports multi-lingualism, it is easier to maintain interfaces in multiple languages. For example, Kyrgyzstan's Open SDG platform is currently available in English and Russian languages; support for the Kyrgyz language is forthcoming.

In the future, when Kyrgyzstan start using an SDMX-API endpoint for their data, they will no longer need to send their SDG data to each organization that requires it, individually. Instead, all of the organisations that require Kyrgyzstan's SDG data can get it from the API, and organizations which have the functionality can even use the API to automatically feed data into applications e.g UN Global SDG platform.

Transparency

The uniform DSD structure allows users to read and correctly interpret SDG data from any source: national, regional or global.

Implementing SDMX throughout NSC

As SDMX is a great format for data and metadata exchange, next Kyrgyzstan is planning to implement SDMX in other areas of statistics. It will help the NSC to create uniform data models, easily integrate data to different dissemination tools and use new technologies for data visualization.

3. Using SDMX

SDMX Version

The IAEG-SDGs SDMX-SDGs Working Group are currently maintaining SDMX artefacts, such as the global SDG DSD, using SDMX v2.1. As Kyrgyzstan are adopting the global artefacts and customising them for national use, they are also using SDMX v2.1.

Tools

- OECD SDMX Matrix Generator is used to edit/update Kyrgyzstan's SDG DSD:
  1. Simple Excel sheet includes harmonised concepts and attributes easy to map, reuse in systems
  2. Matrix Generator creates new DSD structure with extended code lists.
- Eurostat's SDMX Converter is used to create SDMX-ML files:
  1. SDG Data in an Excel file is mapped to the SDG DSD. The mappings are saved as a worksheet in the file.
  2. The Excel file is processed with the SDMX Converter and converted to SDMX.
  3. The SDMX-ML file is outputted and validated against the DSD, with any errors listed

SDMX dataflows

Currently Kyrgyzstan has two SDMX dataflows for SDGs set up:
1. National dataflow, which uses the national customised DSD and includes the full set of their SDG data, and is used to populate the Open SDG platform
2. Global dataflow, which uses the unmodified global SDG DSD and allows Kyrgyzstan's national SDG data to be transferred to the SDG Global Platform and compared to the global SDG dataset.
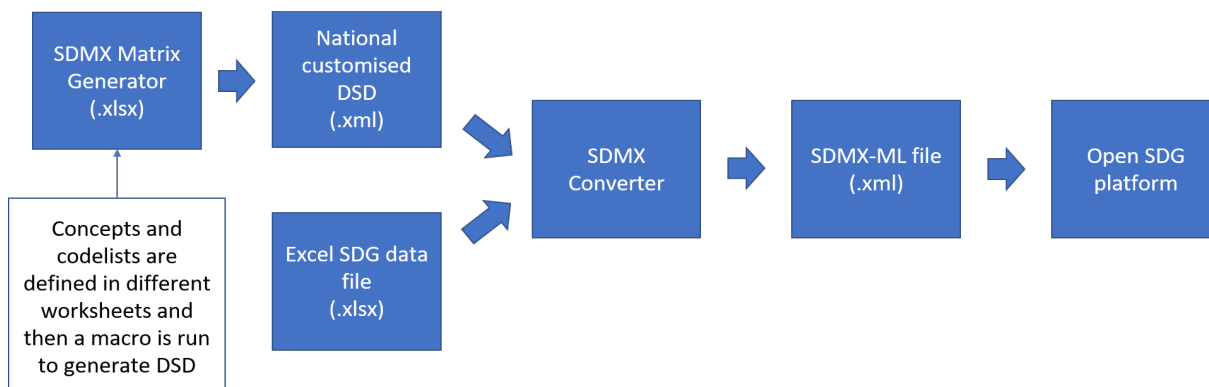

Figure: Diagram showing Kyrgyzstan's Open SDG platform use of the national SDMX dataflow. Diagram by author.
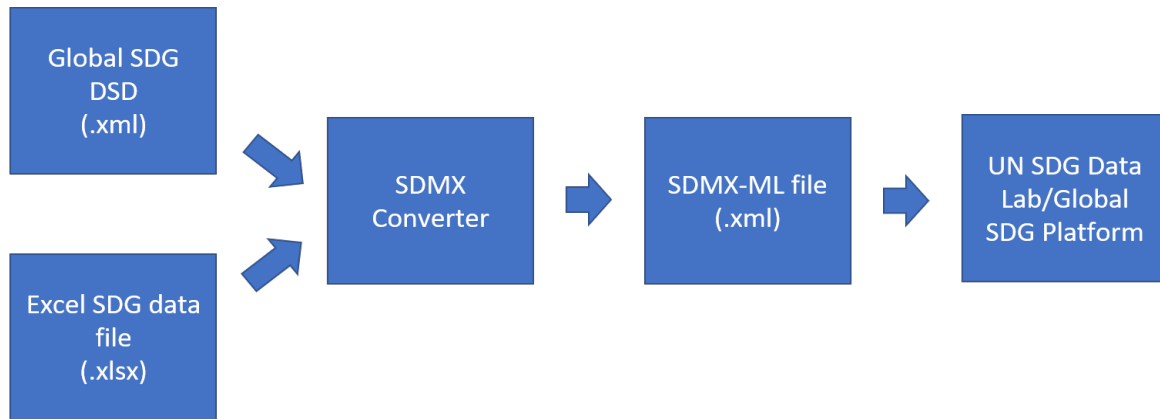
Figure: Diagram showing Kyrgyzstan's UN SDG Global Database using the global SDMX dataflow. Diagram by author.

4.   Implementing SDMX

Workshops

UNSD/UK SDG team joint mission
During May 2019, UNSD and the UK SDG team delivered a joint workshop to a number of participants. The purpose of UNSD's part of the workshop was to introduce the team to SDMX, including how to maintain a DSD and how to create SDMX-ML files from Excel files. The purpose of the UK SDG team's part of the workshop was to introduce the team to their Open SDG platform, including how to upload their standardized data.

UNSD workshop
In October 2019 and January 2020, UNSD delivered their second and third SDMX workshop, where the focus was on fully customizing the global DSD to meet national needs and converting all available global and national SDG indicators to SDMX format (90 global indicators and 57 national indicators).

UK SDG team workshop
In February 2020, the UK SDG team delivered their second Open SDG workshop, where they covered how to change the DSD that their Open SDG platform uses and how to use the new SDMX-ML input functionality. Following on from this, the Kyrgyzstan team has provided feedback to the Open SDG team of how the feature can be improved. In additionally, within the mission all available global and national indicators in SDMX were uploaded to Open SDG Platform.

Customising SDG Global DSD

Since the 2030 Agenda for Sustainable Development is rooted in national ownership, and the SDG indicator framework needs to be adapted to the national context, it is not possible for all countries to use the same DSD. Therefore, the global DSD had to be customised for national use. Some of the changes made include:
● adding national indicators e.g. *1.1.1.1a Extreme poverty rate* (in total 57 national indicators)

- adding disaggregations that aren't in global DSD e.g. regions of the Kyrgystan (Batken, Bishkek (city), Chui, Issyk-Kul, Jalal-Abad, Naryn, Osh, Osh (city), Talas)
- adding breakdowns that aren't in global DSD e.g. *17 to 20 years old* and *17 to 24 years old* under the *Age* disaggregation.

5.   Resources required/used

Team

The team who work on implementing SDMX include:
- Data specialists – supporting DSD structure, data converting.
- Statistical specialists – data collecting, data mapping, data converting.

Support

As mentioned above, support is provided by two organisations:
- UNSD supports Kyrgyzstan in maintaining DSD and converting the data to SDMX.
- The UK SDG team supports Kyrgyzstan in maintaining their Open SDG platform, including using SDMX with Open SDG.

Software and tools

Technical equipment/tools used to implement SDMX are:
- Microsoft Excel - Microsoft Excel has been needed since the original data is in Excel format, and the SDMX Matrix Generator is an Excel-based tool, which uses Excel macros to create a DSD XML file.
- Shared network drive - having a shared network, where all of the data files are stored helps to ensure that everyone is working from the same file, rather than multiple versions of the same file being worked on at the same time.
- Eurostat's SDMX Converter to convert mapped Excel files to SDMX-ML.
- OECD SDMX Matrix Generator to maintain Kyrgyzstan's national SDG DSD.

6.   Challenges

There were a few challenges when implementing SDMX. These include:
- SDMX is a technical subject with a lot of different concepts and tools.
- Highly disaggregated SDG dataset mandated by the UN General Assembly.
- Adapting to the national context means a global DSD cannot be used as is, and needs to be customised for each country.
- Creating national DSD structure requires a systematic approach.

7.   Lessons learnt

Lessons learnt while implementing SDMX include:

- Both technical and statistical experts need to be involved in the process.
- Requires a human capacity with understanding data structure.
- initially requires regular external support (can be remote support) but sustainable capacity can be created
- While implementation of SDMX requires an initial investment into training, this investment pays off with the benefits SDMX brings.
- The range of tools available in support of the implementation of SDMX makes it adaptable to various scenarios and infrastructure environments in statistical offices.

8.   Next steps

Following on from the work the Kyrgyzstan SDG team has already done to implement SDMX, below are some next steps.

SDMX-API

The NSC team is planning to start using an SDMX API using a platform such as the .Stat Suite or another. As mentioned above, this will allow any organization that requires Kyrgyzstan's SDG data to access it, and even allow the data to be pulled into some systems without any manual action required. In addition, a back end system such as .Stat simplifies data and metadata management.

Output of SDMX files from Open SDG platform

In the near future, the Open SDG team will be adding functionality to allow SDMX data to be outputted from platforms. Once this functionality is available, it will be added to Kyrgyzstan platform so that users can access Kyrgyzstan SDG data in SDMX format directly from the Open SDG Platform.