

Workshop on Statistics for SDGs

29 - 30 March 2022, Geneva

Session 1: Data transmission including SDMX and automation

**UK Experience of Statistical Data and Metadata eXchange (SDMX) Sustainable
Development Goals (SDGs) Reporting**

Prepared by Otis Bath, Office for National Statistics (ONS)

Abstract

This paper reviews and provides an update on the progress made by the UK in disseminating data in the Statistical Data and Metadata eXchange (hereafter SDMX) standard over the previous year. The UK SDG reporting platform, built using Open SDG, is outputting indicators in the SDMX standard. The data and metadata for some of these indicators has been uploaded to the SDGs Data Lab. Technical detail will be provided on each of the steps taken to achieve this and, where applicable, guidance will be made available. Whilst notable progress has been made, there remains several avenues for further development which will be outlined. Finally, the paper concludes with a discussion of the successes and challenges that were faced and resolved regarding SDMX.

I. OVERVIEW

1. The work on SDMX within the UK has aimed to create a standardised dataset with accompanying metadata that can be uploaded to the SDGs Data Lab. Once this has been completed, our data and metadata can be compared with the custodian agencies' data to evaluate any discrepancies.
2. The UK platform is currently outputting more than 20 indicators' data in the SDMX format with the uploading of this data and accompanying metadata to the SDGs Data Lab underway. This has come as the result of months of preparation and iterating the Open SDG platform, indicator data and metadata.
3. The process of uploading SDMX data and metadata to the SDGs Data Lab will continue until all the current output has been published. Once this has been completed, the longlist of indicators that could not previously be constrained to the SDG Global Data Structure Definition¹ (hereafter DSD) will be revisited and evaluated.
4. The UK has also been an active member of and contributor to the IAEG-SDGs working group over the previous year. The working group is responsible for maintaining the SDMX guidance and structures utilised by the UK including developing the DSD and Metadata Structure Definition (hereafter MSD) for SDG indicators.
5. There is also work underway to explore the potential to output indicators in a devolved format. Currently the data and metadata uploaded represents the UK in its entirety however

¹ <https://unstats.un.org/sdgs/iaeg-sdgs/sdmx-working-group/>

there is scope to disaggregate and report this for each administration. The devolved administration for the UK refers to the respective delegation of governance to England, Scotland, Wales and Northern Ireland.

II. STEPS TO OUTPUTTING SDMX DATA AND METADATA

6. The following sections detail the steps taken in the UK to publish data and metadata in the SDMX standard.

Proof of Concept

7. The first step was to formulate a proof of concept. Five indicators were identified which could be easily mapped to the DSD. This was completed to ensure the process was viable and worth pursuing further. The data for these indicators was modified to fit within the confines of the DSD and then converted to the SDMX format using an Open SDG test environment. These indicators were uploaded to the SDGs Data Lab in a provisional trial to ensure all access and functionality was working as intended.

Standardising disaggregations

8. This was followed by a comprehensive look at the terms and disaggregations used across the UK data reporting. We had to reach a consensus on the terms used throughout our indicators to standardise the fields being mapped to the DSD. The Open SDG disaggregation report identifies all disaggregations used on the platform so was invaluable in identifying non-standardised terms. For example, with age, the terms '0-15', 'less than 16' and '15 and under' are all equivalent. However, we need a consensus on their representation in the data otherwise a single mapping to the SDMX code 'Y0T15' in the DSD will not work.

Identifying compliant and non-compliant indicators

9. The UK SDG national reporting platform² publishes data for 208 of the 247 indicators listed in the global indicator framework. From this, a longlist of 62 indicators was created containing those which could plausibly be output in the SDMX standard. Indicators were excluded for their non-compliance if they did not follow the UN specification for the indicator, were being used as an approximation or were not present in the DSD.

10. A live monitoring document was used to track changes made to the longlist and to host discussions and capacity building for those making these indicators SDMX compliant. However, not all the indicators that were categorised in the initial longlist could be output in SDMX. For some, the series reported on matched the UN specification, however differences in the definitions of the disaggregations ensured these remained approximations. As such, the longlist of 62 indicators regressed to a potential shortlist of 40 to be explored further.

Meeting SDMX requirements

11. Once this shortlist of 40 indicators had been compiled, the next stage was to work collaboratively to make changes to these indicators to fit within the confines of the DSD. By using the DSD and the SDG Series Content Constraints Matrix³ (hereafter content constraints), changes were made to each of the indicators to ensure they had the necessary fields and coding

² <https://sdgdata.gov.uk/>

³ https://unstats.un.org/sdgs/files/SDG_Series_Content_Constraints.1.7.csv

within. The content constraints detail the criteria that are required for each indicator with the DSD stipulating the full list of concepts and codes that can be used across the SDG framework.

12. Whilst this was being completed for the indicators, each of the variables (for instance Units, Age, Sex etc.) and terms being input into the data (for instance Percentage %, 16 to 24, Female etc.) were recorded and coded based upon the DSD. This formed the basis of the code and column mapping documents, .csv files used by Open SDG to automatically map the fields shown on the platform to the code in the DSD. Indicators were iteratively changed to meet the SDMX requirements and, if their fields were recorded in the mapping files, then the platform could publish it in SDMX.

Open SDG SDMX output and validation

13. When the proceeding steps had been followed the initial result was the publication of seven indicators from the platform in the SDMX standard. Once some iterating had been completed this number increased to 21, a total which has been incrementally increasing since. Given the Open SDG platform checks for and raises errors before publishing, the need to validate the SDMX output using a converter is negated. The process of validating using an external converter can be time consuming, hence, this inbuilt function helped streamline the process of publishing data in SDMX.

14. Since the final objective is to compare the UK data with that of the custodian agencies on the SDGs Data Lab, any upload of SDMX data would be obsolete without the accompanying SDMX metadata. For this, UN Word metadata template forms⁴ were utilised, these are constrained documents which can be converted to the SDMX format using Open SDG allowing them to be uploaded to the SDGs Data Lab. This began with a trial of a single indicator and accompanying metadata document with the plan to gradually increase this output over the coming months.

III. WHAT WE HAVE LEARNT

15. There have been several successes and challenges over the past year that are important to reflect upon and share for other countries and organisations exploring SDMX-based data dissemination.

Open SDG's built-in validation

16. Once the substantial preparation had been completed, a number of indicators were published in the SDMX format. These indicators, by virtue of their publication, fit within the confines of the DSD and could be uploaded to the SDGs Data Lab since this validation is built into Open SDG.

UN metadata template forms

17. Using the UN Word metadata template form negates the need for a separate authoring tool. It ensures the metadata complies with the MSD and is comparable with the metadata provided by custodian agencies.

Column and code mappings

⁴ <https://github.com/sdmx-sdgs/metadata/#floppy-disk-download-the-latest-version-of-the-template>

18. The text used in the global DSD is not always what we wanted to show on the UK platform. Hence, the option with Open SDG to alter this text based on the column and code mappings or hide columns altogether makes SDMX reporting more flexible than would be possible otherwise.

Open SDG's constraints

19. The specific nature of the code and column mappings ensures that maintaining quality data is critical since typos can result in an indicator not being output in SDMX. Open SDG mitigates against this by identifying columns and rows being dropped when constraining to the DSD in the GitHub workflows, showing why it might not be published in SDMX.

Consistent disaggregations

20. The disaggregation report in Open SDG displays all the disaggregations present in the reporting platform allowing for easy merging of similar disaggregations as per the example in section 2.2 of this paper.

21. The Open SDG platform has helped to mitigate the challenges that have been faced and by continuing to push the agenda regarding SDMX, the UK continues to demonstrate its commitment to SDMX-based SDG data dissemination.
