Fourth Regional Workshop on the Use of Statistical Data and Metadata Exchange

In SDG Reporting – Advanced Level II

Beirut – Lebanon 27 June-1 July 2022

Summary

The Economic and Social Commission of Western Asia (ESCWA) in collaboration with the United Nations Statistics Division (UNSD), organized the Forth Regional Workshop on the Use of SDMX in SDG Reporting to improve countries' skills in mapping and converting SDG data and metadata, raise awareness on latest developments and promote the implementation of SDMX at the national level.

The workshop, a fourth in a series of ESCWA/UNSD joint workshops, aims to familiarize participants with techniques of mapping and converting SDG data as per international standards, respecting a content constraint for disaggregation adhering to official SDGs framework, using conversion tools including web-based ones, and process of customizing a national DSD. The workshop briefed participants on SDMX-RI (Reference Infrastructure) tool, DSD constructor and Matrix Generator.

The training encouraged interactive dialogue, and sharing of national experiences in SDMX including challenges, queries, and concerns.

1. Statistical Data and Metadata eXchange (SDMX) is designed to automate data and metadata exchange between two or more entities within the same entity. The process of normalizing data exchange has improved the efficiency of sharing both statistical data and metadata across statistical organizations and entities, providing an integrated approach and enabling the interoperable implementation of exchanging, reporting, and disseminating statistical data and metadata within and between systems.

2. The United Nations Economic and Social Commission for Western Asia (UN ESCWA) and the United Nations Statistical Division (UNSD) jointly organized the 4th Workshop on the Use of the Statistical Data and Metadata Exchange (SDMX)-Advanced Level II, from 27 June to 1 July 2022. The training is the fourth in a series of capacity-building workshops organized jointly by UNSD and ESCWA and was made accessible through the Zoom platform to enable all interested experts to follow the training remotely.

3. The workshop is in response to the resolution (A/RES/70/1) on the adoption of the 2030 Agenda of Sustainable Development in September 2015, to strengthen the capacity of national statistical offices and data systems to ensure access to high-quality, timely, reliable, and disaggregated data by income, sex, age, race, ethnicity, migration status, disability, and geographic location and other characteristics relevant in national contexts. It also responds to Member States' request to the 14th Statistical Committee for ESCWA to hold advanced training on SDMX emphasizing the importance of SDMX.

4. The objectives of the workshop are to improve countries' skills in mapping and converting SDG data, raise awareness of the latest developments and promote the implementation of SDMX at the national level.

5. The Workshop agenda covered the following items:
    - Review of SDMX Information Model
    - Review of SDMX converter basics, data, flows and content constraints.
    - Introduction to DSD constructor and Matrix Generator.
    - Advanced techniques for mapping and converting SDG data.
    - Customizing a national SDG DSD.
    - Introduction to SDMX Reference-Infrastructure.
    - SDG data transfer to ESCWA through SDMX.

6. The five-day workshop was attended by 15 participants (7 statisticians and 8 Information Technology and 7 females and 8 males). ESCWA also made the workshop available online to facilitate attendance by other SDG team members in national statistical offices. The selection criteria for participating in the specialized workshop were based on the

successful demonstration of understanding and implementation of correctly mapping and converting SDG data using the SDMX content constraints.

7. There were also 16 experts registered online 7 statisticians and 9 Information Technology and 11 females and 5 males). ESCWA engaged the online participants and shared the exercises during the workshop to enable participants to follow up closely with those participating in person. Participants who successfully completed the exercises during the workshop (online, in person) were awarded a certificate.
8. Discussion: main points

A quick review of the SDMX information model was made including SDMX main artefacts such as concepts, concept scheme, cross-domain concepts, Data Structure Definition, Data Flow, Content Constraints, and the representation of concepts (coded, uncoded/formatted, free text). It was noted that when a country is designing an SDG DSD for reporting and/or dissemination, it is highly recommended to check the cross-domain code lists and use standard codes for optimal interoperability. It was also noted the code list for FREQ dimensions in SDG DSD stores codes for non-annual frequencies, however, SDG data are annual in nature, and in SDMX 2.0 and 2.1 FREQ is a mandatory dimension.

With regard to REF_AREA, this is a standard dimension in the SDMX code list, and codes should not be removed; however, SDMX constraints can be used to only allow codes for ESCWA member States. ESCWA can create individual provisional agreements with countries, where a country will submit its data to ESCWA directly.

On data validation, it is recommended to use the latest DSD / data flow version to validate data exchanged since data validation in previous versions may not pass, because newer versions may include more or newer codes. It is worth noting that a dataset that passes validation against DSD doesn't mean it will pass validation against data flow because DSD validates all dimensions and mandatory attributes have valid values, and all optional attributes have valid values and position types, while validation against the data flow uses content constraints, which in addition validate relationships among dimensions.

With regard to attributes, some are optional and can be SKIPed, or just removed, however, mandatory attributes must always be provided. Countries were advised to consult the new standard classification for the cross-domain code list for the degree of urbanization and read the recommendation on methods of delineating cities, and urban and rural areas for international statistical comparisons. The recommendation is to follow the international standards for the data to be interoperable as much as possible. Deviations from the international standard which might be necessary for the country must be noted in footnotes and reference metadata to explain the differences with international standards to avoid misinterpretation.

The meeting explored also different mapping position types for mapping the data. Participants mapped and converted SDG data using DSD and data flow as a validation tool. Any non-SDG indicators are considered national indicators and countries need to create their own series codes and not use the global series which are meant only for global indicators

Participants faced challenges when it came to COMPOSITE_BREAKDOWN dimension UNIT_MEASURE, and UNIT_MULT attributes. It was pointed out that code 0 must be used where the observation value is in simple units and recommended to refer to the SDG series content constraint matrix to review valid UNIT_MEASURE, AND COMPOSITE_BREAKDOWN for each series transmitted.

The meeting also covered tools for customizing a national DSD, mainly DSD Constructor and Matrix Generator, it stressed the importance of taking ownership when changing or modifying any artefact. The versioning process includes, first, using the global structure, second, adding national codes to the code lists, third, updating the code list version and taking ownership, fourth, updating the structure version and taking ownership, and finally, publishing your customized structure.

It was also noted that there are certain standard codes that start with "_" and are known as "Generic Cross-Domain Codes", for example, _T is used for Total, _U is used for an unknown, _X for unspecified and _L is used for a local extension to global code lists. Since each country and in each survey the naming of the variable differs such as (DOB, Age, or even age range) it is important to assign the correct code when summarizing / aggregating / transforming the survey results regardless of how data was collected. The ILO DSD Constructor can be used to create and maintain non-SDG DSDs, such as for demographic data.

Deleting dimensions/mandatory attributes or using SKIP is not allowed, they must be mapped, in that case to a fixed value usually, _T, or 0 for the UNIT_MULT attribute (when applicable); optional attributes can be deleted if not in use or use SKIP.

Content constraints are usually attached to data flows, as in the case of the SDGs, nevertheless, they can be attached to DSD. The main reason why content constraints are often attached to data flow is because data flows often impose additional restrictions, and to ensure that data flows in the future stay harmonized to content constraints.

The meeting also presented advanced techniques to map and convert SDG data using SDMX, including transcoding sheets, and mapping multiple data sheets applying advanced techniques. It was noted that SDMX is a means, not the end; SDMX is used to standardize data, and DSD is designed to support the global indicator framework. It was also noted that the "Mapping Assistant" tool can be downloaded from the Eurostat website. However, the latest official release of mapping assistant has a bug, it does not work with

maintenance agencies that have a dash (-) such as example IAEG-SDG, therefore, it is recommended to download the 9.3.0 version or above.

ESCWA featured a demo of data exchange to the ESCWA website, using ESCWA SDMX Converter. The ESCWA SDMX generates the same errors and details attached as provided by the SDMX converter. ESCWA can add an option to send the log file to the users' email. ESCWA will provide a playground for countries to test their data before sending them to ESCWA. Moreover, the ESCWA SDMX converter allows users to connect to APIs.

9. The Workshop conclusions and recommendations are as follows:
    - All participants were requested to complete the SDG data conversion using the SDMX tools to update the National Reporting Platforms by September 2022.
    - To ensure the utmost benefit from the workshop some participants at the workshop, including those who did not attend in person and/or attended online, to complete additional exercises to continue their learning and participate in future workshops.
    - Countries' SDG teams should be formed of both statisticians and IT experts for the compilation and dissemination of SDG data through SDMX tools. Statisticians SDGs-experts and IT experts should work closely to validate and exchange with regional and international partners their national SDG data through SDMX.
    - ESCWA through collaboration with countries will set regional code lists based on a common list of indicators. ESCWA will also facilitate the development of additional data verification as a regional content constraint to improve the dissemination and accuracy of the SDG data from the region.

10. ESCWA also provided simultaneous interpretation in the Arabic and English languages during the period of the workshop. The workshop presentations and resources are available at the ESCWA website, accessible at the following link: https://www.unescwa.org/events/sdmx-training-series

11. The results of the feedback received from 19 participants (100% from 15 in person and 4 online) were as follows: 74% of the participants rated the workshop as excellent and 25 as very good, and 1% as good. In terms of workshop quality, 74% of the respondents found the workshop quality excellent and 26% gave a very good rating. Participants were asked if the workshop objectives were met, 79% of the participants gave an excellent rating and 21% gave a very good rating. Regarding presenters' inputs, 74% of respondents rated the presentations as excellent and 21% rated the presentations as very good, 5% rated presentations as good. Finally, 68% of the participants rated the logistics and organization of the workshop as excellent, and 32% as very good.

12. Further useful resources and links to past workshops:

- Guidelines for the Global DSD for SDGs
- Guidelines for the Customization of the Global DSD
- SDG Dataflows and Content Constraints
- SDMX SDG Page
- SDMX Global Registry
- SDG Data Matrix v 1.9
- Metadata Online Converter
- Tutorial & Materials
- SDMX Crash Course I
- SDMX Crash Course II
- SDMX Series of Trainings
- SDG Metadata Template
- SDMX Desktop Converter
- Europa SDMX Online Converter
- SDMX Tools
- SDMX Converter Error Messages

13. Group photo