United Nations E/CN.3/2023/29

Economic and Social Council

Distr.: General
12 December 2022

Original: English

Statistical Commission
Fifty-fourth session
28 February–3 March 2023
Item 5 (g) of the provisional agenda*
Items for information: common open standards for the exchange and sharing of data and metadata

Statistical Data and Metadata Exchange sponsors

Note by the Secretary-General

In accordance with Economic and Social Council decision 2022/324 and past practices, the Secretary-General has the honour to transmit the report of the Statistical Data and Metadata Exchange (SDMX) sponsors. The report contains a summary of achievements, recent developments and planned activities by the SDMX initiative, including the release of a new version of the standard, SDMX 3.0, the development of software tools and platforms in support of statistical data and metadata exchange and dissemination, and the implementation of data and metadata exchange in various subject matter areas. An overview of capacity-building activities in the area is also included. The Statistical Commission is invited to take note of the report.

* E/CN.3/2023/1.

22-28055 (E) 221222

Report of the Statistical Data and Metadata Exchange sponsors

I. Introduction

1. The aim of the Statistical Data and Metadata Exchange (SDMX) initiative, established in 2001, is to create and maintain technical and statistical standards and guidelines, together with an information technology architecture and information technology tools, to be used and implemented by the community of official statistics. Combined with modern information technology, the SDMX standards and guidelines should improve efficiency when managing statistical business processes. After 20 years of SDMX experience, they are now mature enough to be broadly implemented in various statistical domains and globally by statistical organizations.

2. The Statistical Commission has received SDMX progress reports since 2002. The Commission recognized and supported the SDMX standards and guidelines in 2008 as the preferred standard for the exchange and sharing of data and metadata, requesting the SDMX sponsors to continue their work and to encourage national and international statistical organizations to increase the use and implementation of SDMX.

II. New developments

3. A new version of the standard, SDMX 3.0, was released in September 2021. Following the previous release of SDMX 2.1, in 2011, the new version represents a major upgrade, with significant changes to the information model, data and metadata exchange protocols, and others. The following major new features have been introduced:

- Simplification and improvement of the reference metadata model
- Support for microdata
- Support for geospatial data
- Support for code list extension and discriminated union of code lists
- Improvements to structure mapping
- Improvements to code hierarchies for data discovery
- Improvements to constraints
- Improvements to versioning of structural metadata artefacts
- Improvements to the REST web services application programming interface
- Improvements and simplification of data and metadata exchange formats

4. The need for supporting microdata exchanges is highlighted in recommendation 14 of the new Group of 20 Data Gaps Initiative of September 2022. In particular, the Group recommends investigating the potential development of an international microdata standard. Such a standard would cover principles, enablers, use cases of best practice microdata-sharing (e.g. access for research purposes, cooperation between public and private sectors, open data initiatives and good practices from non-statistical examples of data-sharing) and measures of success.

5. In November 2022, an SDMX user forum was launched.[1] The forum, established in a collaborative effort of national statistical offices, the SDMX sponsors agencies and regional commissions of the United Nations, serves to connect users and experts, share information and provide support to the global SDMX user community. It significantly strengthens support options available to SDMX practitioners and should lead to faster resolution of any issues and improved implementation of data exchange and dissemination. The forum is part of the Statistics Division-led Global Network of Data Officers and Statisticians[2] and benefits from synergies with and connection to an active online statistical community of more than 2,500 members.

III. Ongoing statistical and technical work

6. The Working Group on Statistical Data and Metadata Exchange for Sustainable Development Goal Indicators was established by the Inter-Agency and Expert Group on Sustainable Development Goal Indicators with the mandate to develop a solution for the exchange and dissemination of Sustainable Development Goal indicator data and metadata. The Working Group released the first official data structure definition for Goal indicators in June 2019; the definition has since been regularly updated in synchronization with the Global Sustainable Development Goal Indicators Database. Data exchange has been established with about 36 countries and six custodian agencies, and the database is published in its entirety through the Statistics Division SDMX application programming interface. The first official metadata structure definition for Goal indicators was published in February 2022. Reference metadata exchange is facilitated by a set of tools developed by the Division in cooperation with the World Bank. The reference metadata set for the global Goal indicators is now published through an SDMX application programming interface. The reference metadata application programming interface facilitates the computer-assisted translation of Sustainable Development Goal reference metadata into a number of languages, as part of a World Bank project. The interface is also used by a beta version of the Sustainable Development Goal Metadata Database to facilitate and simplify the dissemination of global Goal metadata, and it has made it possible to link to and disseminate metadata along with data through the upgraded interface of the Global Sustainable Development Goal Indicators Database.

7. As part of the revision of the *International Merchandise Trade Statistics* manual of 2010 by the task team on international trade statistics of the Committee of Experts on Business and Trade Statistics, the topic of SDMX international merchandise trade statistics was added to the research agenda to reflect the expected new variables and take advantage of SDMX version 3.0, notably on microdata support. However, given the focus on the statistical framework in 2022, it was decided to reschedule discussions on the manual update to 2023.

8. With regard to the System of Environmental-Economic Accounting, as part of the work of the Committee of Experts on Environmental-Economic Accounting, the Statistics Division conducted a pilot data collection exercise of energy accounts with six countries using SDMX-enabled Excel questionnaires. The questionnaires were mapped to the global System of Environmental-Economic Accounting for Energy data structure definition, in cooperation with Eurostat. Data received from the countries were converted into an SDMX format and further validated and processed. The Division plans to launch a regular data collection of System of Environmental-Economic Accounting energy accounts using SDMX-enabled questionnaires in 2023. The data are planned to be disseminated through the Division's SDMX application

[1] See www.yammer.com/unstats/#/threads/inGroup?type=in_group&feedId=121500557312&view=all.

[2] https://unstats.un.org/capacity-development/global-network-of-data-officers-and-statisticians/.

programming interface and visualized through the Artificial Intelligence for Environment and Sustainability for System of Environmental-Economic Accounting Explorer.[3]

9. The International Monetary Fund (IMF) has continued to facilitate data dissemination using SDMX through the implementation of its Data Dissemination Standards initiatives. Over the past two years, IMF assisted 14 countries with the implementation of SDMX-enabled National Summary Data Pages – eight participants of the Enhanced General Data Dissemination System, one subscriber to the Special Data Dissemination Standard and five adherents to the Special Data Dissemination Standard Plus. As at the end of 2022, 106 countries will be disseminating data in SDMX format through their National Summary Data Pages.

10. The statistical data warehouse is the main dissemination channel of European Central Bank statistics; in 2021, the Bank worked on the revamp of the statistical data warehouse portal, the ECB data portal. The portal is metadata-driven, and an SDMX data model is the structure used to disseminate statistical data. As part of the portal, the Bank also launched its new search engine, which leverages ElasticSearch, one of the most modern and widely adopted open-source search engines. The new engine offers a novel way to search for statistical data by using the SDMX-modelled metadata.

11. With regard to education statistics, a project on the implementation of SDMX for the global exchange of education statistics has been finalized, and the resulting artefacts are actively used in data exchange. The project involves Eurostat, the Organisation for Economic Co-operation and Development (OECD) and the Institute for Statistics of the United Nations Educational, Scientific and Cultural Organization.

12. With regard to labour statistics, since 2017, the International Labour Organization (ILO) has been assisting countries with the implementation of Labour Market Information Systems. Its projects include the development of a series of capacity-building activities and the provision of a toolkit composed of open-source software tools based on SDMX, namely, .Stat Suite, SDMX Constructor and SMART. The piloting of a draft labour global data structure definition developed by OECD, ILO, Eurostat, the European Central Bank and the World Bank is ongoing at Eurostat and the European Central Bank.

13. Global SDMX data structure definitions in the area of macroeconomic statistics continue to be used for data exchange and to be maintained. The domains covered include national accounts, balances of payments, foreign direct investment, consumer price indices and environmental-economic accounts. A project for the development of a global data structure definition for residential property prices has also been launched under the coordination of the Bank for International Settlements.

14. The SDMX Technical Standards Working Group and the SDMX Statistical Working Group are responsible for the management and enhancement of the SDMX technical and statistical standards and the related SDMX information technology applications. The Technical Standards Working Group has primarily focused on the development of SDMX 3.0, until its successful launch in September 2021. Work has continued after the release to address a series of outstanding issues and develop new features that, for various reasons (e.g. late submission or unclear definition), were not part of the new version but are still considered of importance for the standard. In addition, the “federation of registries and top-level agency scheme” and the coordination of the development of “SDMX 3.0 reference implementation” are two very important items that have been also included in the current workplan.

[3] See https://seea.un.org/content/aries-for-seea.

15. The Bank for International Settlements, OECD and Eurostat have joined forces to improve the way SDMX software tools are delivered, coordinating workplans to avoid duplication of effort and integrating their joint portfolio of products to better address user needs. A key initiative is the provision to users of the OECD .Stat Suite of a way to maintain the tool's SDMX structures interactively by integrating the Bank's Fusion Metadata Registry SDMX metadata management product. Similar work is under way to explore how the Eurostat SDMX reference infrastructure could benefit from the SDMX 3.0 capabilities provided by the latest Fusion Metadata Registry (version 11). The reference infrastructure is used extensively by organizations in Europe and across the world for statistics publication and reporting and also forms an engine for .Stat Suite.

16. The Statistical Working Group has been working on a number of guidelines and improvements to the standard, including:

- A number of new cross-domain code lists and the revision of existing code lists
- Working with the Technical Standards Working Group on developing guidelines for SDMX 3.0 implementation
- Working on guidelines for SDMX structural metadata governance
- Working on guidelines to harmonize units of measure
- Applying an open-source licence to the SDMX content-oriented guidelines
- Working on a revision of the guidelines for SDMX modelling

IV. Capacity-building

17. Responding to growing demand for capacity-building, SDMX sponsor agencies have substantially increased their capacity-building activities. Owing to the constraints imposed by the coronavirus disease (COVID-19) pandemic in 2020 and 2021, many capacity-building activities were conducted online. In addition, more efforts are made to develop e-learning resources, which are made available to the global audience.

18. Eurostat has created regular courses on SDMX for both beginners and information technology developers. While the courses target first and foremost European Union member States, they are also open to participants from other countries. Within the framework of its international cooperation activities, Eurostat has also financed SDMX capacity-building activities in Eastern Europe. In addition, Eurostat maintains two portals: the SDMX InfoSpace and the Portal on Collaboration in Research and Methodology for Official Statistics (CROS). The former provides access to general information about the SDMX standard in three languages (English, French and German), while the latter includes detailed information on the SDMX tools developed by Eurostat, forthcoming training initiatives and offline training material, such as tutorials and explanatory videos. Information on past and ongoing SDMX implementations in the European Statistical System is also available.

19. The Statistics Division has carried out online and in-person capacity-building activities in such areas as the Sustainable Development Goal indicators and international merchandise trade statistics. As part of a project on Sustainable Development Goal monitoring funded by the Foreign, Commonwealth and Development Office of the United Kingdom of Great Britain and Northern Ireland, the Division conducted online training in SDMX for Goal indicators and provided technical assistance to a number of project countries in 2021 and 2022. Together with the Economic and Social Commission for Western Asia, the Division also ran a series of large-scale online training events in SDMX for Goal indicators for countries of the

Arab region, which were followed by an in-person event in June 2022. Similarly, the Division conducted a series of online workshops with the Asian Development Bank for the Thailand National Statistical Office, followed by an in-person event for the national statistical offices of Kyrgyzstan, Maldives and Thailand in November 2022. The Division also provided technical assistance to Cambodia, Kyrgyzstan and Maldives with establishing data warehouses based on the .Stat platform hosted on the United Nations Global Platform. The Asian Development Bank, in cooperation with the Division, the Economic and Social Commission for Asia and the Pacific (ESCAP) and the Statistical Institute for Asia and the Pacific, launched SDMX Foundation, an online course, which ran in April and May 2022 and attracted more than 500 participants from across the globe. Lastly, the Division has developed an e-learning course on SDMX for Sustainable Development Goal Indicators, to be launched in December 2022, which will provide an opportunity to the global audience for familiarizing itself with the Sustainable Development Goal data structure definition and data flows and will facilitate Goal reporting and dissemination. The Bank, the Division, the Statistical Institute for Asia and the Pacific and ESCAP are planning to start the development of an e-learning course on SDMX international merchandise trade statistics in 2023.

20. The European Central Bank held two SDMX training events in 2021 (with 160 participants), primarily targeted at the European System of Central Banks participants, followed by a training session held in October 2022. The training activities were conducted in a webinar format, owing to the COVID-19 pandemic, and covered an introduction to SDMX, the tools, and validation and transformation language, including the introduction of a real-life SDMX project at the Central Bank of Iceland in 2021. Overall, the webinars were very well received by the attendees. Similar events are planned for 2023.

21. The Bank for International Settlements has launched a new ongoing SDMX training and capacity-building programme open to all organizations and individuals working in the official statistics domain. Started in June 2022, the programme includes scheduled live webinars with question-and-answer sessions and self-paced e-learning courses. Topics include practical SDMX structural metadata maintenance, data collection, data validation and data transformation, all with concrete hands-on examples using the Bank's Fusion Metadata Registry SDMX software tool. The programme meets the needs of people who have varying skill levels, with some events and e-learning specifically designed for novice users, while others tackle more advanced use cases. Five webinars have been delivered to date, with an average of 70 participants each. Two e-learning courses have also been developed, with additional courses planned for 2023.

22. OECD has made the .Stat Academy platform[4] available, offering free online training and resources to support capacity-building in the .Stat Suite and data modelling in SDMX for data toolers and producers. The platform comprises several courses, each of which includes interactive quizzes and provides user certification.

V. Software tools and applications

23. Eurostat has continued to upgrade the SDMX tools that it develops and maintains; in particular:

- Various upgrades to the error reporting functionalities of the Eurostat SDMX converter and of its SDMX structural validation service have been implemented.

[4] https://academy.siscc.org/.

- The Mapping Assistant, which is an integral part of the Eurostat SDMX reference infrastructure, was made available as a web application. Data browsing functionalities were also incorporated within the Mapping Assistant.
- Eurostat has begun the implementation of SDMX 3.0 in its various tools and in the SDMXSource library. The tools are available as open-source.

24. The Fusion Metadata Registry is a mature, free-to-use SDMX "structural metadata registry" developed and maintained by the Bank for International Settlements. The Registry helps organizations to externalize, centralize, govern and control their statistical metadata with the benefits of maintainability, reuse, standardization and harmonization. The Registry supports the standard SDMX application programming interfaces and provides a graphical interface helping users with the task of authoring and maintaining SDMX structural metadata artefacts, including concepts, code lists and data structure definitions. In addition to its core metadata management role, the Registry provides SDMX data validation and transformation functions. It also helps to simplify the data collection workflow by providing data reporters with customizable Excel forms into which only the observation values need be entered. The latest version of the Registry, FMR 11, was released in January 2022, with support for the core elements of the SDMX 3.0 information model, version 2.0 of the REST application programming interface, SDMX 3.0 data set mapping and a subset of the revised structure and data transmission formats. Backward compatibility has been maintained with the majority of the SDMX 2.1 specification to ease the transition by organizations to both FMR 11 and SDMX 3.0. To date, FMR 11 implements about 70 per cent of the SDMX 3.0 specification, with the remainder planned for completion within 18 months. While it has always been free to use, the aim is to publish FMR 11 as an open source in 2023 under a standard permissive licence. The Java "SDMX core" library that underpins the Registry is expected to be of particular interest to SDMX application developers, providing an up-to-date and SDMX 3.0-compliant replacement for the popular but aging SDMXSource open-source library originally released in 2013. The Registry's ambitious three-year development road map is to significantly expand the product's functionality and improve support for demanding production workloads. New features include publication tables (online publication of curated dynamic tables of statistics), transformation and mapping microservices, new use-case specific graphical user interfaces and more efficient handling of big data and metadata. The Registry is in production use by national central banks, national statistics offices and other official statistics organizations worldwide. The "containerized" version, which allows users to create a fully functioning installation in less than 10 minutes, has proved very popular and helps to reduce barriers to entry to both the Registry and SDMX.

25. The Statistical Information System Collaboration Community, led by OECD, is a group of statistical organizations with common goals, sharing similar needs in terms of data dissemination and production, that are interested in mutualizing costs and sharing knowledge and committed to fostering common standards. The Community develops open-source digital solutions to support official statistics for the world at large. The solution, the .Stat Suite, is an SDMX native standard-based, componentized, open-source platform for the efficient production and dissemination of high-quality statistical data. The product road map is progressively covering the full (macro) data life cycle, enabling organizations to bring added value as "trusted quality data" producers, empowering them with "quality assurance by design" and "advanced workflow and reporting" mechanisms, "high performance calculation" and import and export tasks to offer efficiency gains and good quality in user experience in data operations. The Community also works in close partnership with Eurostat by contributing and enabling the reuse of open-source and common modules from the

SDMX reference infrastructure package, to which the Community contributes; the implementation of SDMX 3.0 in .Stat Suite will rely on that collaboration. The .Stat Suite is developed under a continuous user-driven product delivery model, with a state-of-the-art software delivery (DevSecOps) mechanism to combine transparency, agility, multiplicity of contributors, simplicity of deployment on the cloud, software quality and security. The .Stat Suite continues to be rolled out in a number of entities and organizations[5] and to be successfully configured on the United Nations Global Platform by the Statistics Division, OECD and ESCAP. It is currently being used by Cambodia as its official statistical data warehouse, while Kyrgyzstan and Maldives are at the pilot stage, with further avenues being explored for making the .Stat Suite on the United Nations Global Platform available for use by other national statistical offices. The .Stat Suite is also the platform powering the ILO-led labour market information systems, now active in more than 20 countries.

26. Since January 2020, the European Central Bank has been running an information technology project, SPACE, which implements new statistical data production and compilation environment using big data technology, such as the full Hadoop stack, and leveraging SDMX tools, such as the Fusion Metadata Registry. SPACE provides functionality to ingest, access and manipulate SDMX data with such programming languages as Python and R, implementing business processes logic on top of the big data platform. In 2022, the SPACE project is delivering statistical processes in a staged approach and under a new working model, using Scrum as framework for agile software development.

27. The IMF SDMX Central continued to support member countries in the conversion, validation and registration of SDMX data in their National Summary Data Pages. Since 2016, IMF has been providing website and web services free of charge, promoting SDMX as an efficient exchange medium for countries that adhere to the Enhanced General Data Dissemination System and subscribe to Special Data Dissemination Standard and Special Data Dissemination Standard Plus. SDMX Central plays an essential role in the automated data collection processes between IMF and its member countries.

28. ILO continued to upgrade its SDMX Constructor tool (formerly known as the Data Structure Definition Constructor), adding new features to create and maintain more artefacts, and added support for the computer-assisted translation of structural metadata. A new feature called “table modeller” allows a user producing all the SDMX artefacts required to model a statistical table in SDMX by dragging and dropping concepts in a panel, in an intuitive way, leveraging the advantages of SDMX information model to foster the harmonization of concepts. A new map service definition editor now enables users to create and maintain reference metadata structures. A new version of the SMART tool has been released, including an Excel connector that enables the ingestion and processing of multidimensional Excel tables in order to recode and reformat the output on the basis of a specification given by a data flow or data structure definition. Both tools can now interoperate with any standard SDMX web service and, in the case of .Stat Suite platforms, push metadata and data into the data warehouse.

[5] At the national level: Australian Bureau of Statistics, Federal Statistical Office of Switzerland, Federal Competitiveness and Statistics Authority of the United Arab Emirates, National Institute of Statistics of Chile, Ministry of Labour and Social Security of El Salvador, National Institute of Statistics of Tunisia, National Bank of Belgium, Statistics New Zealand, Statistics Canada, National Institute of Statistics and Economic Studies of Luxembourg, UK Data Service of the United Kingdom and Thailand National Statistical Office; at the international level: Food and Agriculture Organization of the United Nations, ILO, Pacific Community, United Nations Children’s Fund and ESCAP.

VI. Events

29. The tenth SDMX Expert Group Meeting was hosted virtually by IMF from 25 to 28 January 2021. The workshop offered an excellent opportunity for SDMX experts in the community of official statistics agencies, central banks and beyond to discuss and influence the further development of SDMX standards and various statistical guidelines related to SDMX. Taking advantage of the virtual format, the meeting was open to anyone interested in the technical discussions on SDMX and attracted 580 participants from 84 countries. The meeting was rich in content, with 35 presentations covering six critical areas of the standard: (a) the new features of SDMX 3.0; (b) SDMX use cases from countries and international organizations; (c) a comparison of alternative modelling approaches; (d) a stocktaking of existing and potential SDMX tools; (e) referential metadata models; and (f) the relationship of SDMX with other international standards. All presentations are available on the meeting website hosted by IMF.[6]

30. The eighth SDMX was hosted virtually by the National Institute of Statistics and Geography of Mexico, from 27 to 30 September 2021. The Conference was organized for the first time in the Latin American and Caribbean region, after previous editions in South Asia, Africa and Eastern Europe. The theme of the 2021 Conference was "data without barriers", which showcased how SDMX can help national and international agencies to simplify, strengthen and modernize data and metadata exchange. Furthermore, SDMX 3.0 was officially launched at the Conference. Given its virtual format, registration for the Global Conference was opened to everyone and attracted more than 1,000 participants. Registered participants represented a wide range of countries, with a strong representation from Latin American and the Caribbean region (about 350). The Conference agenda offered 57 presentations, comprising contributions from sponsor agencies (37 per cent), national statistical offices (34 per cent) and central banks, other international organizations and the private sector (29 per cent). As in previous editions, the 2021 Conference also offered capacity-building training on SDMX, from 20 to 23 September 2021, with sessions covering the training needs of both beginner and more advanced SDMX users. All presentations and video recordings of the conference and training sessions are available on the Conference website hosted by the National Institute of Statistics and Geography.[7]

31. The eleventh SDMX Expert Group meeting was held from 28 November to 1 December 2022 and hosted by ILO and the National Institute of Statistics and Geography of Mexico on the premises of the Institute, in Aguascalientes, Mexico.

VII. Action to be taken by the Statistical Commission

32. The Statistical Commission is invited to take note of the present report.

[6] www.imf.org/en/News/Seminars/Conferences/2021/01/25/10th-statistical-data-and-metadata-exchange.

[7] https://sdmx.snieg.mx/globalconference/english/index_english.html.