



SDMX Roadmap 2021-2025

Executive Summary

The Statistical Data and Metadata eXchange (SDMX) is a standard developed to facilitate the exchange of data and metadata in the international statistical community. It is widely used by international organizations, national statistical offices, and other data producing agencies. The development of SDMX is guided by a roadmap that sets forth the objectives for a span of five years.

The SDMX Roadmap 2021-2025 (“the Roadmap”) provides a strategic plan for the continuous development of SDMX over the next five years. The Roadmap has two main drivers: addressing a broadening set of needs and improving the usability of the standard. These two drivers come together under the umbrella of the SDMX 3.0 project. The new roadmap continues to group the objectives into four pillars: (i) Implementation, (ii) Simplification, (iii) Modernisation and (iv) Communication.

The Roadmap aims to strengthen the implementation of SDMX. SDMX will be expanded to new statistical domains. Existing SDMX packages will be upgraded based on updates of international classification standards. The adoption of established global code-lists and data structure definitions (DSDs) will be encouraged. Finally, SDMX tools will be finalized to facilitate the exchange of referential metadata.

Simplification will be required to facilitate the use of SDMX. Main directions of work will be to simplify referential metadata exchange mechanisms, explore links between SDMX and other standards, and develop tools that simplify the complexities of SDMX for novice users.

SDMX will be leveraged to modernise statistical processes and IT infrastructure. The development of SDMX 3.0 will incorporate modern features of data management. Other priorities are to develop cloud-based SDMX applications and tools compatible with the Validation and Transformation Language (VTL).

The Roadmap envisages to step up efforts to improve communication and better interaction with the broader community. Training materials for SDMX users and trainers will be expanded and upgraded. Core training materials will be translated to broaden use in non-English speaking countries. An SDMX Users' Group will be created to disseminate knowledge of the standards.

The SDMX Sponsors commit to pursue the goals outlined in this Roadmap and invite the wider community of statisticians, analysts and data professionals to contribute to the development of SDMX.

Introduction, motivation

SDMX, short for Statistical Data and Metadata eXchange, is a standard developed by seven international organisations¹ to facilitate the exchange of data and metadata in an international, official statistics context. Since its inception and through various editions published since 2004² the standard grew in acceptance and it became fit to various use-cases beyond what had been originally identified or anticipated.

The declared purposes SDMX were to:

- a. develop, maintain and promote internationally recognised standards, guidelines and implementation tools that can facilitate the exchange of statistical data and metadata using state-of-the-art information technology;
- b. bring together and leverage the existing expertise on statistical data and metadata exchange in the Sponsor Organisations and their constituencies and avoid duplication of effort in respect of these activities;
- c. develop and promote more efficient processes for exchanging and sharing statistical data and metadata among the Sponsor Organisations and their constituencies as well as by any other collectors, compilers, processors, analysts and disseminators of statistical information;
- d. identify, share and promote good practices in the exchange and sharing of statistical data and metadata across a wide range of social and economic statistics.

The original scope for SDMX was humble, it aimed at standardising the data exchanged between reporting agencies on one hand (typically national statistics offices and central banks) and macro-data aggregator recipients on the other hand (typically international organisations). At the time the landscape of organisations showed varied technological readiness, had partially harmonised statistical methodologies and terminologies, and all had largely differing reporting and data presentation templates. Therefore, both reporters and recipients had to multiply work to cater for the variability and volatility of the landscape they each faced.

However, in nearly two decades SDMX matured, acting as a catalyst for many of the harmonisation efforts. Today SDMX is considered for other uses and proved its fitness for providing a backbone for enterprise scale applications, such as data warehousing, a data aggregator's workflow management or dissemination platforms aiming at wider audiences.

Recent experiences and the recognition of the broadening scope of SDMX applications warrant a renewal of the standard. The next five years will be thus characterised by the preparation and adoption of a new major version of SDMX. This coming five years will also be a busy period for many statistical domains – the co-ordinated renewal of methodologies in domains of macro-economic statistics, international classification revisions and progress in global harmonisation initiatives, such as the Sustainable Development Goals - will all set implementation and adaptation actions for SDMX.

¹ The Bank of International Settlements (BIS), the European Central Bank (ECB), Eurostat, the International Monetary Fund (IMF), the Organisation for Economic Co-operation and Development (OECD), the United Nations Statistics Division (UNSD) and the World Bank.

² Version 1.0 released in 2004, version 2.0 released in 2005, version 2.1 released in 2011.

About the standard

The seven Sponsor agencies guide the standard via a governance structure composed of multiple standing bodies, with two main pillars: a technical working group (TWG) and a statistical working group (SWG). Under the auspices of the two working groups, SDMX reached the status of an ISO standard composed of:

- a. An information model describing typical statistical entities, and a number of related concepts that play a role in the processing and exchange of data and metadata in the international statistical system.
- b. XML, JSON and CSV representations useful for diverse use cases in the exchange, dissemination and analysis of statistical information.
- c. A description of registry services and a webservice-API (application programming interface) for data and metadata querying. These interfaces/services when implemented facilitate automation, facilitate the design of modular applications, the adoption of a modern service oriented architecture (SOA), and not least provide adopters a high level of compliance with Open Data principles³.
- d. Beyond the narrowly interpreted technical standard, the SDMX initiative is complemented with *content oriented guidelines* to harmonise statistical content: concepts, terminology and even reporting templates, so that the variability of the content - which needs to be described by the standard - decreases, helping both reporters and recipients. Coordination efforts accompany the guidelines to help put SDMX into practice.

With time, based on the technical standard, software components, and complete tools were developed, often in collaboration between multiple agencies, and often resulting in open, reusable, public products. Much of the ensuing work is now available publicly in the SDMX global registry.

Developments, achievements, lessons from the last 5 year roadmap

As the end of the previous planning cycle approaches it is time to take stock of developments. Progress and achievements are listed by the four pillars of the 2020 roadmap.

Implementation

The implementation section had the ambition to support the implementation of SDMX tools and exchanges in various statistical domains, but also to promote technical adoption beyond the context of specific domains. Progress was monitored via surveys prior to the biennial SDMX Global Conferences. Many of the envisaged implementation actions were completed:

Global SDMX packages have been published, and for many of them, data exchanges have started, for example for Sustainable Development Goals (SDGs), Price Indices, the System of Environmental Economic Accounts (SEEA), Education and Labour statistics, and International Merchandise Trade Statistics (IMTS).

³ Contributes to open data principles 3 and 4 as in: <https://opendatacharter.net/principles/>

Besides the globally harmonised data packages, various international organisation specific programs generated opportunities for SDMX implementation in data providers' systems.

Data sharing pilot studies under the auspices of the IAG looked into reporting template harmonisation and organising a cascading collection of data in National Accounts, Balance of Payments. Most recently, Price statistics and SEEA related data sharing arrangements were discussed. These data sharing pilot studies investigated not only the data templates to be exchanged but also exchange modalities⁴ and while exchanges were successful, it was clear that the standard and the tools will have to evolve to support the desirable pull mode sharing with notification, as well as time based differential updates, so that a much larger scope of data can be covered by harmonised exchanges.

New and updated guidelines were issued for “Modelling Statistical Domains with SDMX”, and a broad range of other topics, such as versioning or coding patterns in code-lists, just to name a few.

Progress in implementing referential metadata exchanges was slow, nonetheless some progress was made in the form of the release of a new version of the “SDMX Glossary”, and guidelines on how to map individual reporting templates to and from the glossary (and a global concept scheme).⁵ Progress in tools development is going to be needed for significant improvements in SDMX referential metadata exchange adoption.

Simplification

Efforts to ease adoption of the standard in various contexts (such as data visualisation or in countries with less mature IT production infrastructures) are regrouped under this pillar. A significant achievement was to publish two new formats: SDMX-JSON and SDMX-CSV in support of new use cases and simplifying the interaction with other tools outside of the standard.

Another main development was the evaluation of SDMX with respect to other standards such as linked-open data, data cataloguing standards, XBRL, DDI, geospatial data standards, GSIM, CSPA etc. The evaluation outcomes provided inputs towards the new major update of the standard which will better fit or provide better surface/interface for connectivity with other standards (e.g. improving geospatial data incorporation, preparing for more straightforward microdata and administrative data representation).

Finally, with encouragement of the sponsors, nonetheless fairly independently, the easy access in various analytical environments has improved. Tools were produced to encapsulate SDMX web services and integrate data communication into R, Python, Stata, Eviews and Java.⁶ This is an efficient way to relate to various user communities, which simplifies the use of SDMX in commonly used statistical packages.

⁴ Push vs pull – with notification, gateway vs sharing mode, differential vs. complete cube-region

⁵ See [Standardising Reference Metadata Reporting in SDMX](#)

⁶ See the list of tools in: https://sdmx.org/?page_id=4500

Modernisation

Tools and software development continued in the SDMX community, e.g. BIS is sponsoring Fusion Metadata Registry, a free tool for managing structural metadata. Similarly, the Eurostat maintained Reference Infrastructure has evolved and has been widely used, and re-used, for example in OECD's .Stat Suite (which besides being a collaborative development exercise, has become open source during the last five years). A number of other utility tools were produced and provided for the SDMX community (ILO, IStat, etc.)

A new version of the SDMX RESTful API was published, and a Technology Compatibility Kit (TCK) is under development to help implementers verify the coverage and compliance of their web-service implementation with SDMX.

The technical and statistical working groups have jointly prepared a portfolio of 14 work packages⁷ which will bring a major improvement to the standard. These work packages have reached an advanced stage of technical specification. Their implementation and publication as SDMX 3.0 will be due during the period covered by the next roadmap.

To make SDMX more dynamic and integral part of the statistical production process a Validation Transformation Language (VTL) was designed to complement the more structural, static components of SDMX. Acknowledging that SDMX exists in an ecosystem of statistical modernisations standards (including such standards as GSBPM, CSPA, and GSIM) VTL 2.0 was designed to work with multiple environments and in particular: GSIM and SDMX. A major remaining step is the connection of VTL to SDMX, so that SDMX messages can invoke VTL expressions and connect the expressions to SDMX artefacts. This work envisaged to be part of SDMX 3.0.

Communication

A review of the two main recurrent events took place: the alternating biennial Global Conference and Experts Meeting and their respective role in communication and capacity-building has been delineated, the former aiming at diverse audiences, with a strong outreach and communication element as well as capacity building, while the latter focuses on forward looking aspects of the standard with a more select audience, and in terms of capacity building addressing training needs of trainers and advanced users.

Besides communication via and around the conferences and recurrent events the www.sdmx.org website has been revamped, and serves as a central location for all content around SDMX, while additional communication channels have been opened via Twitter and LinkedIn.

Sponsor agencies, with the help of other international organisations – especially regional agencies – have carried out trainings and workshops, often in the context of a statistical modernisation or data initiative (SDGs, SDDS).

⁷ Following public consultation and prioritisation by Sponsor organisations 14 work-packages were retained out of the original 23 proposals.

The roadmap 2021-2025

It follows from the previous section that, many of the developments of the last 5 years need follow up, either because the successes of the statistical domains already working with SDMX are worth replicating in other domains, or indeed because the work done so far paved the way for a natural next step or developments opened new opportunities. Due to the success of the previous roadmap, the new one is organised around the same four pillars (Implementation, Simplification, Modernisation, Communication).

The SDMX Sponsors commit to pursue the laid out goals, and at the same time, invite the wider community of statisticians, analysts and data professionals to contribute to the set goals and the cause of SDMX. The ultimate goal is to establish a thriving open ecosystem of SDMX, in which the Sponsors are open for user feedback and ready for continued SDMX support via communication, training and consultancy, nonetheless in which support is also provided by a broad range of SDMX adopters and industry partners, who can act as SDMX ambassadors and contributors.

The 2021-2025 roadmap has two main drivers: addressing a broadening set of needs and improving the usability of the standard. These two drivers come together under the umbrella of the SDMX 3.0 project:

- On one hand work will aim at materialising SDMX's potential to play a key role in animating/powering the international statistics ecosystem; becoming the central piece in the integration of the data-workflow of a typical data aggregator from data collection, via validation, data re-shaping, mapping, computations, aggregation, all the way to data dissemination; while extending the business case for a broader production cycle coverage, the standard will stretch to better represent micro/administrative and geospatial data, beyond traditional aggregated data
- On the other hand, the standard will be streamlined and adjusted based on experiences gained with an increased number of SDMX system implementations and SDMX based exchanges (e.g. introducing semantic versioning of objects, code-list inheritance and composition, deprecating lesser used formats and flavours);

The recent emergence of Big Data trends was already generating new demands to treat data that are more granular, in increased volumes and timelier. However the context of the COVID-19 pandemic re-iterated how much the international community needs the capacity to not only quickly react to emerging data needs in ad-hoc ways, but to be able to do so by mounting efficient and maintainable data pipelines – supported by broadly adopted data standards.

Strengthening the implementation of SDMX (Implementation)

1. Based on the success of the coordination exercises to construct global SDMX packages for entire statistical domains – supported by modelling guidelines and project management check-lists exercises should be opened in new domains (e.g. Business Statistics, Income and Consumption, Population, Housing and Migration, Regional Statistics, Energy and Health). There are a number of initiatives where work is already ongoing (e.g. Research & Development) where a boost or relaunch may be needed for completion, or partially completed domains can step to a second stage

broadening scope (e.g. the Price statistics community having completed Consumer Prices will work towards a Residential Property Price Statistics package).

2. While maintenance and regular minor updates of the existing SDMX packages continue, some domains already prepare for a major update (e.g. National Accounts and Balance of Payments) in the timeframe of the roadmap; driven primarily by the planned release of key classification updates such as the International Standard Industrial Classification, but opportunistically packaging other withheld critical updates. This work will cut across various packages – and will be coordinated by the Ownership Group for Macro-Economic Statistics.
3. Extending the adoption of the existing global SDMX packages should be pursued. It is when practical data exchange arrangements are based on the established global code-lists and DSDs when the benefits of the coordination exercises materialise.
4. Recognising that despite the considerable domain-by-domain efforts to harmonise concepts, code-lists and modelling patterns, differences between domains emerged. These differences may exist for reasons such as:
 - a. combining far away subdomains of statistics with un-harmonised conceptual frameworks: e.g. social statistics vs. economic statistics;
 - b. statistics produced for different purposes – dissemination vs time-series analysis, or data-warehousing.
 - c. cross cutting assemblies of data: e.g. short terms statistics, regional statistics, SDGs etc. vs the original tightly integrated statistical domains from which they are assembled.
 - d. legal and related reasons, such as reference areas based on customs territories rather than countries.

The data appearing in these various objects may be grouped or represented differently, however relationships, overlaps between different representations will prevail. To facilitate connections between these separate collections/presentations of data, and indeed to allow effortless transformations – *data mappings should be used to express linkages explicitly*. Although this goal temporarily existed in the previous roadmap, progress has not been made because of the lack of tools and the complexity of (machine optimised) mapping representations. This is likely to change with data mapping improvements planned in SDMX 3.0, the likely emergence of tools supporting mappings and transformations. In anticipation to the tools and standard evolutions reusable maps should be provided for code-lists e.g. between the three flavours of ISO codes: 2 digit alphanumeric, 3 digit alphanumeric or M49 numeric codes and mappings should be maintained between main global SDMX packages and related international shared collection artefacts.

5. In the last 5 years groundwork has been done to build an all-encompassing referential metadata concept scheme and guidance has been provided as to how to relate existing metadata templates to this global scheme. Nonetheless work is needed to improve the readiness of SDMX software tools to enable (facilitate) the creation and exchange of referential metadata. The objective hence for the next 5 years is threefold: improve tools availability for referential metadata manipulation, simplify the model for referential metadata, and finally increase the number of actual SDMX-based exchanges, or referential metadata dissemination via SDMX.

Making data usage easier via SDMX (Simplification)

1. As SDMX use for data dissemination and data visualisation is getting more widespread, research should be carried out to identify the best ways to approach these business cases without increasing the complexity of the standard. The avenues to explore are: develop guidelines that specifically address these cases; evaluate the fitness of existing formats under dissemination/visualisation scenarios – and consider adjustments; evaluate annotations for process metadata usage, dedicated attributes or referential metadata models to be used to store presentation metadata, process metadata and para data).
2. Simplify and improve referential metadata exchange mechanisms to facilitate both its usage and development of tools in its support.
3. Phase out of old formats that can cause confusion and put additional burden on software development aiming for complete support of the standard.
4. The 2020 SDMX roadmap put a lot of emphasis on the exploration of links between SDMX and other standards (e.g. XBRL, RDF, DDI etc.). The new roadmap calls for follow-up implementation actions for improving cataloguing, navigation and discoverability (e.g. the publication of SDMX metadata assets as Linked Open Data, DCAT or Schema.org catalogues), as well as the exploration of how SDMX may relate to emerging standards for new use cases (e.g. Big Data/Microdata, Internet of Things etc.).
5. Consumption of SDMX by chatbots and other NLP (Natural Language Processing) enabled conversational data experiences is a new field of research. These applications, which are just emerging, will generate needs like those generated by interactive applications designed for dissemination and data visualisation. SDMX support for semantics has the potential of decreasing the cost of the implementation of such bots and making easier to transfer knowledge from one organisation to the other, one domain to the other or one language to the other. Leveraging SDMX to make data more accessible through chatbot conversations is set as an objective for the community.
6. Develop, maintain and promote tools designs that hide some of the complexities of SDMX for the novice users, but nonetheless maintain accessibility for a deeper SDMX layer for expert users.

Using SDMX to modernise statistical processes, as well as continuously improving the standards and IT infrastructure (Modernisation)

1. The main deliverable for the next 5 years is going to be the completion of 14 packages that constitute SDMX 3.0, and publication of the new standard. This large undertaking will aim at:
 - a. Introducing semantic versioning
 - b. Improving support for microdata and geospatial data
 - c. Modernising and extending the functionality of the web API
 - d. Simplifying and yet generalising referential metadata support
 - e. Making more user-friendly and more versatile the management of mappings and code-lists

2. As a follow-up task, changes should be reflected in the ISO standard specifications.
3. Consecutively the full implementation of SDMX 3.0 in the various SDMX tools maintained by the Sponsor organisations and by other key stakeholders is expected. Ideally, each concerned organization should set out milestones for the adaptation of their tools to SDMX 3.0.
4. Prepare cloud based reusable SDMX applications to lower costs of SDMX implementation. The tool may be prebuilt with standard DSD's used in most national statistical offices and central banks.
5. The update of the technical standard should be followed by a modernisation and update of the content-oriented guidelines. For example, many aspects of the modelling guidelines will need to be updated to reflect the availability of new features, such as the possibility to compose / combine code lists. At the same time, new guidelines may need to be created for the new features made available or reinforced by SDMX 3.0 (e.g. guidelines on microdata modelling and on mapping artefacts).
6. As a stable version of the new VTL language was released in 2018, the development of VTL-capable tools is also one of the priorities for this road map.

Improving communication and better interaction with the broader community (Communication)

1. The development of regional implementation road maps will remain a focus for the next SDMX road map
2. The improvement of standard training materials for SDMX users and trainers shall continue. Consider joining forces under an umbrella of an 'SDMX Academy'.
3. Over the past years, there has been an increase in requests for translations for SDMX specifications or training material: work in this direction, identification of core content to be translated, and organising a volunteer network to accomplish the translations will be considered.
4. Reach out to new stakeholders (chambers of commerce, ministries, and the private sector in general) and integrate traditional and new stakeholders; reach out and address a broader community - to date, the primary focus has been on member organisations of national and international statistics services.
5. Reach out and forge partnerships with the private sector, find synergies, promote the benefits of the standard more actively, help ease the adoption of the standard via offering seminars and knowledge transfer sessions, and thus enable the sector to tap into the increasing volume of business opportunities around the standard.
6. Within the traditional official statisticians' network, step up efforts to bring back harmonisation via SDMX to the agenda – both in subject matter networks and technological modernisation networks where the agenda recently was predominantly driven by Big Data and related topics, without connecting to SDMX.

7. Foster the creation of an "SDMX Users' Group" as a way of disseminating the knowledge on the standard by providing support and building capacities on the standard itself and related topics.
8. The Sponsors shall develop a sustainable model for the funding and management of common promotion, training and support initiatives in the next 5 years.