THE BUSINESS CASE FOR SDMX

Version 1.0 - July 2020

Contents

Introduction

The purpose of this document is to introduce the business case for SDMX, a standard developed by the main international actors in official statistics[1] with a view to standardising and modernising ("industrialising") the mechanisms and processes for the exchange of statistical data and metadata.

The expected audience for this document is senior and middle management as opting for SDMX is a critical business choice rather than a technical choice.

Since the publication of version 1.0 of the technical specifications in 2004 the SDMX tools have been continuously improved and fine-tuned, and the number of global implementations is constantly growing, covering such domains as education statistics, labour statistics, national accounts, and sustainable development goals. The current revision work of the standard will be a step further towards a tailor-made solution for the exchange of statistical information.

In this context any statistical organisation will at some point be confronted with the question of considering SDMX as a solution for harmonising and automating their data and metadata exchanges with international organisations or within their organisation, e.g. in the case of National Statistical Systems. Even countries that in the past did not adopt SDMX may reconsider their decision, taking into account the significant progress made over the years.

It is hoped that this business case will help the reader in making a well-informed analysis on the SDMX implementation.

The first section of the document presents typical use cases and scenarios where SDMX can assist in modernising business processes. The second part of the document presents the "*seven benefits*" of SDMX that a statistical organisation may expect from the implementation of SDMX as well as the challenges that it may face.

SDMX is an integrated solution consisting of three main elements:

- technical standards (including the Information Model);
- statistical guidelines;
- IT architecture and tools.

These three elements can be implemented in a stepwise approach. This aspect can be decisive for countries with varying capacity levels.

---

[1] Bank for International Settlements (BIS), European Central Bank (ECB), Eurostat (Statistical Office of the European Union), International Monetary Fund (IMF), Organisation for Economic Cooperation and Development (OECD), United Nations Statistical Division (UNSD), and World Bank.

Typical use cases and scenarios for data and metadata exchange

SDMX can support various use cases and implementation scenarios, as described below. Whatever the scenario envisaged the essence of SDMX is to provide a set of common tools, common processes, common terminologies, and common methodologies that will facilitate the exchange of information between producers and consumers.

The most typical use case for SDMX is **data and metadata reporting** where a statistical organisation wants to improve its data and metadata collection procedures from other organisations. Besides the enablers described in the previous paragraph, the infrastructure proposed by SDMX to achieve this objective is made of a network of registries which make publicly and centrally available globally or internationally harmonised Data Structure Definitions, a set of reporting formats as well as various IT tools to assist implementers. Full implementation of such infrastructure will end up in direct machine-to-machine dialogue.

Another domain where SDMX can deploy its full potential concerns **data discovery and visualisation**; in this case the objective pursued by a statistical organisation is to make statistical data and metadata more easily discoverable and accessible by external users. Under this scenario SDMX is used as a queryable data source that drives website presentation of data and metadata. Based on user selection the application retrieves the requested information (data and metadata) from SDMX structural repositories, and transforms it into tables, graphs, charts, etc.

SDMX can also be used for automating the loading of data and reference metadata into databases, automating the creation of database tables, enabling a database to be compatible with SDMX Web Services, and implementing data and metadata warehousing.

The seven benefits of SDMX

Due to the similar nature of the statistical activities, all national or international statistical organisations face similar challenges, namely:

- harmonising statistical data and metadata content and structure across and within domains (e.g. similar terms can have a different content, different names and codes can refer to the same content, codes can be based on different structural rules);
- reducing the reporting burden for data providers (e.g. providing data in the most efficient way: reporting data only once, or even stop reporting altogether and offering data through web services or data hubs);
- reducing costs for end users who build applications to access data made available by various institutions;
- reducing development and maintenance costs;
- delivering data faster by still maintaining high data quality;
- improving data quality through better and faster validation;
- improving harmonisation and streamlining of statistical business processes;
- addressing the increased demand for data;

- making data collection and data sharing simpler, less labour-intensive, less manual, and cheaper;
- reducing the maintenance costs linked to different data collection and data sharing channels (surveys, web queries, data files, metadata files);
- reducing the number of formats used for transmitting data (paper, MS-Excel sheets, web forms, flat files, etc.);
- reducing the number of media used for transmitting data (email, CD-ROMs, file uploads, etc.);
- reducing errors and inconsistencies resulting from the difficulty to validate and process unstructured or poorly structured data.

SDMX can respond to all these concerns as will be demonstrated below.

In short SDMX is about changing from a multiple, diverse and complex exchange system, to a common, harmonised and standardised exchange system. The main benefits of SDMX can be grouped into seven major categories:

1. SDMX inspires trust;
2. SDMX improves coherence and comparability;
3. SDMX supports modernisation;
4. SDMX improves timeliness and accessibility;
5. SDMX reduces costs and reporting burden;
6. SDMX removes barriers to implementation and data accessibility;
7. SDMX provides access to a global community of practitioners.

Benefit 1 : SDMX inspires trust

SDMX is a **global** response that addresses how to improve the exchange of statistical data and metadata. The seven international organisations sponsoring SDMX are major official statistical data producers, and they are collaborating closely with countries throughout the world.

SDMX is here to stay. **Stability** for the long run is guaranteed by the commitment taken by the sponsor organisations, and their strong and continuous involvement in SDMX for more than 15 years now.

The reputation of the SDMX sponsors and their commitment to fundamental principles governing official statistics, such as the United Nations Fundamental Principles of National Official Statistics, the African Charter on Statistics, the Code of Good Practice in Statistics for Latin America and the Caribbean, the European Statistics Code of Practice, and the Public Commitment on European Statistics by the European System of Central Banks (ESCB), **foster trust** in the standard. These fundamental principles aim to ensure that statistics produced by international organisations are not only relevant, timely and accurate but also comply with strict professional principles.

SDMX has a strong and well established **governance**. The SDMX sponsor organisations have put in place a governance model with various levels of responsibilities: the Sponsors Committee for the strategic decisions, the Secretariat for the operational management, and two working groups (the Technical Working Group and the Statistical Working Group) for the development of the technical standard and the statistical guidelines. These two groups strive to be as reactive and proactive as possible, meaning that SDMX is **driven by users' needs**. SDMX also provides a proven governance model to manage the life-cycle of a domain's SDMX objects and versions.

SDMX is an **ISO** standard (17369:2013). The International Organization for Standardization (ISO) was founded with the idea of answering a fundamental question: "What is the best way of doing this?" International standards mean that users can have confidence that products are reliable and of good quality. They can also help ensuring that separate practical implementations (e.g. IT modules/applications supporting statistical production) are harmonised in terms of inputs and outputs, can interoperate with each other, and can successfully be developed collaboratively and then shared among agencies. Furthermore, international standards also facilitate international comparability of statistics.

Earlier on in 2008 the United Nations Statistical Commission at its thirty-ninth session "*recognized and supported SDMX as the* **preferred standard** *for the exchange and sharing of data and metadata, requested that the sponsors continue their work on this initiative and encouraged further SDMX implementations by national and international statistical organizations*".

SDMX is a standard based on **shared experiences** among the sponsor international organisations and their constituencies.

SDMX is **specifically targeted to the needs of official statistics** with all the necessary metadata involved. The use of modern, internet-based formats with extensive support for statistical metadata is clearly an advantage above sending MS-Excel sheets or csv files by email, referring to loosely coupled metadata specifications.

SDMX and its sponsoring organisations provide capacity building and training in various forms (dedicated websites, videos, webinars, international conferences).

Challenges: Challenges include building institutional capacity and ensuring sustainability of SDMX-based data processing, exchange, and dissemination. There is a learning curve associated with an entry to SDMX, and training a reasonable number of experts in an environment where staff turnover is relatively high, particularly on the IT side, is often a challenge whichever technology is used. In the case of SDMX, however, the availability of documentation and reliance on standards rather than proprietary applications, is instrumental to ensuring the sustainability of statistical information systems.

Benefit 2 : SDMX improves coherence and comparability

SDMX provides a **standardised** way of organising and exchanging data and metadata, enabling interoperable implementations within and between systems concerned with the exchange, reporting and dissemination of statistical data and related metadata.

SDMX provides a common **information model** for describing statistical data in an environment where there is a wide range of different data models and transmission formats used for the exchange of data and metadata. It is not the format, but the information model that is the power behind SDMX; it is syntax- and format-agnostic, and consequently most processes and functions can be developed around the model, and not around the syntaxes.

SDMX is supported by a **common terminology**. Concepts used for data and metadata exchanges are harmonised, clearly defined, and made publicly available. This also holds true for code lists which are a central element of data exchanges. This strong terminological foundation makes it possible to describe data and metadata in any statistical domain.

Standardisation makes **tool development much easier** as SDMX objects are fit-for-purpose, i.e. the relationships between the objects and their usage are clearly defined. The extremely structured information model of SDMX can serve to build highly interoperable tools while at the same time making it easy to use a part of it rather than implementing the full standard. The standard is very responsive to new technologies as new syntax representations can be constructed easily while still based on the information model.

The **advantages of harmonising and standardising** statistical data and metadata content and structure are numerous:

- implementers <u>and</u> users speak the same language;
- reusing existing material saves time and resources; reuse is facilitated by the existence of SDMX registries;
- less mapping and data processing;
- numerous existing SDMX objects are available for defining data and metadata;
- tools based on a commonly agreed format have wider audience.

The **harmonisation** triggered by SDMX extends **beyond** the mere scope of **SDMX**, e.g. the SDMX Glossary is used beyond the SDMX sphere.

SDMX improves **interpretability** because it harmonises structural metadata (the identifiers and descriptors of data, such as table columns); it harmonises reference metadata (the content and quality of data, such as the methodological principles applied in a survey); it harmonises its terminology (the SDMX Glossary); it should be noted that even if you have not yet implemented SDMX you might already use its terminology as the Glossary is more and more used beyond the mere scope of SDMX. SDMX contributes modestly to the development of a

global statistical language (along the lines of the models developed under the auspices of UNECE: GSBPM, GSIM, GAMSO, CSPA[2], etc.).

SDMX can also be said to improve **coherence** through the use of cross-domain concepts, shared code lists and harmonised statistical guidelines, and the extensive reuse of SDMX objects across domains and agencies.

Another aspect of coherence is the ambition of a single figure dissemination, so-called "data sharing", where e.g. financial data collected by Eurostat and the ECB from their member countries are first validated and then shared with OECD which in turn collects data from non-EU constituencies, validates it and make it available to IMF which collects only data from non-OECD countries.
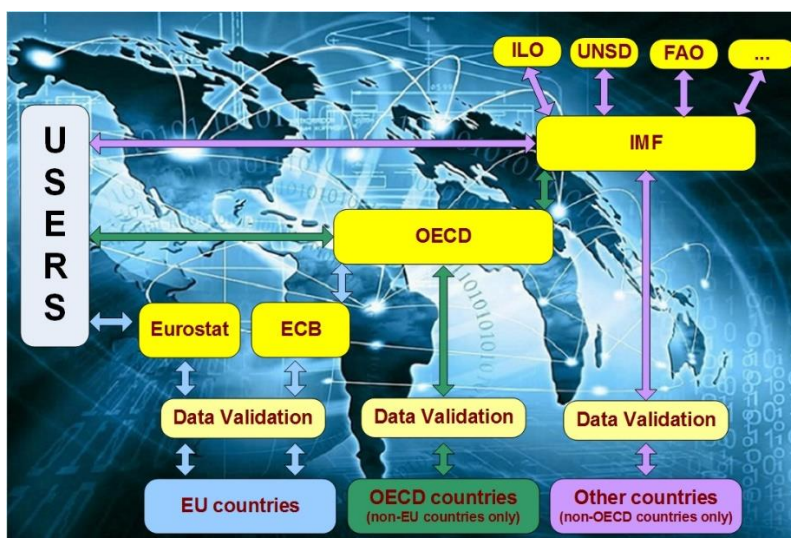
Example of data sharing process

There will be **more and more global data initiatives**, e.g. in areas such as SDGs (Sustainable Development Goals), so implementation of data exchange standards will increase accordingly on a global level.

Challenges: Harmonisation comes at a cost. On the technical side, unifying disparate systems is usually a major project that needs proper planning and allocation of resources. On the substantive side, developing common classifications and code lists requires an ongoing organisation-wide effort that individual departments may see as interfering with their work. These challenges are inherent in any harmonisation effort but SDMX reduces the cost by enabling organisations to make use of existing standard classifications and software as well as best practices instead of starting from scratch and potentially using approaches that have proven suboptimal elsewhere.

---

[2] GSBPM: Generic Statistical Business Process Model; GSIM: Generic Statistical Information Model; GAMSO: Generic Activity Model for Statistical Organisations; CSPA: Common Statistical Production Architecture.

Benefit 3 : SDMX supports modernisation

SDMX provides a **standardised** way of organising and exchanging data, enabling interoperable implementations within and between systems concerned with the exchange, reporting and dissemination of statistical data and related metadata.

SDMX can be a central element of the **modernisation of official statistics** (MOS) as promoted by the United Nations Economic Commission for Europe (UNECE). The main aspects involved in the "industrialisation" of statistics are the use of common tools, common processes, and common methodologies. Industrialisation recognises that all statistics are produced in a similar way and that no statistical domain is special. Finally, the adoption of these common elements provides increased flexibility to adapt to new sources and to produce new outputs.

SDMX is a **business choice** (as opposed to a technical choice), which is about improving the quality of exchanges through standardisation, automation, validation and data sharing.

We have more and more **corporate services** and SDMX is just one of them. The basic premise is that SDMX implementation must be seen in the context of a wide range of corporate institutional, infrastructure and statistical initiatives currently underway in almost all statistical agencies around the globe to improve the quality and relevance of the service they provide to government and non-government users of their outputs.

SDMX **supports** many **statistical activities** and the processes supporting these activities:
- data collection – data registration and data retrieval, data validation;
- data reporting and data mapping;
- data dissemination – data discovery, data query, data portal;
- structural metadata repository for metadata management, persistence, query, and retrieval;
- reference metadata reporting and dissemination, and linking metadata to data points.

SDMX is a good way of **avoiding "stovepipe" or "silo" approaches**. SDMX is an interface which is fully agnostic to the data and is thus applicable across domains and encouraging common solutions.

SDMX improves the **coordination of statistical activities** across the national statistical system (NSS) and among development partners.

The harmonisation of statistical business processes induced through horizontal and vertical integration also facilitates the **transferability of skills**. This holds true in both cases where organisation is structured by domain or along the business process.

Challenges: As with any change, modernisation is often met with resistance. While breaking stovepipes is unquestionably beneficial to a statistical organisation or system it inevitably leads to a restructuring of existing processes and coordinating activities that were previously carried out at the discretion of various organisational units. The transferability of skills facilitated by SDMX helps reduce the complexity of organisation- or system-wide changes while the management can refer to best practices to argue for the need for and direction of change.

Benefit 4 : SDMX improves timeliness and accessibility

A standard technical architecture promotes **timelier and better quality data**. Timeliness improves because less manual conversion is needed, and automated checks are fast, thus reducing "time to users". The quality of data improves because automated processing means less human error and faster access to data.

SDMX also triggers increased automation (automated workflows for exchange of statistics, thus reducing manual intervention errors; less "wait states"). Automation allows unattended workflow execution and promotes standard classifications, code lists, concepts schemes, etc. which reduces mapping and transformation errors.

SDMX improves **accessibility** due to the variety of data exchange channels offered. These include bilateral exchange, multilateral exchange, and data sharing. Data can also be accessed using push and pull modes and data hub architecture[3].

SDMX can **reduce data errors** thanks to automated validation (including the intrinsic structural validation based on the information model), agreed structures for transmission, time saved on conversion and mapping, and reduction in manual intervention.

Challenges: Among challenges to achieving higher quality and timeliness is a shortage of tools to facilitate automation of the entire statistical value chain. For example, there are few tools to link production micro-data systems and the dissemination/exchange side of the house. However, work is ongoing to address this issue, in particular on the Validation and Transformation Language (VTL). It is expected that solutions and tools to facilitate the automation of the statistical value chain will rely in large part on SDMX.

Benefit 5 : SDMX reduces costs and reporting burden

SDMX can **reduce IT** development and maintenance **costs**, mainly through its open source approach (benefit from a large community offering free tools and sharing expertise around a standard); furthermore, sharing resources worldwide is better than working nationally as it promotes cross-fertilisation of ideas and practices. Open source approach thus means no licensing costs, shared toolbox, improved interoperability between systems and applications, and sharing of development burden among the international community.

---

[3] For a precise definition of these concepts, please consult the SDMX Glossary.

SDMX can **reduce** the **reporting burden** via a series of features such as the pre-validation of the data message against SDMX data structures that will avoid multiple exchange of mail between data producers and consumers; the automation of publication via a web server by the reporter; the significant reduction in the number of reporting systems and exchange agreements to be maintained; and the possible use of the "pull" mode by collecting agencies.

Challenges: Implementation of automated data reporting requires mapping of internal database structures and codes to global reporting Data Structure Definitions, which can be a fairly complex exercise. Experts with knowledge of the subject-matter domain (e.g. National Accounts, Demography, SDGs) need to collaborate closely with data experts to understand the global structures and how they can be mapped to in-house databases. However, once the initial mapping has been completed, subsequent maintenance is much less complex, while the substantial effort on completing the various, often overlapping, questionnaires provided by international organisations, is reduced or eliminated. Mapping to and disseminating in accordance with the global DSDs also greatly improves interoperability of data being published.

Benefit 6 : SDMX removes barriers to implementation and data accessibility

SDMX has always taken seriously the idea that different organisations will implement at their own speed, and with their own objectives. The result of this is the **"toolkit" approach**: SDMX offers many different tools, but they need not all be adopted or used together. Differentiated implementation strategies are thus possible, making the standard accessible to countries or entities of varying capacity levels.

SDMX is indeed made of distinct components such as:
- technical specifications;
- statistical guidelines;
  - code lists;
  - glossary;
  - implementation guidelines;
- registry;
- governance;
- tools;
- capacity building.

These components can be used independently from each other, thus providing high flexibility in implementation pace.

The apparent complexity of the syntax representation of the data and metadata should not be a barrier to entry in SDMX as there are plenty of open-source **tools and code** that hide this complexity and thus enable to use SDMX to address most issues.

SDMX takes advantage of open software advancements and encourages the free sharing of SDMX tools. This facilitates the implementation, not only by data procedures, but it also helps end-users, data vendors, research institutes, universities, ministries, government agencies and policy makers to easily adapt open source software tools or easily build applications using free

library packages to access data produced by various institutions and made available in a uniform SDMX-based manner.

Challenges: Among the challenges linked to SDMX implementation one is the availability of documentation in national languages. Extensive documentation and guidelines on the implementation of SDMX is available but a majority of it is in the English language only. However, the SDMX community includes experts from across the globe, who speak many languages and can often help those who are not fully proficient in English. Furthermore machine translation has made tremendous progress over the last years and can now be considered as a relatively reliable solution for understanding the basics of SDMX.

Benefit 7: SDMX provides access to a global community of practitioners

As mentioned above, organisations benefit from gaining access to communities around the standards as well as tools. The organisations can also benefit from participating in events such as the SDMX Global Conferences and Expert Group Meetings, which further facilitate the sharing of experience and best practices.

The advantage of being part of the SDMX community goes beyond the implementation of data and metadata exchange and dissemination. Members of the community include data experts from National Statistical Offices, Central Banks, International Organisations, and the private sector, who possess vast amounts of knowledge of and experience in data modelling and implementation of statistical information systems using various tools and standards not limited to SDMX. Both positive and negative experiences are shared, i.e. solutions and approaches found to work well as well as those where difficult to overcome challenges were encountered. In short nothing replaces a good address book and bilateral contacts for progressing with SDMX.

By gaining access to the community, statistical data producers and users benefit from this global expertise, which can help them develop architecture and use tools that are most appropriate for their particular circumstances, including available infrastructure and skillsets.

Challenges: Participation in global or regional events can be very useful for joining the SDMX community but requires funding for travel. Developing countries can often find such funding from various development partners including the SDMX Sponsor Agencies, country offices of various UN agencies, and development banks. In any case, most of communication within the community takes place over the internet while remote participation in the events is increasingly common.

Giving SDMX a try

The most efficient way to assess the fitness for purpose of SDMX is to conduct a small-scale but representative pilot showing how SDMX provided significant improvements for that pilot and how that can translate into other processes in the organisation. By doing so, the strategic decision will be based on a concrete experience.

The first step is to assess what SDMX use case applies to your needs and which level of implementation may be needed. Different use cases and needs may require different levels of implementation and knowledge, ranging from the simple use of some ready-made tools (such as converters) to the local implementation of a full-fledged SDMX infrastructure.

You are invited to contact the SDMX Sponsors at contact@SDMX.org for further details and information

References

- Governance of commonly used SDMX metadata artefacts
- SDMX Glossary
- SDMX IT tools
- SDMX "Learning" section
- SDMX official website
- SDMX Roadmap 2020
- SDMX Starter Kit for National Statistical Agencies
- Validation and Transformation Language