The Role of SDMX in Official Statistics

A European Central Bank Perspective

Alessandro Bonara
European Central Bank, Frankfurt, Germany – Alessandro.Bonara@ecb.europa.eu

Zlatina Hofmeister
European Central Bank, Frankfurt, Germany – Zlatina.Hofmeister@ecb.europa.eu

Abstract

In 2001, seven Sponsor Organisations, the Bank for International Settlements (BIS), the European Central Bank (ECB), Eurostat, the International Monetary Fund (IMF), the Organisation for Economic Cooperation and Development (OECD), the World Bank and the United Nations joined forces to develop more efficient processes for the exchange and sharing of statistical information. This initiative was called Statistical Data and Metadata eXchange or SDMX. This paper has four objectives: it offers a short introduction to the SDMX information model used in official statistics; it gives an overview of the numerous initiatives supporting international data sharing using the standard; it presents the implementation of the standard from reception to dissemination at the European Central Bank; and introduces the future developments in the SDMX standard.

Keywords: SDMX; international data sharing; data structure definitions; web services.

1. Introduction

The "Statistical Data and Metadata Exchange" was established in 2001 by seven International Sponsor Organisations - the Bank for International Settlements (BIS), the European Central Bank (ECB), Eurostat, the International Monetary Fund (IMF), the Organisation for Economic Cooperation and Development (OECD), the World Bank and the United Nations. The SDMX international cooperation initiative is aimed at developing, implementing and executing more efficient processes for the exchange and sharing of statistical data and metadata among international organisations and their member countries. SDMX is the technical standard used as foundation for infrastructures serving official statistics.

SDMX was created to aid official statistics. "Official statistics" consist of data collected, produced and disseminated by governments and international organisations as a public good, according to defined legal requirements or other types of agreements. This data provides quantitative or qualitative information in a variety of fields, ranging from economic, demographic, social, environmental, and serves the government, the economy and the public at large for policy decisions and research. The data is "official" because it comes with the reputation of the world's governments and international institutions publishing it. One of the most important quality criteria for official statistics which will assure data reliability and comparability is that these statistics are compiled based on approved international standards. As such SDMX has established itself as the international standard in the technical aspects of official statistics data exchange.

We live in a world that is increasingly interconnected and the exchange and sharing of data is becoming easier, more frequent and important. SMDX is a key enabler in this context: its world-wide adoption makes it possible for compilers of statistics to gain in data quality and to international

organisations to increase the timeliness of publications and at the same time minimise the risks of errors in the transmission process. Furthermore, the SDMX strong emphasis on metadata guarantees the information is immediately understandable and useful.
Surely, more is expected from the standard in the years to come, in this respect the Sponsors Organisations are committed to sustain the evolution of the standard. The SDMX 2020 roadmap is the concrete example of this support. With ideas clearly specified to cover the next improvements cycle and a vibrant community actively discussing implementation steps, the roadmap is the next investment the SDMX community needs.

2. The main elements of SDMX

SDMX consists of an information model with content-oriented guidelines, standard formats, and IT architecture for data exchange.

2.1. The SDMX information model and its content-oriented guidelines

An information model is a framework to organise and classify any resource in the reality that is being modelled. It is a representation of concepts, relationships and constraints applicable to a specified domain. From a SDMX perspective, the information model is a set of classes (and interrelations among them) defining things like data- and metadata- sets, structures, flows, organisations and their specific roles. Those classes are agnostic as to a specific implementation and are used to model the statistical data, the associated metadata and the data exchange processes.
The data and related metadata for a particular statistical domain are structured according to a "Data Structure Definition" (DSD). A DSD describes the structure of a particular statistical data flow through a list of main concepts called dimensions, attributes and their associated code lists. SDMX also allows for additional explanatory metadata, which are often referred to as reference metadata. Reference metadata are generally in a textual format, for example describing the content, methodology and quality of the data. The reference metadata are structured according to a "Metadata Structure Definition" (MSD).
Each DSD and MSD is accompanied by content-oriented guidelines[1] which are recommendations for categorising and describing data. The content-oriented guidelines are intended to be applicable to all statistical domains. The guidelines focus on harmonising specific concepts and terminology that are common to a large number of statistical domains.

2.2. Technical foundations of SDMX - standard formats and architecture of data exchange

The SDMX standard includes data exchange formats based on the XML syntax (SDMX-ML) and EDIFACT syntax (SDMX-EDI). The first SDMX message format was the GEneric Statistical MESsage for Time Series GESMES/TS, which was introduced in the 1990s. With time the XML syntax took over as it had the advantage of being a widely-used open standard, which can be processed with a various range of IT applications, including free and/or open-source software.
In addition to the robust exchange format, the SDMX technical specifications, currently version 2.1, contain web-services recommendations, expanded to include a RESTful interface, standard functions and error messages. In this way now, it is possible to develop generically interoperable applications based on the SDMX standards.
SDMX supports two complementary modes for data sharing: the “push” mode (where data are transmitted from one organisation to another) and the “pull” mode (where one organisation retrieves

[1] More information on the SDMX Content-Oriented Guidelines can be found on the official SDMX website- https://sdmx.org/?page_id=4345

data from another organisation's server). SDMX also supports the "hub" concept, where users obtain data from a central hub, which itself automatically assembles the required dataset by querying other data sources.
The SDMX IT architecture also comprises specifications for registries, i.e. ways to realise organised repositories for metadata used to implement data sharing models based on easy location discovery of data and metadata and the ways this data can be accessed. A SDMX registry can be seen as an index of a distributed database and is essentially an application, central in the messaging architecture. Over the internet this application is accessible by programs that can accept query messages, returning as results the locations of SDMX-compliant data as well as structural and reference metadata. SDMX has developed specific registry standards in order to enable statistical organisations to perform efficient data and metadata sharing.

3. SDMX and its role in official statistics

SDMX follows a not-for-profit policy. All the tools produced are made publicly available free of charge. With this mandate, the Sponsor Organisations promoted the usage of SDMX and launched activities aimed at facilitating the SDMX adoption world-wide. A milestone in the acceptance of SDMX in official statistics was reached in 2008 when the United Nations Statistical Commission and the Committee for the Coordination of Statistical Activities (CCSA) adopted SDMX as the standard for data and metadata exchange.
SDMX now is used by a wide range of international data exchange activities and is adopted around the world. Many national and international modernisation initiatives use SDMX to improve and standardise data and metadata exchange and dissemination, to make data usage easier, to maintain data quality and to reduce production and dissemination costs. This progress has been acknowledged by the international community, for instance by the Data Gaps Initiative endorsed by the G20.
It should be stressed that the creation of global DSDs played an important role in pursuing the worldwide interoperability in the field of official statistics. Global DSDs are a necessary condition for optimal business processes. Undoubtedly, this is a major step forward in modernising official statistics. After several years of development in 2014 the first global DSDs in the field of national accounts and balance of payments went live. Since then, global SDMX structures are being developed in a number of other statistical domains.
Another milestone was reached with the SDMX Global Registry, which provides the SDMX community with the authoritative source for SDMX material. Thanks to the Global Registry, national and international agencies that wish to use SDMX as standard for data management or data exchange, have an easy way to access global DSDs, cross-domain concepts and code lists.
Significant progress in applying SDMX was achieved in the European Union (EU) by both the European Statistical System and the European System of Central Banks who apply SDMX on a wide scale in their day-to-day work. For example currently all statistical data exchange between ECB and Eurostat is based on the SDMX model. In addition all official statistics data exchange between ECB and EU member states is also using the SDMX information model and its technical specification. But there are many further visible outputs. One of them is the Principal Global Indicator website hosted by the IMF and sponsored by the Inter-Agency Group on Economic and Financial Statistics (IAG) composed of the same seven international organisations as the SDMX Sponsors. The PGI Website is providing comparable data mainly for the G-20 economies. Equally, SDMX takes its place in the Data Gaps Initiative endorsed by the G-20 Finance Ministers and Central Bank Governors.
After the successful implementation of the first global DSDs, the IAG established the Task Force on International Data Cooperation (TF-IDC). TF-IDC's objective is to determine the procedures that could be applied for a successful data cooperation arrangement across international agencies that would streamline and improve the efficiency of data collection, sharing, and dissemination. Its first pilot went live in mid-July 2015 and included the exchange of national accounts aggregates. The

outcome of this pilot showed that SDMX is a powerful technical tool essential for efficient real- time data exchange. More pilots of the task force are planned for the future.

4. SDMX at the European Central Bank, from reception to dissemination

In order to undertake the tasks of the European Systems of Central Banks, the ECB collects the necessary statistical information either from the competent national authorities or directly from economic agents. When doing so, the ECB tries to harmonise, where necessary, the rules and practices governing the collection, compilation and distribution of statistics in the areas within its field of competence. For example the ECB tries to use the established cross-domain code lists from the global DSDs as much as possible in all data collections and disseminations. For instance the sectoral code list used in the global DSD in national accounts is re-used in other data collections defined after the global DSD implementation where institutional sector breakdown is required. This way only very few sectoral classification are maintained in the system, reducing maintenance resources not only internally but also externally (for national authorities sending data and final users using the data).
Another important consideration is that the statistical infrastructure in place at the ECB uses the SDMX standard in all of its base elements, from reception, through production and finally to dissemination. SDMX artefacts are embedded in all the main parts of the statistical process and represent the building blocks used to create statistical IT solutions. The ECB, being a relatively young organisation, implemented IT solutions starting with a green field approach and evolving them alongside the development of SDMX, hence achieving great efficiency with maximum automation. Firstly, the SDMX information model guides the representation of data used in house and with partners. Secondly, all exchanges of data files are using the SDMX technical specification. Thirdly, all data produced is compliant with the SDMX structures used in the data exchange. Last but not least, the internet based dissemination and data-sharing activities are centred on the concept of registries and make use of the SDMX format for the message codifications.
The common denominator, and most distinctive feature, of the approach above is the philosophy followed when defining and processing data structures: the same data structure definitions are to be used unchanged from reception to dissemination with no intermediate mappings at any stage of the statistical process. This paradigm brings in numerous benefits; apart from minimising the maintenance work and guaranteeing better performance, it facilitates the activities of end users because keeping the same data structure format when they work, allows to integrate easily data from different sources, limiting the amount of mapping required to get data comparable.
The focus on the user dimension permeates the definition of interfaces to access data. A large number of applications are already in production for that purpose and provide important benefits to users through the availability of pull technologies. External users of ECB statistics can automate their processes and “take” from the ECB web site data available in various SDMX formats depending on their needs. An example is the Principle Global Indicator table’s website, a result from the G20 data gap initiative, and maintained by IMF. Both the euro area and euro area country data on financial transactions are retrieved by IMF in a fully automated manner by pulling the data from ECB’s Statistical Data Warehouse (SDW) using the available SDMX web services. In addition, the data disseminated by SDW is using the Global DSDs (for national accounts and balance of payments, see the DSD example in table 1) which makes the exercise even easier for the pulling agent as no additional conversions are required compared to the established data exchange DSDs. This way an institution could use the same IT system to process statistics coded with the Global DSDs and originating from different institutions. In addition, in the Eurosystem context, the approach of using the same DSDs reduces the reporting burden not only for ECB side but also for the euro area countries sending data to their central bank. More work is required to have a wider adoption of global DSDs, and this is a priority for the Sponsors organisations, to bring the full benefit to end user of a harmonised view on statistics.

The table below offers a schematic view on the Global DSDs in the field of National Accounts and Balance of Payments (BOP) statistics used at the ECB.

Conecpt/Dimension	National Accounts			BOP	BOP/NA
	Main Aggregates	Sectors and Government	All		
Frequency	FREQ	FREQ	FREQ	FREQ	Shared
Seasonal Adjustment	ADJ.	ADJ.	ADJ.	ADJ.	Shared
Reference Area	REF AREA	REF AREA	REF AREA	REF AREA	Shared
Counterpart Area	COUNT AREA	COUNT AREA	COUNT AREA	COUNT AREA	Shared
Reference Sector	REF SECTOR	REF SECTOR	REF SECTOR	REF SECTOR	Shared
Counterpart Sector	COUNT. SEC	COUNT. SEC	COUNT. SEC	COUNT. SEC	Shared
Consolidation		CONSOLIDATION	CONSOLIDATION		
Flow Stock				FLOW_STOCK	
Accounting Entry	ACC. ENTRY	ACC. ENTRY	ACC. ENTRY	ACC. ENTRY	Shared
Stocks/Transactions and other changes	STO	STO	STO		
				IAI	
Functional Category				FUNC CAT	
Instrument / Asset		INSTRUMENT_ASSET	INSTRUMENT_ASSET	INSTRUMENT_ASSET	Shared
Maturity (Original/Remaining)		MATURITY	MATURITY	MATURITY	Shared
Activy	ACTIVITY		ACTIVITY		
Activity (Destination)			ACTIVITY_TO		
Product			PRODUCT		
Product (Destination)			PRODUCT_TO		
Expenditure	Expenditure COICOP	Expenditure COFOG	EXPENDITURE		
Pension Manager			PENSION_MANAGER		
Pension Claimant			PENSION_CLAIMANT		
Pension Fundtype			PENSION_FUNDTYPE		
Unit Measure	UNIT MEAS.	UNIT MEAS.	UNIT MEAS.	UNIT MEAS.	Shared
Currency / Denomination	CURR DENOM	CURR DENOM	CURR DENOM	CURR DENOM	Shared
Valuation	VAL	VAL	VAL	VAL	Shared
Prices	PRICES	PRICES	PRICES		
Transformation (Time)	TRANSFORMATION	TRANSFORMATION	TRANSFORMATION		
Custom Breakdown		CUST_BREAKDOWN	CUST_BREAKDOWN		
Compilation method				COMP. METH	

The large number of dimensions captures more complicated data. In addition, there is a large degree of consistency between the NA and BOP DSDs. This assures better comparability of the data by final users and the use of common tools for implementation of the DSDs. Less maintenance burden due to shared concepts and dimensions.

More examples of applications that automatically retrieve data from the ECB using the SDMX web services are the various interactive dashboards presented on the ECB website that give overview on the latest monetary policy and financial stability developments. Another implementation is the ECB's mobile application that allows you to access and display statistical data disseminated by ECB. The app uses SDMX 2.1 web services to retrieve data from SDW and offers data visualisations predefined in the ECB SDMX registry on topics such as key euro area indicators, exchange rates, prices, government finance, monetary aggregates and bank interest rates.

The ECB and its internal users are looking forward to more implementations of the SDMX standards by statistical organisations from all over the world. This would very much facilitate the automatic downloading of detailed official statistics for economic analysis and research by all the community of users of statistics.

5. SDMX 2020- vision for the future

In 2011, the SDMX sponsors published an action plan for the years 2011 to 2015. This plan was instrumental in channelling development efforts and providing SDMX users with a clear view of how the standard was developing. Key strands of the action plan included setting up a Global SDMX Registry to share metadata assets, implementing Data Structure Definitions (DSD) for local and global use in macro-economic statistics, improving Content-Oriented Guidelines, producing a new generation of IT tools and rolling out new training services.

SDMX has already achieved an optimal level of maturity and it is widely recognised as a global standard for data exchange. Building on the positive experience of the previous action plan and in response to the needs expressed by users, the SDMX sponsors have decided to develop a "Roadmap 2020" presenting a vision of where SDMX is heading in the next few years.

One of the main strategic objectives in the roadmap is to make data usage easier via SDMX (especially for policy use). Although SDMX was originally focused on the exchange of aggregated time series data between organisations, SDMX has gradually widened its scope to become a full metadata standard, also usable for other types of data. In this respect it is important to observe the increasing importance of micro-data in statistics, and assess how SDMX could fit this space. One of the foreseen strategic goals of the SDMX 2020 roadmap is to work on more compatibility of SDMX with other standards that are used for micro-data, such as XBRL, widely adopted for individual business reports, and DDI, which is a reference metadata standard used for describing questionnaires and other types of statistical micro-data. The importance here is not to talk about “competing standards” but instead to realise that various standards can co-exist, and that it is left to the acumen of the technical solutions to ingest in their information models data represented in alternative manner.

Another objective for the future is the increased support of SDMX adoption in more domains. In this respect work is undergoing to create more global DSDs in areas of statistics currently not covered by the standard. At the same time, efforts are spent in promoting infrastructural packages to facilitate technology adoption; ensuring minimum duplication from overlapping programs and a pooling of efforts from the outset; providing cross mappings and other models of co-ordinated actions.

Not to be underestimated is also the effort being put to create a standard Validation and Transformation Language (VTL) to enable in a tool agnostic way the definition of business rules to be applied to data.

The initiatives above are only a small portion of the future work aimed at enhancing the SDMX standard and promoting its usage. More details and the full list of activities under Roadmap 2020 can be found on the official SDMX website[2].

6. Conclusions

SDMX now supports a wide range of international data exchange activities and is widely used around the world to exchange official statistics. Many national and international modernisation initiatives use SDMX to improve and standardise data and metadata exchange and dissemination, to make data usage easier, to maintain data quality and to save production and dissemination costs. The action plan for further development of the SDMX standard introduces new strategic objectives that aim to make SDMX one of the leading information models used in official statistics.

[2] https://sdmx.org/wp-content/uploads/SDMX-actions-summary-web_2017_jan_13.pdf