Data exchange in business statistics (part 2): SDMX

Preface

This article provides a general outline of the **Statistical Data and Metadata eXchange** (**SDMX**) initiative, which is being implemented across a wide range of statistical domains. It explains what SDMX is, why it matters, and who is using it. The SDMX initiative and the various implementation projects are explained in comprehensive detail in the [SDMX section of Eurostat's website](), which also provides several tutorials.

Since it is a general introduction, this article does not go into detail about each of the tools that support SDMX exchanges or technical issues. However, it does provide links to other, more detailed sources of information on these tools.

The article is part of the online [European Business Statistics Manual,]() a comprehensive guide to methodologies and how business statistics are produced within the European Statistical System (ESS)

Contents

1. What is SDMX?

The internet and the world-wide web has made the electronic exchange and sharing of data easier and more frequent. However, exchanges often take place in an ad-hoc manner, using all kinds of formats and non-standard concepts. This creates the need for common standards, guidelines and tools to enable more efficient processes for exchanging and sharing statistical data and metadata [1].

Step forward SDMX!

Statistical Data and Metadata eXchange (SDMX) is the name given by seven international sponsor organisations (including Eurostat) to an initiative designed *to manage and automate the process of data and metadata exchange*.

---

[1] Metadata are data that define and describe other data and processes. Metadata can either be of a 'structural' form — to identify, use and process data matrices and data cubes — or 'reference' form — describing the contents and quality of statistical data.

SDMX is a standard (indeed an [ISO standard, 17369:2013](#)) designed to describe statistical data and metadata, to normalise their exchange, and to enable them to be shared more efficiently among organisations.

To meet these three requirements, SDMX has three key components:

i) a model — the **Information Model** — to describe data and metadata

ii) a standard for automated communication (called the **Content-Oriented Guidelines**)

iii) an **IT architecture** and set of tools for data and metadata exchange.

SDMX is a model

The SDMX Information Model forms the core of SDMX. It describes statistics in a standard way, it identifies objects and their relationships and it allows central management and standard access. In other words, statistical data, metadata and the data exchange process are all modelled.

How so?

Data represent concrete observations of a particular statistical phenomenon at a given moment. A data set is a collection of related observations, organised according to a predefined structure. In themselves, data are meaningless unless accompanied by a description. For instance, what does 2347 mean? It means nothing without concept descriptors that explain its meaning! These descriptors can be modelled according to whether they are:

- *dimensions* — identifying and describing the data

- *attributes* — providing additional information about the data, such as whether they are estimates

- *measures* — representing the phenomenon to be measured.

These structural descriptors are brought together in something called a **Data Structure Definition** (**DSD**). The DSD identifies the *dimensions*, *attributes* and *measures* in a data set, associates them with common code lists and is integrated within concept schemes.

In addition to the structural descriptions of data sets, there are also reference metadata. These refer to information about quality descriptions, process descriptions, methodological descriptions and administrative descriptions. Reference metadata are described in a standard way using the **Metadata Structure Definition** (**MSD**).

SDMX is a set of guidelines

The **Content-Oriented Guidelines** (**COGs**) are a set of recommendations within the scope of the SDMX standard that are designed to maximise interoperability. They are intended to be applicable to all statistical domains.

The COGs focus on harmonising specific concepts and terminology that are common to a large number of statistical domains. Such harmonisation helps achieve an even more efficient exchange of comparable data and metadata, and builds on existing experience from implementation.

SDMX standards thus provide essential support to statisticians: they maximise the amount of information through to users, enable the process to be automated, and allow web-service queries.

The COGs cover *cross-domain concepts*, *code lists*, *subject-matter domains*, a *glossary*, and *implementation-specific guidelines*.

Cross-domain concepts in SDMX describe concepts relevant to many, if not all, statistical domains. SDMX recommends using these concepts whenever feasible in SDMX structures and messages to promote the reuse and exchange of statistical information and related metadata between organisations. Examples of concepts include 'Reference area', 'Statistical Unit' and 'Time Period'. Each concept is described in a standard way with an ID, description, context and presentation.

Code lists are predefined sets of terms from which some statistical coded concepts take their values. SDMX cross-domain code lists are used to support cross-domain concepts.

A *statistical subject-matter domain* refers to a statistical activity that has common characteristics with respect to variables, concepts and methodologies for data collection and the whole statistical data compilation process.

The SDMX *Glossary* contains concepts and related definitions used in structural and reference metadata of international organisations and national data-producing agencies. It recommends using a common terminology to facilitate communication and understanding. The overall message of the SDMX Glossary is: if a term is used, then its precise meaning should correspond to the Glossary definition.

SDMX is an IT architecture and set of tools for data and metadata exchange

To support more automated, efficient exchanges of data and metadata, standard tools and an IT architecture are required. In practice, this means that SDMX:

- promotes the use of standard SDMX-compliant formats (such as .xml)

- provides the necessary tools to support the Model, to create SDMX-compliant files, to store SDMX-related artefacts, to map and transcode from existing databases, and to validate the structure (and in future the content) of data files

- provides the necessary architecture to connect IT systems to the SDMX world, enabling data to be shared more easily.

The tools that support the SDMX exchanges fall into three broad groups:

1. tools for 'data receivers', covering the Data Structure Wizard (DSW) tool that creates SDMX artefacts and the Euro SDMX Registry that stores those SDMX artefacts

2. tools for 'data and metadata providers', covering the SDMX Converter that converts files to and from SDMX-ML files, the SDMX Reference Infrastructure (SDMX-RI) that creates and disseminates SDMX_ML files directly from databases using a set of pick-and-choose building blocks and tools, such as the Mapping Assistant, and the ESS Metadata Handler (ESS-MH) that processes reference metadata, and

3. tools for the 'IT developers' in SDMX, covering the SDMX source that is the source code for SDMX, the SDMX-RI web service that disseminates SDMX-ML data, and the SDMX Converter API that converts files from your own code. These tools are detailed in a series of articles on the SDMX Info Space under [SDMX IT tools](#).

Choosing the relevant tools to use in an SDMX implementation is a business choice made by the statistical production unit together with the relevant organisations in the EU countries that weighs the possible benefits of implementation against the necessary investment. As each SDMX implementation is different, and as the tools are frequently updated (to provide new functionalities, for example), Eurostat provides support to its partners.

Myth-busting: SDMX is therefore much, much more than just a data transmission format!

2. Why is SDMX used?

The use of SDMX is a **business** choice, as opposed to a technical one.

What makes it a business choice?

Firstly, decision makers need to understand background issues. **Supply-side** issues associated with the exchange of statistical data and metadata include the following:

- such exchange is complex, resource-intensive and expensive, with data being collected in multiple ways and transmitted in various formats, across various media.

- multiple organisations can collect similar or the same data.

- similar concepts can have a different content.

- the manual nature of data collection can lead to errors and inconsistencies.

There are also **demand-driven** issues: there is an increasing demand for data, faster and more frequent exchanges and a growing number of types of information exchange.

Secondly, decision makers need to know what advantages SDMX can offer. Here is some help:

- SDMX improves t**imeliness**, with faster access to data and metadata and the possibility of automated exchanges.

- SDMX improves **accessibility**, with bilateral, gateway and data-sharing possibilities.

- SDMX improves **interpretability**, with standardised structural metadata (the identifiers and descriptors of data) and reference metadata (the content and quality of data).

- SDMX improves **coherence**, by using standard cross domain-concepts, shared code lists and standard guidelines which are reused across statistical domains and agencies, and can support single figure dissemination.

- SDMX can **reduce data errors**, through automated structural and content validation, agreed structures for transmission, and **save time** on conversion and mapping, with less manual intervention.

- SDMX can **reduce the reporting burden** on agencies through the use of pre-validated content, common formats, automated publication, and the possible 'pull' of data by collecting agencies.

SDMX can **cut the costs of IT development and maintenance** by using open-source software, eliminating licensing costs, having a shared toolbox and improving interoperability between systems and applications.

> In short, SDMX responds to a business need, it improves the quality in exchanges, is an international standard and offers cost-efficiencies.

3. Who uses SDMX?

SDMX was launched for exchanges of official statistics among international organisations (such as central banks and statistical agencies) and their member countries (particularly government departments). However, SDMX is also now being used by organisations outside the world of official statistics. In theory, it may be of interest to any organisation that collects, processes, analyses and disseminates statistical data and metadata.

There are generally three roles and teams in an SDMX project:

1. initiators, represented by a business unit that has a business case for an SDMX project

2. facilitators, represented by IT units that are involved with either developing tools or establishing the necessary IT architecture, and, in international organisations, also SDMX specialists who develop the DSDs or MSDs

3. implementers, who provide the data (implementing the project to provide SDMX-compliant files).

An SDMX project typically brings together statisticians, economists, methodologists, and experts in dissemination and information technology. This is why it is vital for people to 'speak the same language', i.e. use shared standards and a shared vocabulary.

4. How is SDMX implemented?

An SDMX project follows successive project management steps:

1. preparation

2. compliance

3. implementation

4. production.

There is a working checklist of the steps required in an SDMX project, which are summarised

here.

The **preparation phase** of an SDMX project is arguably the most critical. This is the phase in which the initiators and facilitators determine the project's objectives, scope, expected benefits and outputs. It is the moment to specify needs, plan and organise. Some of the key questions to be asked are: Why do you want SDMX? What is the timetable? What risks are involved? What production systems, file formats and code lists are currently in use? What is the frequency of data flows? Who will be involved? By the end of this phase, the goals of the SDMX project should be clear. So should the timetable for implementation, a draft project plan and roles and responsibilities. The key decision is whether to go ahead with the project or not.

The **compliance phase** is arguably the most time-consuming. This is the phase in which the initiators and facilitators design the system and plan the sequence of the workflow. Steps are taken to analyse the current exchanges, decide what can be reused, define the concepts, define the DSD matrix and design supporting artefacts.

The **implementation phase** brings together facilitators and implementers. At this building stage, SDMX artefacts (particularly the DSDs) are made available (in something called the SDMX Registry), the appropriate IT infrastructure is established, pilot projects are conducted (testing and review), last changes made, the roll-out schedule agreed and support provided.

The **production phase** is the ultimate goal, when SDMX-compliant data and metadata can be used in exchanges. SDMX artefacts will continue to need regular maintenance, reflecting the need to be flexible to accommodate new coverage, new needs, new codes, etc.

So how do government departments and other implementing bodies in EU countries get started? The first steps are all about communication; Eurostat's business units and SDMX facilitation team share information with their counterparts through the appropriate Working Parties. Topics for discussion include the project's rationale and goals, and what SDMX tools to use for implementation.

After the planning stage, the real work for the national institutions starts! It begins with Eurostat providing them with a set of guidelines. In addition to the background information underpinning the project and various contact points, these guidelines provide information about the code lists, the DSD(s) and how to conduct the data (or metadata) transmission(s). A straightforward step-by-step guide provides instructions on how to design the input file format and download and use the appropriate SDMX tool.

Eurostat is there to help. The SDMX support email address is always provided, so counterparts in EU countries can ask questions in the knowledge that every effort is made to respond within 48 hours. Furthermore, with increasing experience at national level, local points of contact also emerge to help with implementation.

5. Where can you find out about the status of SDMX in statistical domains?

All statistical domains for which there is some level of SDMX implementation are described in the new SDMX section of Eurostat's website. A [table showing the relevant DSDs](), their version

number, their location and the agency responsible for their maintenance is already available.

6. See also

Overview of methodologies of European business statistics: [EBS manual](#)

Legal provisions related to Data exchange: SDMX can be found in the following [overview](#)

7. Further Eurostat information

SDMX within the European Statistical System (ESS) has its own section on Eurostat's website: http://ec.europa.eu/eurostat/web/sdmx-infospace/welcome

This replaces the previous web portal that was available and updated through until the end of 2016.

The new site has information about:

- SDMX tools: http://ec.europa.eu/eurostat/web/sdmx-infospace/sdmx-it-tools

- training schedules and tutorials on SDMX: http://ec.europa.eu/eurostat/web/sdmx-infospace/trainings-tutorials/trainings

- implementation projects: http://ec.europa.eu/eurostat/web/sdmx-infospace/sdmx-projects

- associated validation services: http://ec.europa.eu/eurostat/web/sdmx-infospace/validation-transformation

8. External links

SDMX is a global initiative. The sponsors have their own website: https://sdmx.org/

A specific checklist of steps to be taken in an SDMX implementation is provided by the SDMX.org: https://sdmx.org/?sdmx_news=checklist-for-sdmx-design-projects

SDMX is an ISO standard: http://www.iso.org/iso/catalogue_detail.htm?csnumber=52500

9. Contacts

Eurostat, Directorate B

Methodology; corporate statistical and IT services

SDMX support: ESTAT-SUPPORT-SDMX@ec.europa.eu