

**The Business Case for SDMX**

**Version 1.0 - November 2019**

Contents

[Introduction 3](#_Toc25230064)

[Typical use cases and scenarios for data and metadata exchange 4](#_Toc25230065)

[The six benefits of SDMX 6](#_Toc25230066)

[Benefit 1 : SDMX inspires trust 7](#_Toc25230067)

[Benefit 2 : SDMX promotes standardisation 8](#_Toc25230068)

[Benefit 3 : SDMX supports modernisation 10](#_Toc25230069)

[Benefit 4 : SDMX triggers more timely and better quality data 11](#_Toc25230070)

[Benefit 5 : SDMX reduces costs and reporting burden 12](#_Toc25230071)

[Benefit 6 : SDMX removes barriers to implementation 12](#_Toc25230072)

[Management buy-in 13](#_Toc25230073)

[Some challenges of SDMX 13](#_Toc25230074)

[References 14](#_Toc25230075)

[Annex 1 - Example of data discovery and visualisation using SDMX 14](#_Toc25230076)

© SDMX 2019

<https://sdmx.org>

# Introduction

The purpose of this document is to introduce the business case for SDMX[[1]](#footnote-1).

This document is intended for statistical organisations considering SDMX as a solution for modelling statistical data and metadata, and harmonising and automating their data and metadata exchanges.

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 six benefits of SDMX that a statistical organisation may expect from the implementation of SDMX.

As there is no such thing as a free lunch the document also points to some challenges that can be faced when implementing the standard.

The information presented in this document may be freely re-used, partially or wholly, in drafting the specific business case of an organisation.

SDMX is an integrated solution consisting of three main elements:

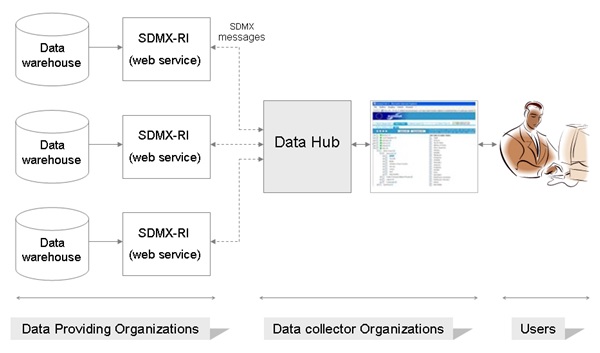
* technical standards (including the Information Model);
* statistical guidelines;
* an IT architecture and tools.

These three elements can be implemented in a stepwise approach. Depending on the scope of the SDMX implementation envisaged, project managers can thus use all or some of the arguments listed in this document.

# Typical use cases and scenarios for data and metadata exchange

SDMX can support various uses cases and implementation scenarios, as described below.

* **data and metadata reporting** (e.g. to international organisations)
  + SDMX is used as a "push" reporting format for data and metadata: the information provider generates a file and sends ("pushes") it to the receiver. The elementary processing activities cover extracting data from the source, writing a data set, and validating it (using the powerful Validation and Transformation Language VTL[[2]](#footnote-2)). The push model already offers automation possibilities: generalisation of SDMX file creation, creation of SDMX-ML files directly from the database, and conversion between formats using open source tools.
  + SDMX is used as a "pull" reporting format for data and metadata: the information provider opens a web service to the data, and the receiver organisation downloads ("pulls") the information at its best convenience; the elementary processing activities cover extracting data from the source, writing a data set and validating it, placing the file at a URL location, and publishing the existence of the data set by means of an SDMX registration.
  + Data hub architecture. This is a variant of the pull mode based architecture where the data is not previously collected and stored in a central repository but is directly accessed from the Member States' databases by the receiver upon end user request, as shown below.



* **reference metadata reporting**: the elementary processing activities cover extracting metadata from the source (database, file, spreadsheet), writing a metadata set, and validating it.
* **load data into a database**: the elementary processing activities cover reading an SDMX data set, validating it, and writing the data to the database.
* **load reference metadata into a database**: the elementary processing activities cover reading SDMX-ML metadata set, and writing the metadata to the database.
* **database administration**: SDMX can be used to automatically generate database tables. The elementary processing activities cover accessing a Data Structure Definition (DSD), creating database tables, and loading data into the database.
* **data dissemination**: SDMX offers standard interfaces for data and metadata dissemination
* **data discovery and visualisation** (GUI browsing and machine-actionable): SDMX offers a metadata-driven way of querying data sources using an open standard and can be used to drive website presentation of data and metadata. The elementary processing activities cover querying an SDMX structural repository, creating SDMX queries from the user selections, accessing a metadata repository to extract the metadata pertaining to the data, and transforming the data and their associated metadata into tables, graphs, charts, etc.

Annex 1 presents a real-life example of data discovery and visualisation based on an SDMX architecture.

* **enabling a database to be compatible with SDMX Web Services**: the elementary processing activities cover accepting and processing SDMX structure queries, accepting and processing SDMX data queries, and writing SDMX data sets.
* **data and metadata warehousing**: SDMX is used as a model for structuring a data warehouse or metadata repository.

# The six benefits of SDMX

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

* standardising statistical data and metadata content and structure across and within domains (e.g. similar concepts in wording can have a different content, different concept names and codes can refer to the same content, concepts and codes can be based on different structural rules)
* reducing the reporting burden for data providers (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 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 simpler, less labour-intensive, less manual, and cheaper
* reducing the maintenance costs linked to different data collection 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 of these concerns as will be demonstrated below.

This section presents various arguments that can be used for drafting a business case for SDMX. This list is not to be used "as is" because not all arguments will fit under all contexts. Furthermore, as stated earlier, there can be several steps in an SDMX implementation. Therefore, the number and type of arguments used will depend on the specific kind of implementation considered. Broadly speaking the list below should be considered as a list of ingredients for making recipes, with the number of ingredients used depending on the complexity of the recipe chosen.

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

1. SDMX inspires trust
2. SDMX promotes standardisation
3. SDMX supports modernisation
4. SDMX triggers more timely and better quality data
5. SDMX reduces costs and reporting burden
6. SDMX removes barriers to implementation

## 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 the major official statistical data producers, and they are collaborating closely with countries throughout the world. The SDMX sponsors are the Bank for International Settlements (BIS), the European Central Bank (ECB), the Statistical Office of the European Union (EUROSTAT), the International Monetary Fund (IMF), the Organisation for Economic Co-operation and Development (OECD), the United Nations (UN), and the World Bank.

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](https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx), the [African Charter on Statistics](https://au.int/en/documents/20121228/african-charter-statistics), the [Code of Good Practice in Statistics for Latin America and the Caribbean](https://www.cepal.org/deype/publicaciones/externas/6/47276/codigo-regional-buenas-practicasALC-ENG.pdf), the [European Statistics Code of Practice](http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-32-11-955), **foster trust** into 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 basically **driven by users' needs**. SDMX also provides a proven governance model to manage the life-cycle of a domain's SDMX artefacts.

SDMX is an [**ISO** standard](https://www.iso.org/standard/52500.html) (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 ensure 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 could successfully be developed collaboratively and then shared among agencies. Furthermore, international standards also facilitate international comparability of statistics.

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 an XML-based, easy to parse, internet-based language with extensive support for statistical metadata is clearly an advantage above sending data 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).

## Benefit 2 : SDMX promotes standardisation

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 syntax, 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;
* it provides a common terminology (concepts, code lists);
* it can describe data and metadata in any statistical domain;
* it makes tool development much easier;
* objects are fit-for-purpose: the relationships between objects and their usage are clearly defined;
* it is responsive to new technologies as new syntax representations can be constructed easily as the base of these is the information model;
* it serves as basis for building formats and tools
  + interoperable tools;
  + the information model is highly structured, i.e. it is easier to use a part of it rather than implementing the full standard.

**Standardising** statistical data and metadata content and structure has various advantages:

* speaking the same language;
* reusing existing material saves time and resources; reuse is facilitated by the existence of SDMX registries;
* facilitate mapping/data processing;
* shopping list of artefacts when defining structure;
* tools based on a commonly agreed format will have wider audience

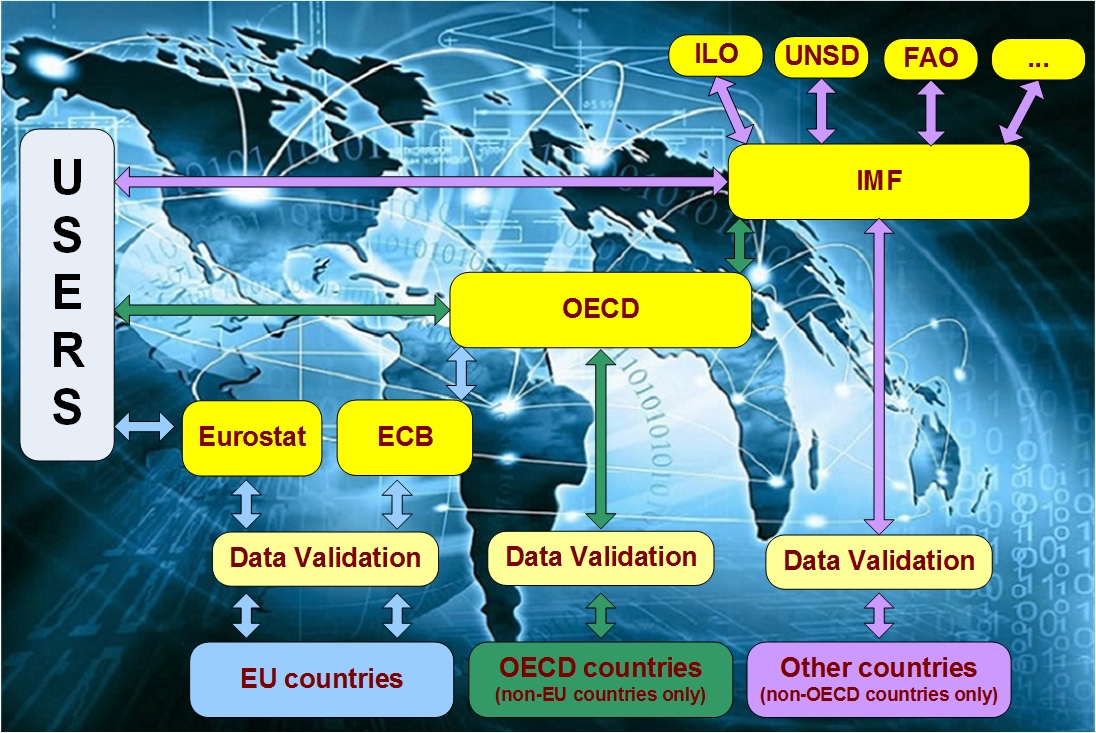
The **harmonisation** triggered by SDMX extends **beyond** the mere scope of **SDMX**, e.g. the SDMX Glossary is used beyond the SDMX sphere.

SDMX improves **interpretability** because:

* it standardises structural metadata (the identifiers and descriptors of data, such as table columns);
* it standardises reference metadata (the content and quality of data, such as the methodological principles applied in a survey);
* it standardises 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 to the development of a global statistical language (along the lines of the models developed under the auspices of UNECE: GSBPM, GSIM, GAMSO[[3]](#footnote-3), etc.).

SDMX improves **coherence**:

* through the use of cross-domain concepts;
* through the use of shared code lists;
* through the use of harmonised statistical guidelines (so-called "Content-oriented Guidelines");
* through reuse of SDMX objects ("artefacts" in SDMX language) across domains and agencies;
* through the ambition of a single figure dissemination (so-called "data sharing").



**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.

## 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](https://www.unece.org/stats/mos.html) (MOS) as promoted by the United Nations Economic Commission for Europe (UNECE). The processes 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.

## Benefit 4 : SDMX triggers more timely and better quality data

A standard technical architecture promotes **more timely and better quality data**:

* timely because less manual conversion is needed, and automated checks are fast, thus reducing "time to users";
* quality because automated processing and validation means less human error;
* faster access to data;
* move towards automation (automated workflows for exchange of statistics, thus reducing manual intervention errors; less "wait states"). Automation allows unattended workflow execution;
* promotes standard classifications which reduce mapping and transformation errors.

SDMX improves **accessibility**:

* bilateral, gateway and data sharing;
* push and pull modes.

SDMX can **reduce** data **errors**:

* automated validation;
* agreed structures for transmission;
* time saved on conversion, mapping;
* less manual intervention.

## Benefit 5 : SDMX reduces costs and reporting burden

SDMX can **reduce IT** development and maintenance **costs**:

* 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;
* no licensing costs;
* shared toolbox;
* improved interoperability between systems and applications;
* development burden is shared among the international community.

SDMX can **reduce** the reporting **burden**:

* the data message content can be pre-validated by SDMX data structures;
* the reporter can automate publication of SDMX via a web server;
* avoids the burden of maintaining many different reporting systems and exchange agreements;
* possible "pull" by collecting agencies.

## Benefit 6 : SDMX removes barriers to implementation

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 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

that 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 solve your problems.

# Management buy-in

It is essential that the decision to implement SDMX across an organisation is supported by senior management. SDMX is often perceived as IT-oriented and can be quite technical at the detailed level, so there is a risk that senior management may lose interest if they do not understand the business case in order to justify the funding of implementation. Therefore, managers and policy-makers should rather focus on the business aspects, especially on how SDMX facilitates data transfer faster and at a lower cost by using a common open standard and shared tools that can be reused by many.

The most efficient way to get senior management buy-in 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 not be based only on an abstract business case but also on some concrete experience.

It should however be noted that SDMX adoption is not always a must. Let us consider the case of a small organisation, possibly active out of the statistical sphere, that sends each year one single data set to an international organisation. In this specific case, does it really make sense to expect that they implement SDMX? This organisation could simply interact with SDMX using transparent tools such as web forms or converters.

# Some challenges of SDMX

SDMX is not investment free: it means analysis, it means training; it means changes. A pilot project, whatever its size, will incur costs. Training will also trigger expenses through the participation in training events, international conferences, etc. Migrating from a legacy system to an SDMX compliant system may mean a significant cost.

By definition change also implies reluctance to change. Changing business processes might even have a more significant cost and meet more reluctance than IT changes.

Standards are also costly to maintain in terms of improving fitness for purpose and evolving as business needs change.

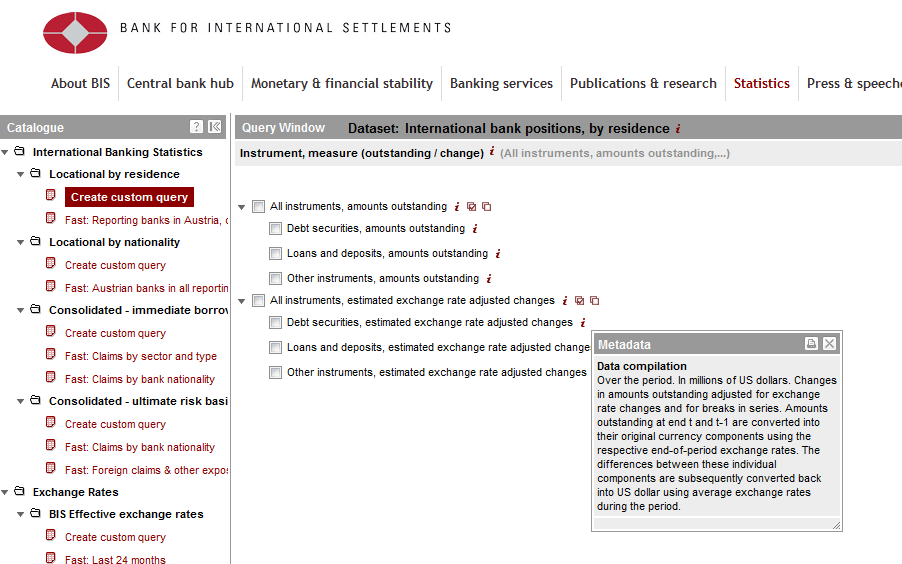
SDMX is dynamic: software versions are updated to increase functionalities and overcome bugs. However there is a danger that too much change may discourage adoption.

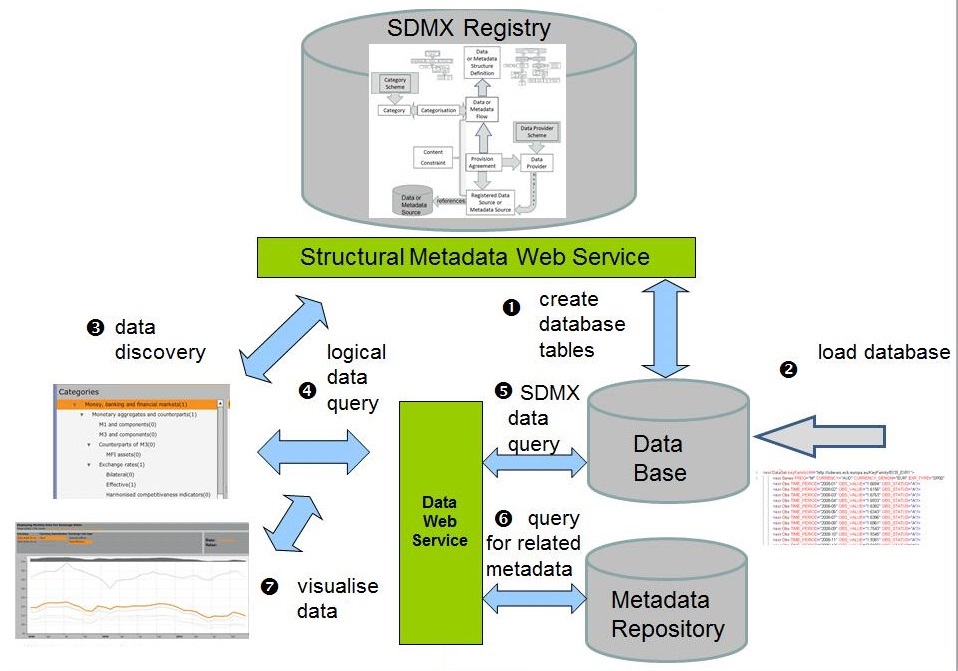
# References

* [Governance of commonly used SDMX metadata artefacts](https://sdmx.org/wp-content/uploads/Governance_of_commonly_used_SDMX_metadata_artefacts_v1.3.docx)
* [SDMX Glossary](https://sdmx.org/wp-content/uploads/SDMX_Glossary_Version_2_0_October_2018.docx)
* [SDMX “Learning” section](https://sdmx.org/?page_id=2555)
* [SDMX official website](https://sdmx.org/)
* [SDMX Roadmap 2020](https://sdmx.org/wp-content/uploads/SDMX_roadmap2020_FINAL.pdf)
* [SDMX Starter Kit for National Statistical Agencies](http://sdmx.org/wp-content/uploads/SDMX_Starter_Kit_Version_23-9-2015.pdf)
* [Validation and Transformation Language](https://sdmx.org/?page_id=5096)

# Annex 1 - Example of data discovery and visualisation using SDMX

This section shows an example of data and metadata discovery and visualisation using SDMX. The first picture shows a real-life visualisation of a query request, while the second picture depicts the SDMX architecture needed to achieve such a result. The last section describes the various process steps needed for such discovery and visualisation.





**Process description**

1. Retrieve the DSD from a structural metadata source (typically an SDMX Registry), and create database tables.
2. Read an SDMX data set file and load the data into the database.
3. Data discovery system uses structural metadata from the SDMX Registry to build the query interface (e.g. search form, navigation tree).
4. The query interface is adapted dynamically to user choices by querying the structural metadata.
5. The query interface creates an SDMX data query that is passed to a web service and the system returns an SDMX data set.
6. Reference metadata relevant to the dataset is retrieved from a metadata repository.
7. The data and metadata are passed to a visualisation tool to display the data in tables, charts, graphs, maps etc. Often a download is offered in various formats. The download options often include also the DSD or MSD.

1. A business case is a document describing the expected benefits for an organisation from the implementation of a new solution, such as a revised business process, a new organisation chart, or a new IT software. In other words, a business case captures the reasoning for initiating a project or task. It is intended to educate decision-makers and convince them to take some sort of action.

   A solid business case leads to well-informed decisions, while the one that fails to provide adequate information leads to decisions that can be very damageable to any organisation, not only in terms of time and money lost, but also in terms of making wrong decisions. It is a crucial communication tool that needs to be written in a language that the target audience understands, and with enough detail to facilitate proper decision-making. [↑](#footnote-ref-1)
2. [VTL](https://sdmx.org/?page_id=5096) is a standard language for defining validation and transformation rules (set of operators, their syntax and semantics) for any kind of statistical data. [↑](#footnote-ref-2)
3. GSBPM : Generic Statistical Business Process Model;GSIM : Generic Statistical Information Model; GAMSO : Generic Activity Model for Statistical Organisations. [↑](#footnote-ref-3)