

# A brief introduction to SDMX

## Statistical Data and Metadata Exchange

*This introduction is intended as a guide to readers who are unfamiliar with the technical details of SDMX and its related technology, but who need to use the standards, or make decisions concerning their use. It is intended especially for the staff and management of statistical organisations, such as national statistical offices (NSOs).*

*A lot of technical information on SDMX standards is already available on the Internet. This introduction will make reference to some pieces of this information. Some of the documents are, of necessity, of a highly technical nature which is more adapted for IT specialists.*

### **1. What is SDMX?**

The name “Statistical Data and Metadata Exchange”<sup>1</sup> refers to an international cooperation initiative aimed at developing and employing more efficient processes for exchange and sharing of statistical data and metadata among international organisations and their member countries. The initiative, started in 2001, is sponsored by 7 international organisations: Bank for International Settlements (BIS), European Central Bank (ECB), Eurostat, International Monetary Fund (IMF), Organisation for economic Co-operation and Development (OECD), United Nations (UN) and the World Bank (WB), who are committed to establish, implement and comply with common standards.

### **2. What are the goals of SDMX?**

The rationale of SDMX is standardisation for statistical data and metadata access and exchange.

With the ever increasing ease of use of the Internet, the electronic exchange and sharing of data is becoming more and more easy, frequent and important. This stresses the need for a set of common standards for exchange and sharing of

statistical data and metadata, and making processes more efficient. As statistical data exchange takes place continuously, the gains to be realised from adopting common standards are considerable both for data providers and users.

The objective is to establish a set of commonly recognised standards, adhered to by all players, making it possible not only to have easy access to statistical data, wherever these data may be, but also access to metadata that makes the data more meaningful and usable. The standards will allow national organisations to fulfil their responsibilities towards users and partners, including international organisations, in a very efficient way, among other things by using their general online databases to give access as soon as the data are released.

The SDMX standards also aim to ensure that appropriate metadata always come along with the data, making the information immediately understandable and useful. For this reason, standards for metadata exchange are extremely important in SDMX. At present, this part of the SDMX standards is only partially developed, but the plans for future editions of the standards comprise full development of metadata standards.

In conclusion, several quality dimensions can be improved through the use of SDMX standards, such as timeliness, accessibility, interpretability, coherence, as well as cost-efficiency.

---

<sup>1</sup> See [www.sdmx.org](http://www.sdmx.org)

### 3. Who can benefit from SDMX?

As mentioned above, the SDMX standards are designed for exchange or sharing of statistical information between two or more partners<sup>2</sup>. Evidently, the SDMX standards have been developed by the sponsors in order to accommodate the constituencies of the sponsoring organisations (national statistical offices, central banks, ministries, etc.). Within and across these constituencies, the standards are intended for reporting (or sharing) statistical data and metadata in the most efficient way. But the SDMX standards can also be used within a national system for transmitting or sharing statistical data and metadata and by private data providers (such as re-sellers of statistical databases). This is particularly interesting in countries with a federal structure or a fairly decentralised statistical system. In such cases, a close link can be established between the national system for data sharing and the international ones, allowing for additional efficiency gains for the involved organisations (for example, the Australian Bureau of Statistics is working on such an integration of systems).

Different forms of exchange can be accommodated, depending on the number of partners involved and the nature of the agreements between them, and on which of the parties is sending the data. Three kinds of exchange can be identified, according to the number of partners and agreements:

1. **Bilateral exchange:** All aspects of the exchange process are agreed between the partners, including the mechanism for exchange of data and metadata, the formats, the frequency or schedule, and the mode used for communications regarding the exchange. This is perhaps the most common process pattern.
2. **Gateway exchange:** Gateway exchange is an organised set of bilateral

exchanges, in which several data and metadata sending organisations or individuals agree to exchange the collected information with each other in a single, known format, and according to a single, known process. This pattern has the effect of reducing the burden of managing multiple bilateral exchanges (in data and metadata collection) across the sharing organisations/individuals. This is also a very common process pattern in the statistical area, where communities of institutions agree on ways to gain efficiencies within the scope of their collective responsibilities.

3. **Data-sharing exchange:** Organisations participate in making their data and metadata available to any organisation that has permission to access it. This requires adherence to certain data and metadata publication standards (such as the need to register the existence of the data and metadata). This model does not mandate a pre-defined agreement, but requires that data and metadata providers and consumers adhere to the standards. The model uses the “Pull” mode described below.

Messages can be exchanged in two different modes, the push mode and the pull mode:

1. **Push** mode means that the party who provides the data takes the necessary action to send the data to the party collecting the data. This can take place using different means, such as e-mail or file transfer, and in some cases the transfer can be supported by systems such as Eurostat’s Stadium and Statel. These are the “traditional” modes of data collection, carried out by international organisations for many years, and this is the method used in the bilateral and gateway exchange environments.
2. **Pull** mode implies that the data provider simply makes the data available via the Internet (this may be as simple as placing a structured (SDMX-ML) file on a

---

<sup>2</sup> This is further described in the document Framework for SDMX standards (version 1.0). See <http://www.sdmx.org/news/document.aspx?id=125&id=49>

website or it may involve accessing a database service, available via the web and capable of processing a standard SDMX query), and the data collector fetches it on his own initiative. In this case, more than one data collector may be allowed to take the pieces of data needed by each collector. This mode also resembles dissemination in the sense that access might be given to final users of information, who will then, according to their needs, access multiple web sites all using the same formats. The pull mode requires adherence to the standards demanded by the data sharing exchange.

While all combinations of the modes above are supported by SDMX standards, it is the aim of the SDMX initiative to further promote data sharing exchange using the pull mode.

#### **4. How can SDMX help with harmonisation of content?**

A major task of international organisations is obviously to agree on standards for compiling statistical data, to make automation feasible and international comparisons meaningful. This involves setting up standard classifications, common definitions of concepts, handbooks describing conceptual frameworks and guidelines for data collection, etc. In many cases this work is carried out in cooperation between several international organisations: for instance the System of National Accounts (SNA) is issued jointly by 5 organisations.

Once the methodologies have been defined, the next level of co-operation among organisations is to determine exactly which pieces of data need to be collected to meet their users' needs. There are a number of important statistical domains with data-sharing agreements at international level: for example education, environment, energy and transport statistics. In several cases, the data sharing is based on common questionnaires between the international organisations, supported by standardised glossaries and other methodological tools. These arrangements imply that each country only reports data to one

organisation, and the organisations subsequently share the original data received from the country ("input data") or the data validated by the international organisation ("output data").

The ultimate step would be to make it possible for countries to make statistical tables available on their web sites, with a breakdown able to satisfy all international organisations' needs, using a common standard. This last step is exactly what SDMX is aiming at<sup>3</sup>, and this is the reason why SDMX is seen by international organisations as the key strategy for developing data and metadata collection or sharing.

An important ingredient in such agreements would be the metadata to be exchanged with data, describing the characteristics of the latter. The SDMX standards are intended to cover the needs of both the data and the metadata using SDMX-ML or by implementing SDMX conformant web services to deliver the data and metadata in response to a request coming from user. To do that, it is necessary to define the data characteristics using a common terminology. Therefore, SDMX is working both to develop technical standards for data and metadata exchange, and to develop a common metadata vocabulary.

#### **5. What are the technical foundations of SDMX?**

The specifications of the SDMX standard formats build on the specifications of the GESMES (Generic Statistical Message) UN/EDIFACT standard and more specifically on the subset of GESMES named GESMES/TS (TS

---

<sup>3</sup> «...to explore common e-standards and ongoing standardisation activities that could allow us to gain efficiency and avoid duplication of effort in our own work and possibly for the work of others in the field of statistical information...», see SDMX "common statement" by participating institutions (source: <http://www.sdmx.org/about>).

## ***SDMX standards version 1.0: an overview***

A full information package on the SDMX version 1.0 standards can be found at <http://www.sdmx.org/news/document.aspx?id=125&nid=49>. The package contains the following standards:

**SDMX-EDI**, designed for the exchange of statistical information between organisations in batch mode. As the name indicates, it is an EDIFACT standard, and it is in fact identical to the GESMES/TS standard which has been used successfully for more than five years for exchange, especially in the central banking sector. There are two types of messages:

1. *Structure Definition*: A message describing the structure of the data and the structural metadata needed to understand and process a data set.
2. *Data/Attributes Message*: This is a message used for the exchange of the actual data and metadata (attributes).

**SDMX-ML** is the XML implementation of the GESMES/TS information model used in SDMX-EDI. This means that the structure is the same, and there exists a one-to-one translation between the two formats. However, SDMX-ML is designed to be used for a wider variety of exchange modes.

The standard contains the following elements:

1. *Structure Definition Message*: a common XML message expressing the structural metadata needed to understand and process a data set. It is completely congruent with the SDMX-EDI Structure Definition Message, and a conversion between the two formats is available.
2. *Full Data Message (or Generic Data Message)*: All statistical data expressible in SDMX-ML can be marked up according to this data format, in agreement with the contents of a Structure Definition message. It is designed for data provision where applications receiving the data may not have detailed understanding of the data set's structure before they obtain the data set itself.
3. *Compact Data Message*: A message optimised for the batch exchange of large amounts of time series. This format is specific to the agreed conventions for the subject matter area of the data set (the key family), and, unlike the above-mentioned Full Data Message, it can only be understood in connection with the metadata defined in the Structure Definition Message; this is because all of the data from the Structure Message are not repeated in the Compact Message. It allows for the transmission of partial data sets (incremental updates) as well as whole data sets. It is completely congruent with the SDMX-EDI data message, and a conversion between the two formats is available.
4. *Utility Data Message*: This message type, like the Compact Data message, is specific to the key family of the data set, but is designed to support validation and other expected XML schema functions.
5. *Cross-Sectional Data Message*: This message is similar to the Compact Data Message, but it allows for transferring data which are not organised strictly as time series but where there is more than one observation per time period.
6. *Query Message*: Data and metadata are often published in databases which are available on the web. Thus, it is necessary to have a standard query document which allows the databases to be queried, and return an SDMX-ML message. The Query document is an implementation of the SDMX Information Model for use in web services and database-driven applications, allowing for a standard request to be sent to data providers using these technologies.

for time series). This latter standard has been successful in standardising several statistical data flows. GESMES/TS has been incorporated into SDMX as SDMX-EDI and messages can be easily transformed between SDMX-EDI and the SDMX-ML formats, thus safeguarding investment in GESMES/TS.

The main SDMX formats (SDMX-ML) use XML syntax. In addition, the standards contain guidelines for the development of web services.

## 6. Implementations

As part of the SDMX initiative, a number of initial implementations are underway. These include:

**Joint External Debt Hub:** Early on in the SDMX work, external debt statistics were identified as an interesting area for piloting, as these statistics were already the subject of a joint venture between BIS, IMF, OECD and the World Bank. In 2003 a joint demonstration model was set up, showing how data could be shared among the four organisations using preliminary SDMX standards.

**Alignment projects:** These projects are designed to test the standards in “real-world” applications, and generate practical feedback to help guide their further development. They are organised by one or more of the sponsor organisations, and are adopted and described under the umbrella of SDMX.

1. One of the SDMX alignment projects, launched by the OECD, is NAWWE<sup>4</sup> – National Accounts World Wide Exchange. The idea behind the NAWWE project is to use a web based mechanism for reporting an already internationally agreed set of national accounts data. The objective is to allow any user, in particular international organisations, to access directly a set of internationally

comparable data made available by countries. If all the involved international organisations were to agree to use this mechanism, it would reduce the reporting burden on member countries and improve the accuracy, coherence and timeliness of the data. NAWWE is based on the common questionnaire agreed by the OECD and Eurostat for the collection of national accounts data and the tables are expressed using the SDMX-ML standard formats. At the moment three countries -- Australia, Canada and France -- have made sample data available in this format and the data are presented in the NAWWE web site. It is planned in 2005 to move to full production for countries willing to join in, meaning that these countries would no longer have to submit national accounts questionnaires to the OECD.

2. The SODI (SDMX Open Data Interchange) project, launched by Eurostat in 2004, has been designed to test the feasibility of simultaneous publication at national and EU level of short-term statistics from EU Member States, aimed particularly at making those short-term statistics available more quickly and making them easily accessible to users in formats which facilitate further use. The SODI project is especially concerned with “Principal European Economic Indicators” (PEEI), a set of statistical indicators which are used in the management of economic and monetary union in the EU. SODI will investigate various approaches, based on SDMX standards and tools, for streamlining the collection of these indicators. The data model and metadata standard will be based on SDMX. An important concern for SODI is to ensure the underlying harmonisation of the indicators. The first SODI pilots in 2005 will involve five EU Member States (Germany, France, Netherlands, Sweden and UK) and two indicators (quarterly GDP and the monthly industrial production index). There will be regular exchange of

---

<sup>4</sup> See <http://stats.oecd.org/nawwe/>

experience with the NAWWE project, in order to ensure the coherence of technical developments in the two projects.

3. Joint UN/OECD trade project (ComTrade): Following agreement to share responsibilities for collection of annual foreign trade data, the OECD and UN have additionally agreed to work jointly to establish a common system for managing annual foreign trade data, using an SQL-based data model designed by the UN Statistics Department. Development effort for data collection, validation, processing and management software is being shared by the OECD and the UN. Responsibilities for collecting and validating data will also be shared, with data periodically replicated from one site to the other. The OECD and UN agreed to use the SDMX-ML schema for replication of annual foreign trade data between the databases maintained at UN/ New York and OECD/ Paris.
4. Dissemination project of the Eurosystem: The European Central Banks (ECB), working together with the national central banks (NCBs) of the Monetary Union, has set up a project aiming at improving accessibility to national contributions to Euro area aggregates. The project foresees that national contributions to euro area aggregates shall be presented jointly with euro area aggregates and disseminated simultaneously on the ECB's and NCBs' websites (for those NCBs wishing to do so) so as to guarantee full statistical consistency from all entry points. Furthermore, the mechanism allows that data can be presented on each web site with the general look and feel of that website, including some translation into national languages. The workload for NCBs can be kept to a minimum while the maintenance costs for the statistics are mostly centralised at the ECB. A pilot has been conducted with several

NCBs of the Eurosystem to validate the concept.

5. The IMF Metadata Repositories Project intends to contribute to SDMX technical and content standards to facilitate the open exchange of reference metadata. The immediate objectives of the project are twofold: (1) identify the commonalities in the metadata structures of macroeconomic datasets that are collected and stored in existing repositories; (2) build on these commonalities to standardize format, structure, and vocabulary for the exchange of reference metadata. More specifically, the project will implement the SDMX exchange standards to the metadata repositories maintained by the IMF, allowing reference metadata to be received and updated from countries and make metadata available in SDMX-ML format. This work will contribute to the development of version 2.0 of SDMX standards, which is expected to be available for public review by end-Spring 2005.  
The project will lead to the development of a SDMX reference metadata information model (work in progress, available at [www.sdmx.org](http://www.sdmx.org)) that provides the framework for defining the structure of metadata to be reported for any subject matter domain. The Metadata Repositories Project will also provide the first short-list of "core statistical concepts" for the exchange of reference metadata under the SDMX. These core metadata concepts will need to be adapted to the needs of all SDMX partners and extended to suit the needs of specific exchanges.

## Metadata and terminology

Metadata standards are important components of SDMX. SDMX metadata standards build on the distinction between “structural” and “reference” metadata.

**Structural metadata** are those metadata acting as identifiers and descriptors of the data, such as names of variables or dimensions of statistical cubes. **Reference metadata** are metadata that describe the contents and the quality of the statistical data (*conceptual metadata*, describing the concepts used and their practical implementation, *methodological metadata*, describing methods used for the generation of the data, and *quality metadata*, describing the different quality dimensions of the resulting statistics (e.g. timeliness, accuracy).

The idea is that it should be possible, using the SDMX standards, to exchange or share the data and the metadata that will allow a thorough understanding and interpretation of the corresponding statistical data. In the first version of the standards, there is limited support for reference metadata (principally conceptual metadata and some methodological metadata that typically accompanies the data). More comprehensive documentation, notably methodological reference metadata describing definitions, sources, shortcomings, transformations, etc, will have to be exchanged outside these messages. Version 2 of SDMX to be released in 2005 will comprise a fuller set of standards for metadata, including common structures for reference metadata.

SDMX also supports common terminology to be used when exchanging and sharing data. For this purpose, SDMX comprises a Metadata Common Vocabulary (MCV), see <http://www.sdmx.org/knowledge/document.aspx?id=66>. The MCV contains a comprehensive set of definitions of metadata terms used for describing the statistics and processes used in their compilation by national agencies and international organisations.

## 7. How is SDMX an advancement over earlier attempts to standardise data flows?

There are several reasons for having confidence in the soundness of investing in SDMX:

1. The organisations sponsoring the initiative are leaders in the world of international exchange of statistical information, and they are strongly committed to make the standards work in practice. They are also committed to working together, sharing the same mechanisms wherever this is relevant.
2. The sponsors are eager to ensure that earlier investments in standardisation in this field are not wasted.
3. GESMES/TS has been the most successful attempt so far to standardise statistical data flows, even though it still has a long way to go to become *the* world standard for exchange. Central banks have employed it successfully on a large scale, and Eurostat is gradually migrating more and more data flows to GESMES. SDMX incorporates GESMES/TS (SDMX-EDI) as a standard when using the EDIFACT syntax, and the newer XML standard, SDMX-ML builds fully on the same data model as GESMES/TS. This means that there is a one-to-one correspondence between SDMX-EDI and SDMX-ML messages, and the one can be transformed into the other, using available tools.
4. XML is a technology standard which is easy to use and widely accepted, not only in the statistical world. Consequently, it is expected to be a clear winner, and investments in it are comparatively low risk.
5. All the involved international organisations, as well as many data providing organisations in member countries, are already working with XML-based data and, in some cases, web services, so it will be a familiar environment.
6. Tools for conversions and manipulation are supplied by the consortium.

## **8. Organisation and contacts**

As mentioned, the SDMX initiative is a consortium of 7 international organisations, the sponsors. The “ownership” of the project lies with the high level group called the Sponsors Group, working as a kind of Board. This group, which typically meets four times per year, sets the targets of the initiative and supervises its implementation. The Board’s chair rotates among sponsor organisations and is currently held by Enrico Giovannini, Chief Statistician of the OECD.

Daily work is monitored by a Secretariat Group, consisting of one person from each sponsor organisation, and chaired by Stuart Feder, BIS, who is the SDMX Project Manager. An enlarged secretariat group, the Staff Conference, meets regularly to oversee development, mainly using teleconferences. Technical meetings take place as necessary.

Each organisation has its own way of organising its cooperation and involvement of its constituency.