



## Economic and Social Council

Distr.: General  
1 June 2011

English only

---

### Economic Commission for Europe

#### Conference of European Statisticians

##### Fifty-ninth plenary session

Geneva, 14-16 June 2011

Item 3 of the provisional agenda

#### **Organization of data collection and sharing, and the management challenges for the implementation of Statistical Data and Metadata eXchange**

### **The experience of the European Central Bank with Statistical Data and Metadata eXchange: (more than) ten years benefiting from the use of the standards**

#### **Note by the European Central Bank**

##### *Summary*

The Statistical Data and Metadata eXchange (SDMX) standards are extensively used in the statistical data sharing activities of the European Central Bank (ECB). This paper discusses the ECB's more than 10 years SDMX experience and some critical success factors to observe in order to optimise the use of the standards and maximise expected benefits.

## I. Introduction

1. The Statistical Data and Metadata eXchange (SDMX) standards<sup>1</sup> are extensively used in the statistical data sharing activities of the European Central Bank (ECB). Moreover, they are exclusively used in all ECB's statistical data exchange with the 27 European Union (EU) national central banks, the Bank for International Settlements (BIS), Eurostat and (in data reporting to) the International Monetary Fund (IMF).
2. The SDMX information model is present in all layers of the relevant ECB's systems and in all its statistical processes including data collection, validation, production and dissemination, fostering efficient and automated statistical activities, supported by its stable format. Given the benefits of the SDMX standards, the ECB, jointly with the BIS, Eurostat, IMF, the Organisation for Economic Co-operation and Development (OECD), the United Nations (UN) and the World Bank actively sponsors and contributes to the maintenance of these standards and related tools.
3. This paper discusses the ECB's more than 10 years SDMX experience and some critical success factors to observe in order to optimise the use of the standards and maximise expected benefits.

## II. Business case and guiding principles

4. According to Alexandre Lamfalussy, former president of the European Monetary Institute (the predecessor to the ECB), "*nothing is more important for monetary policy than good statistics*". Peter Bull (former Director General of the ECB Statistics) provides an unique record of the statistical preparations for Economic and Monetary Union (EMU) in the period 1992-2003 in his book "The Development of Statistics for Economic and Monetary Union" (July 2004). In his foreword to this book, the late Mr. Eugenio Domingo Solans (member of the Executive Board of the ECB from June 1998 to May 2004) mentions "[Many themes run through this record of events.] *Among them is the importance of cooperation and coordination among statistical agencies, with a clear allocation of tasks to avoid wasted effort, unnecessary burden on reporting agents, and competing data.*" Indeed, as a new institution, the ECB acknowledged, already at the time of its establishment, the need to develop efficient mechanisms to exchange statistical data as an essential prerequisite in the statistical preparations context<sup>2</sup>. It was also clear to the ECB that international co-operation and co-ordination should belong into the guiding principles inter alia in the development of the statistical data and metadata exchange facility of the ECB and the European System of Central Banks (ESCB). As an example on this cooperation it could be mentioned that the design and maintenance of GEneric Statistical MESSage for Time Series (GESMES/TS), the predecessor of the SDMX standards, greatly benefited from the intensive co-operation with the BIS, Eurostat, the IMF and other key institutions from the central banking community and the national statistical institutes.
5. Developing standards for the exchange of data in a co-operative manner was the joint interest of the BIS and the ECB (European Monetary Institute at the time) to perform efficient statistical data exchanges within the central banking community. Moreover, the IMF's and Eurostat's commitment to benefit from previous standardisation work done at the time in the field of balance of payments statistics allowed the efforts of the four

---

<sup>1</sup> See <http://sdmx.org> for details.

<sup>2</sup> For more about today's ECB's statistics: <http://www.ecb.europa.eu/stats/intro/html/index.en.html>

institutions to be brought together and converge to the design of the GESMES/TS message. At a later stage, the maintenance of GESMES/TS was brought under the aegis of SDMX as SDMX-EDI format. Practically, the SDMX-EDI format has remained unchanged since its first design (and there is currently no identified need to modify it in future). The SDMX-ML started as an eXtensible Markup Language (XML) representation of SDMX-EDI. However, over time (and as it was not widely used during the first years as GESMES/TS) it benefited from significant improvements, taking advantage of more recent XML developments. The ECB continues using SDMX-EDI in all its batch data transfers (thus, together with the other central banks, continuing exploiting considerable benefits from the respective investments on this highly compact and efficient format), while it uses SDMX-ML in the web dissemination of statistics. Most of the statistics made available on the ECB website can also be downloaded in SDMX-ML.

6. This success story with the SDMX standards established a solid base for the further development of statistical data exchange, production and sharing of ECB's statistics, clearly facilitating automation, integration and optimised processes. As said, this success with SDMX was possible not only because of clear business case but also because of close cooperation with the main partner institutions (the EU national central banks), the European Commission, the BIS and the international central banking community and with other international organisations<sup>3</sup>. This argumentation is further elaborated in the next chapter.

7. In general, it could be suggested that the constant effort to optimise the use of limited statistical resources by a strong high level sponsorship is one of the key success factors in development and deployment of best statistical practices and standards. Given the significant efficiency gains constantly observed, the strong commitment to the continuing use of the SDMX standards by the ECB and the European System of Central Banks is expected to be maintained for the foreseeable future.

### **III. Key characteristics and requirements in development and use of Statistical Data and Metadata eXchange**

8. The research on a state-of-the-art message format started around 1997 involving the experts in data exchange and data modelling from the BIS, Eurostat, some national central banks (e.g. Banca d'Italia), taking into account work done in the European Statistical System by Eurostat and some national statistical institutes (NSIs). It was clear from the outset that the new model and message format should meet several requirements in order to serve the content and efficiency objectives at an affordable investment cost. The key characteristics and requirements identified at the time are still relevant:

(a) Multidimensional support: inclusion of dimensions depending on conceptually relevant data characteristics. Recognition of the "time" dimension in the model, so as to serve "time series" requirements;

(b) A message simple to understand and implement with a strong focus on "fixing" as many elements of the syntax as possible, without sacrificing the generic strength of the model. Simplicity facilitates implementation at partner institutes using by in-house resources and easily integrating with local database systems;

---

<sup>3</sup> It should be added for completeness that besides the very intensive use of the SDMX standards in the context, the ECB statistical work also conforms to the other international statistical standards. See more about today's main features of the ECB's statistics at: <http://www.ecb.europa.eu/stats/intro/html/index.en.html>

(c) Efficient, allowing the compact representation of long time series (e.g. of daily data) or the exchange of only new data or revisions to already reported data (“Updates and Revisions”); of course, full complete databases should also be possible to be described and exchanged;

(d) Rich metadata model, for providing qualitative information about data in a manner that would still allow automation and integration; rich metadata within statistical databases (or closely linked) serve more easily search engines in a data warehouse environment;

(e) Representation of data structures, concept definitions and code lists. These structural “metadata” elements allow for even more advanced automation and integration with local systems; but also describing the data structures without external documentation;

(f) Facilitate statistical harmonisation (e.g. statistical “dimensions” taking values from internationally or broadly accepted code lists), allowing for flexibility where statistical standardisation is still limited or no widely accepted code lists exist yet;

(g) Possibility for aligning with future developments. SDMX-ML supports a rich metadata model for querying and web services. The ECB supports extensively web dissemination of statistics in SDMX-ML<sup>4</sup>. Moreover, the ECB actively contributes to the development of free tools for the interactive visualisation of statistical data in SDMX/ML<sup>5</sup>.

#### **IV. The success of Statistical Data and Metadata eXchange in the European Central Bank and the European System of Central Banks**

9. The most important success factors in the use of SDMX within the ECB and the European System of Central Banks (ESCB) could be identified as follows:

(a) Clear business case, commitment and high level sponsorship in the ECB and partner institutions. The need to respond to extensive and constantly growing statistical requirements with limited resources, required automation and it was natural to deploy SDMX for efficient statistical data exchange facilitating integration with applications to the maximum extent;

(b) As a result, related processes have been optimised and single “entry” and “exit” points have been established. The data flows from reception to dissemination (focusing on aggregated data) life cycle have been optimised in a structured manner based on the SDMX model. The same holds for ECB Statistical Data Warehouse (ECB’s web dissemination tool for on-line statistics) and the web services it supports;

(c) An organisational unit with strong and combined statistical and IT expertise is used to maintain the system and the implementation (technical aspects, structural metadata management, liaison with external parties to ensure harmonisation and road maps when new data flows are considered, integration of SDMX applications with the statistical databases). The ECB regularly organises courses for central banks for developing the statistical and technical expertise essential for the proper understanding of the information model and support of the local SDMX applications;

(d) The simple (but powerful!) and stable nature of the data exchange model and message allowed to cover any statistical domain (mainly time series aggregated data) and to

---

<sup>4</sup> See <http://www.ecb.int/stats/services/sdmx/html/index.en.html>

<sup>5</sup> See <http://code.google.com/p/flex-cb/>

have sufficiently rich metadata for serving search engines and further processing by other tools (e.g. Online Analytical Processing - OLAP). Data files have been easy to generate and understand and (reading/writing) applications easy to develop and integrate. Data Structure Definitions (DSDs) are exchanged separately from the data. The stability of the standards has greatly reduced maintenance cost over the years;

(e) SDMX facilitates complete automation for data and metadata exchange (and web sharing), including the exchange of Data Structure Definitions (DSDs), code lists and definitions of statistical concepts;

(f) Sharing data with many institutions and partners (with SDMX capabilities) increases the benefits enjoyed. For example, when an institution has automated the production of SDMX messages and uses these to report the same data to the ECB, Eurostat and other institutions, the investment pays off even faster. This is already the case, for example, for balance of payments data: the EU national central banks - using their output database systems and through unique exporting applications - report their statistics to the ECB, Eurostat, IMF and the OECD in SDMX formats;

(g) As also earlier discussed, in the ECB, DSD development and maintenance is centralised in a unit combining strong statistical and technical expertise with co-ordinating power. More than 60 DSDs are in use covering all fields of economic statistics. The co-ordinating unit aims at ensuring maximum consistency, while being open to future requirements. It is also involved in European and international standardisation work, liaises with other institutions conducting similar work where relevant and aims at “keeping things pure and simple”. The ECB regularly distributes its “DSD database” (also in SDMX formats) to its partners and other interested institutions. Moreover, based on its experience in designing DSDs, together with other institutions (e.g. BIS, Eurostat, IMF, UN) it contributes to international and regional statistical harmonisation (e.g. balance of payments, national and financial accounts) and generally supports the idea of globally governed Data Structure Definitions and Metadata Structure Definitions (MSDs)<sup>6</sup>;

(h) Reliable and secure telecommunication means have been supporting full automation and monitoring tools allow for problem solving when problems occur; they also support, with adequate security, the exchange of both published and unpublished statistics;

(i) Free SDMX enabled software and SDMX tools have been developed by the ECB and also by other partners and institutions. The ECB has developed the “Checker” suite (for validating and converting data files) and the SDMX-Flex as a presentation tool. The ECB tools are appropriate and mainly targeted to support integration, as the use of intermediate tools (e.g. csv based) involving manual work is strongly discouraged;

(j) Wide acceptance: there is a constantly growing usage of SDMX with international support (e.g. a world-wide central banking community). E.g. apart from the ECB, also Eurostat, IMF, OECD already receive data reported in SDMX formats. The growing community makes the benefits from implementation more apparent;

(k) Consistent support and stability: a key success factor is to keep consistency in the implementation and a stable format. Otherwise, sudden changes would affect the costs-benefits balance of partner institutions. Change management in the SDMX implementations

---

<sup>6</sup> On MSDs, in particular, a global convergence towards one MSD for official statistics that would adequately cover statisticians’ and users’ needs, regardless of domain and institutional partner group, would be very beneficial. For example, the work done by Eurostat in the domain of “high level” metadata (i.e. Euro SDMX Metadata Structure - ESMS) could be a good candidate (or at least a reference basis), as it provides a very comprehensive coverage of high level metadata concepts.

of the ECB is done very carefully, in close co-operation with partner institutions and decisions are taken on a consensus basis.

## V. Observed limitations and some perceived weaknesses

10. SDMX limits applicability. The success of SDMX in the E(S)CB of providing a model for the fully automated exchange of aggregate and time-series oriented statistical data is also its main limitation for exchanging data with a different focus or modelling requirements. Extensive administrative information about individual transaction data (e.g. securities) or (one-off) questionnaire data, especially with answers to open questions, or commercial bills information does not easily or efficiently fit the SDMX model and might affect the way the applications are integrated and “mapped” to internal processing applications. Nevertheless, this limitation was clearly accepted from the outset for reasons of maintaining simplicity.

11. External distribution of large databases. Institutions processing to SDMX-ML may consider that SDMX-ML is verbose and sharing a large database would create heavy network traffic. In the ECB, however, this has not been a problem: first, other physical means (e.g. CDs, DVDs, file transfer) are used for sharing large databases; secondly, the ECB uses mainly the (very compact) SDMX-EDI format for sharing this type of large data files. The batch “large database” sharing process is often used by the ECB for “refreshing” partner institutions’ databases.

12. Not aimed at individual reporting agents and arithmetic checks within the file: As an example, a reporter sending a data file with balance sheet information to a national statistical institution would like to know up front whether the data provided are consistent and plausible, thus expecting built-in checking functionality inside the data file for evaluating the checks and for explaining errors. Both reporter and receiving institute might appreciate tools for showing a data file as a balance sheet on screen. None of this exists in SDMX, however, there are also other means to serve these requirements: it has been assumed (and it is already the case with the ECB systems) that the recipient “loads” the received data and (automatically) runs the checks against the target database. Thus, the results of the checks are eventually much richer, as they include more plausibility checks and, thus, eventually, offer the required (and superior) functionality and still in an automated manner.

13. SDMX is often perceived as complex. SDMX is clearly more complex than the “comma separated values” (CSV) format, but CSV could not represent metadata without using new syntactic conventions that would eventually lead to equivalent extra complexity. SDMX is aimed at being as simple as possible while providing also the necessary “metadata” oriented functionality (i.e. pros exceed weaknesses).

14. Complex user guide, too wide standards. Simple tutorials are available oriented to the basic scope and serve well all ECB partner implementers<sup>7</sup>. Beyond the basic functionalities (batch data exchange, SDMX-ML files on the web), the ECB also supports more advanced facilities like SDMX queries against the ECB statistical data warehouse, but, of course, the usage of advanced functionalities is on a voluntary basis.

15. Limited commercial use and limited support by commercial software providers. Alternatively to the commercial products, it has been envisaged from the outset that the development of open source and the use of free tools should be pursued. In general, formal support by commercial software is not strictly necessary given the purpose of SDMX

---

<sup>7</sup> For example, see <http://www.ecb.int/stats/services/sdmx/html/tutorial.en.html>

(integration with database systems in way that would fit the specific features of internal implementations), but also a large set of computer programs is already available and open source code is constantly developed. Moreover, database systems support easy integration with external data (e.g. writing or reading txt or xml files) and SDMX files can be seen as simple “text” or “xml” files. Nevertheless, it would be desirable to see more desktop tools be capable (with minimum configuration) to “pull” SDMX files from the web and display them almost automatically. This would further increase the user basis expanding it beyond the “application to application” power users and institutions.

## VI. Conclusions

16. This note describes the successful ECB and ESCB implementation of the SDMX standards and discusses some characteristics/factors which have driven the development and contributed to generating considerable benefits for partner institutions for more than ten years already. Most processes, across partner institutions, are already SDMX-model based; more than sixty data structure definitions (mainly concerning economic time series) are in use serving the exchange of harmonised data supporting the ECB’s monetary policy and objectives. Moreover, the ECB contributes to the SDMX maintenance, the statistical harmonisation and the further development of the SDMX-based technologies.

17. As the SDMX information model is present in all layers of the relevant ECB’s systems and in all its statistical processes, new development initiatives require appropriate analysis and debate about merits and costs, the trade-off between the quality and costs and, before proceeding to change, a thorough evaluation of the trade-off between expanding to a larger “scope” vs. simplicity and stability.

18. The ECB considers the use of SDMX for data exchange with its partners and, in the web dissemination context, the SDMX functionalities of the ECB Statistical Data Warehouse as an important key activity. The continuous international co-operation for further deploying and using the SDMX standards in this field could be seen as a global goal of statisticians, responding also to important requirements of a wider community (e.g. easy “pulling” of data from various websites into users’ applications for further processing). The importance of easy and automated access to statistics is ever growing along the constantly increasing demand for efficient and effective exchange of macro and micro level statistical data and metadata between parties. In his foreword to the fifth ECB conference on statistics (October, 2010), the ECB President, Mr. Jean-Claude Trichet, mentions: “[This conference has also made a significant contribution to the process supported by the G20 of providing reliable, timely and, above all, comparable global financial statistics.] *These statistics should focus not just on our own economy, but also on partner economies world-wide. To urgently close the information gaps revealed by the financial crisis, a wide range of economic and financial statistics should be created which are mutually consistent, reliable and timely as well as comparable across countries and economies.*” The use of the SDMX standards can also contribute to achieving this objective, as they are sponsored by all major international and European statistical organisations and are expected to continue supporting efficient, mutually consistent, reliable and timely statistical data exchange across countries and economies, as the ECB example discussed in this paper illustrates.

## VII. References

Peter Bull, (2004), "The Development of Statistics for Economic and Monetary Union", European Central Bank<sup>8</sup>

Jean-Claude Trichet (2010), Foreword for fifth ECB Conference on statistics, "Central Bank Statistics - What did the financial crisis change?" , European Central Bank<sup>9</sup>.

---

---

<sup>8</sup> <http://www.ecb.int/pub/pdf/other/developmentstatisticsemu200406en.pdf>

<sup>9</sup> <http://www.ecb.europa.eu/pub/pdf/other/centralbankstatistics201102en.pdf>