

WP. 8
ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES
(EUROSTAT)**

**ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT
(OECD)
STATISTICS DIRECTORATE**

Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Geneva, 9-11 February 2004)

Topic (ii): Metadata interchange

**Report on the SDMX Initiative on Common Open Standards
for the Exchange and Sharing of Socio-economic Data and Metadata**

Invited Paper

Prepared by the Bank for International Settlements, the European Central Bank, the
International Monetary Fund, the Organisation for Economic Co-operation and
Development, the Statistical Office of the European Communities, the United Nations
Statistical Division and the World Bank

1. Introduction

1. In 2001, the Bank for International Settlements (BIS), the European Central Bank (ECB), the Statistical Office of European Communities (Eurostat), the International Monetary Fund (IMF), the Organisation for Economic Co-operation and Development (OECD) and the United Nations Statistical Division (UNSD) joined together to develop more efficient processes for the exchange and sharing of data and metadata within the current scope of their collective activities. The World Bank joined the initial group of sponsor organisations in 2003.

2. The goal of the SDMX initiative is to explore common e-standards and ongoing standardisation activities that could allow national and international organisations to gain efficiencies and avoid duplication of work. The above sponsor organisations intend to do this by taking advantage of existing and emerging exchange protocols, dissemination formats and e-standards.

3. The Joint UNECE/Eurostat Work Session on Statistical Metadata received a first report on the SDMX initiative during its meeting held in Luxembourg on 6-8 March 2002 (see <http://www.unece.org/stats/documents/2002/03/metis/11.e.pdf>). The current report reviews the latest progress in SDMX initiatives and outlines future plans. In particular, Part II summarises the main aims and achievements of the four projects launched in 2002. Part III presents future steps envisaged over the next two years and the expected implications for the sponsor organisations and the wider community of statistical data and metadata providers and users.

II. The first phase of SDMX (2001 – 2003)

4. The scope of the SDMX initiative is, in general terms, the development and maintenance of standards for statistical data and metadata exchange. These standards should be based on evolving information and communication technology (ICT) developments and would be used in the context of a new co-operative effort of international organisations to:

- minimise the duplication of data and metadata collections;
- maximise the coherence of data sets disseminated by various international organisations on the same subject matter;
- improve the accessibility and interpretability of statistics disseminated by international organisations; and
- automate the data exchange processes and gain efficiency from the use of up to date ICT tools.

5. In particular, the SDMX initiative is focused on the exchange of data and metadata for economic and social statistics based both on the direct exchange of files between parties who have made prior agreements for the exchange, and on the use of Internet technologies, including open exchange standards to make data and metadata accessible to all interested parties. The international organisations sponsoring the SDMX initiative are endeavouring to maintain existing and to create new standards that will suit the needs, not only of themselves, but also of their constituencies and users around the world.

6. From a technical point of view, SDMX has been looking at the development of a common information model that can foster the use of emerging solutions and technologies. In particular, this concerns:

- electronic data interchange (EDI) conformant with ISO and UN/CEFACT norms;
- the eXtensible Mark-up Language (XML) specified by the World Wide Web Consortium; and
- opportunities for web-based services that build on standards now being developed.

In addition, work has been undertaken to develop metadata standards.

7. Several activities have been developed in the first biennium of the initiative, under the chairmanship of Ms. Carol Carson, Director of the Statistics Department of the IMF. Four projects were established in June 2002:

- a practical case study on emerging e-standards for data and metadata exchange;
- the maintenance and advancement of standards for batch time series data exchange;
- the creation of a common vocabulary for statistical metadata;
- the development of a model for metadata repositories.

In addition, a web site¹ has been developed and several presentations have been given to various audiences (Heads of National Statistical Offices, central bankers, ICT experts, etc.).

8. The aim of the **practical case study** was to investigate the ways in which emerging web standards can be used to simplify the collection, compilation and dissemination of statistical information. The case study focused on the Joint BIS-IMF-OECD-World Bank external debt statistics and national debtor data. Key deliverables included:

- identifying e-standards that can foster the achievement of practical benefits;
- demonstrating how web-based technologies may point to possible improvements and advances in statistical information exchange processes;
- recommending ways to go forward with future SDMX activities.

The main documents from the case study and the practical demonstration are available on the SDMX web site. The study revealed how new techniques and processes might work and illustrated potential benefits of emerging e-standards: opportunities for reduced reporting burden for suppliers, more user-friendly access to national and international data published on the web, easier comparative data analysis, and possible reuse of SDMX models and tools for statistical information systems at national and international levels.

9. The second project focused on the **release of GESMES/TS** for data and metadata exchange in the batch EDI mode, a standard that has been used world-wide for more than five years in automating statistical data exchange. During 2003, in light of the expected use of the message by a wide range of national and international organisations, the message was renamed from GESMES/CB to GESMES/TS (GESMES for time series). Moreover, after extensive consultations between the project team and interested users with additional requirements, the updated release of GESMES/TS 3.0 was made available in July 2003. Finally, joint work is under way with the SDMX Case Study Project to develop an XML based standard (SDMX-ML) that can support data and metadata exchange both for EDI (compatible with GESMES/TS 3.0) and for web dissemination.

10. The aim of the third project was the creation of a **Metadata Common Vocabulary (MCV)**, containing a core set of metadata items and their related definitions to improve the standardisation of metadata content for the purposes of data exchange and to promote the use of common nomenclatures that can foster international comparability of international data. This has involved close collaboration with the other SDMX projects. The current version of the MCV (available on the SDMX web site) contains five fields - *term*, *definition*, *source*, *related terms* and *context* (the last one being used to provide additional explanation and to highlight some peculiarities) – and is intended to be of use to international organisations and national statistical agencies.

11. The fourth project is developing a standard approach to the creation and use of **metadata repositories**, allowing access, analysis and reuse of statistical metadata stored in SDMX organisations as well as in other organisations that adopt emerging SDMX standards. The project took as its starting point the IMF's Special Data Dissemination Standard (SDDS) framework. The main deliverables of the project are a metamodel for metadata repositories derived by adapting and expanding the SDDS framework to

¹ More information on SDMX is available at www.sdmx.org or www.sdmx.info

accommodate the needs of other SDMX partners and an XML schema and protocol to enable search and query across metadata repositories. This project has been closely co-ordinated with the MCV project and outputs from both are being integrated into planned future work.

III. The way forward (2004-2005)

12. Even though some technical development and maintenance work is still to be completed, the first phase of the SDMX work is coming to an end. Therefore, the Sponsors discussed in the second half of 2003 how to continue the overall initiative, with the aim of delivering concrete solutions and starting the implementation of these solutions in several statistical subject matter domains. In addition to the technical work, a communication strategy has been designed to make potential users aware of SDMX proposals and solutions, and to involve national and international data providers in the change of current data and metadata exchange arrangements. Finally, the seven Sponsors agreed to work over the biennium 2004-2005 under the chairmanship of Mr. Enrico Giovannini, Chief Statistician of the OECD.

13. The first target of SDMX over the next biennium will be the release of version 1.0 standards for SDMX Common Information Model, Data Formats and Core Metadata. In particular, the work on version 1.0 involves the key standards to support data and metadata sharing and web dissemination. These elements will also be incorporated in a start-up toolkit which will become freely available to all interested users. Sponsors proposed to draft version 1.0 standards using a learn-by-doing approach. "External debt statistics" were identified as the most suitable subject matter for formulating such standards as an integral part of steps to develop a pilot implementation. In fact, this domain involves both international organisations (BIS-IMF-OECD-World Bank) and national agencies in producing and disseminating statistical data (respectively, from the creditor and the debtor perspectives) concerning the same phenomenon. The version 1.0 of SDMX standards will benefit from all results already achieved by the four projects carried out over the last biennium.

14. The Inter-Agency Task Force on Finance Statistics (TFFS), at its September 2003 meeting, has recently agreed to redesign the joint external debt statistics, integrating the currently available creditor sources of the BIS, IMF, OECD and World Bank with new statistics being disseminated (as prescribed under SDDS) using the IMF's compilation guidelines by debtor countries themselves. The intention is to extend the prototype developed by the SDMX case study to build a dynamic joint web hub linking all four organisations and national agencies producing and disseminating debt statistics.

15. Inter-Agency Task Force and SDMX members have agreed to use the resources for the joint external debt project to deliver version 1.0 standards and related tools, which will be generic to all statistical domains. A project proposal has been prepared to address the following deliverables:

- first draft of version 1.0 SDMX core standards for data exchange and web dissemination and their release for public comment;
- endorsement of version 1.0 core standards by the Sponsors Committee and steps to anchor them within the international standards community;
- draft version 1.0 SDMX technical architecture standards and their release for public comment;
- endorsement of version 1.0 of standards for technical architecture by Sponsors and steps for anchoring them within the international standards community;
- release of pilot production infrastructure and release of the associated Joint External Debt Hub implementation.

16. In addition, SDMX sponsors will start preparing the groundwork for implementing the SDMX standards in other statistical domains. The OECD is working on the adoption of SDMX standards to share national accounts data with Member countries and other international organisations. The OECD and the

IMF are exploring the use of web services between internal data base systems at these institutions. For international trade data, the OECD and the UNSD are building a shared database based on SDMX standards. Finally, at the beginning of 2004, SDMX Sponsors will identify one or two additional subject matter domains. The idea is that the implementation of SDMX in these and other statistical domains will promote in future the development of new versions of SDMX standards.

17. Apart from developing a sequenced implementation of SDMX standards in various statistical domains, the SDMX sponsors also envisage to foster the adoption of GESMES/TS (e.g. via training programmes) and the regular update of the MCV. Moreover, they envisage the release of a production version of the metamodel for metadata repositories based on SDMX standards.

18. As it is not possible to develop world standards on statistical data and metadata exchange without having close relationships with other bodies in charge of setting similar standards and with ultimate potential users of the latter, special attention will be paid over the next two years to the communication of SDMX results to potential users and partners. Relationships with national statistical data providers (NSOs, central banks, etc.) and international organisations active in statistics will be reinforced, using existing bodies belonging to SDMX Sponsor Organisations, and the SDMX web site will be redesigned to be more interactive and to solicit feedback. In addition, relationships with international bodies (e.g. ISO) which set world standards linked to ICT matters, international information brokers and groups involved in information policy management, such as groups focusing on e-government standards (e.g. OASIS), accounting standards (e.g. IFAC), etc., will be established or reinforced.

IV. Conclusions

19. The SDMX initiative has the potential to make an important contribution in helping international and national agencies involved in the production of statistical data and metadata to become more efficient in their internal operations, to avoid duplication in data collection and to provide users with a higher quality service. A first set of complete results of the initiative are expected in 2005, even if important deliverables (GESMES/TS 3.0, the Metadata Common Vocabulary, etc.) are already available to users and/or will be ready for external distribution over the next 12 months.

20. New ICT developments can be applied to several aspects of statistical activities, but to fully benefit from them, international and national organisations must also address governance issues. In some cases, SDMX standards may provide a possibility to statistical “communities” to reconsider the current pattern of data and metadata exchange and change the existing division of labour. On the other hand, new models for data exchange and dissemination based on the concept of “hub” (a common registry keeping track of data and metadata available on a network of web sites of partners) can reinforce the role of national statistical offices in co-ordinating other data providers. Finally, international organisations have the opportunity to avoid duplication in data collection, share data and metadata between themselves and other users, and establish new ways of co-operation and co-management of international statistical activities.

21. The METIS Work Session may wish to comment on the expected technical developments and to discuss how these can impact data and metadata exchange practices.