

**Economic and Social Council**Distr.: General
21 April 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**Statistical Data and Metadata eXchange in statistics, complexity, costs and benefits****Note by the former Yugoslav Republic of Macedonia, prepared by the State Statistical Office***Summary*

Statistical Data and Metadata eXchange standards were originally designed for the exchange and sharing of statistical data and metadata amongst international organizations. Therefore, usage of Statistical Data and Metadata eXchange standards is relatively limited within the statistical organizations and is primarily focused towards harmonization and standardization in the dissemination phase of the statistical process. In order to achieve a more efficient exchange of comparable information, the State Statistical Office has made an effort to implement Statistical Data and Metadata eXchange in statistical work. The new Statistical Data and Metadata eXchange compliant reporting template for reference metadata, i.e. the Euro Statistical Data and Metadata eXchange Metadata Structure, was introduced in the State Statistical Office in 2009. Furthermore, the State Statistical Office has undertaken the initial step to adopt its own statistical business process model in accordance with the Generic Statistical Business Process Model. Statistical Data and Metadata eXchange aims through Content-oriented Guidelines to harmonize data and metadata terminology, as well as to provide transmission standards. The Generic Statistical Business Process Model, in offering standard terminology for different phases and sub-processes of the statistical business process, would seem to complement and fit logically within those guidelines. Since Statistical Data and Metadata eXchange has become an accepted concept for future gradual step-by-step implementation at the State Statistical Office, the paper will discuss how its complexity could be managed and how time and resource consuming tasks could be justified from an efficiency point of view and re-usability of metadata.

I. Statistical Data and Metadata eXchange implementation - a managerial challenge

1. Statistical Data and Metadata eXchange (SDMX) consists of technical and statistical standards and guidelines, information technology (IT) architecture and IT tools, to be used for an efficient statistical data and metadata exchange. The harmonization of the dissemination formats of data and metadata sets improves accessibility and clarity in the access to statistics, especially for the consumers of statistics.

2. With regard to the standardization of processes, efficiency, change, and the continuing modernization of statistics production, a model-based approach is needed and the implementation of the SDMX model could be regarded as a possible solution. The implementation of the SDMX model or modifications of the existing models in national statistical organisations (NSOs) is a resource-consuming task and must be a top management level decision in order to promote its use within the organization as well as to justify the investments needed.

3. In order to promote SDMX use within their organizations, top management in national statistical offices needs to be convinced of the benefits gained by using an architecture that enables the exchange of information in a standardized way.

4. Although SDMX implementation is considered as strategic, advantages are not immediately obvious and there is a very small number of SDMX competent staff in NSOs. The importance of analysing, building and managing the standard from a content, and not just a technology, point of view should be stressed.

II. The governance model for future Statistical Data and Metadata eXchange development work

5. The aim of the SDMX initiative is to create and maintain technical and statistical standards and guidelines, together with information technology architecture and tools. On the development side, the SDMX initiative shows ongoing progress, but on the implementation side progress is slower, unless a sponsor organization is confirmed. Even in this category, Eurostat is behind schedule. In NSOs, domain experts are not yet convinced of the benefits of SDMX and consider it an additional burden.

6. The SDMX sponsors decision to strengthen the involvement of NSOs in the maintenance and development of SDMX as a new SDMX governance model may improve the level of SDMX acceptance and its implementation.

7. Therefore, in the proposed governance model, technical and IT standards should not be viewed separately from statistical standards and business process integration. Technology is a prerequisite, but the development should not be IT-driven. Benefits from common standards recognized by statisticians and clear business case should govern the future SDMX development.

III. Barriers to Statistical Data and Metadata eXchange implementation, and how to overcome them

8. On the NSO side, SDMX knowledge and competence have to be developed among subject-matter staff and integrated with particular statistical knowledge already built into the existing production systems. On the other hand, there is a need for insight into SDMX

developments within other SDMX sponsoring organizations in order to be able to provide metadata to more organizations, using as much as possible information extracted from own metadata systems, and reducing manual interventions, double work and inconsistencies.

9. Using SDMX standards does not necessarily mean that national organizations could adopt the same statistical and metadata concepts in the short term. Instead, the existing concepts need to be mapped against the concepts identified in a common Metadata Structure Definition. The strategy of SDMX implementation may differ between statistical organizations depending on the level of development of their meta-information systems and their adherence to the existing metadata standards.

10. Regardless of the implementation strategy, more or less everybody is faced with the fact that SDMX is being treated as a technical issue. To some extent, there is a tendency to view SDMX as just another format.

IV. The complexity of implementing Statistical Data and Metadata eXchange standards

11. Implementing SDMX standards is multi-disciplinary and requires teamwork across a variety of skills (statistical, technical and managerial). The contribution of those with a sound knowledge of the overall statistical business process and information system architecture is crucial for making the most of the use of the SDMX framework of standards and guidelines.

12. As the State Statistical Office (SSO) is committed towards implementing SDMX, a number of employees attended SDMX courses that focused on different aspects of SDMX, both statistical and technical. Summarized participants' views are that SDMX is too demanding to implement.

13. To make the complexity of implementation manageable, future capacity building activities should be focused more on gaining knowledge for the implementation strategy in order to have a clear view of the larger SDMX picture and steps should be defined how to gradually build the environment. This approach is supposed to downsize the complexity of the issue, but also to justify used resources, to see if benefits from an efficiency point of view and re-usability of metadata are really achieved. Introducing the Euro-SDMX structure (ESMS) for describing reference metadata is the first step towards SDMX implementation at the SSO. The next step would be producing SDMX files with reference metadata directly from the ESMS database.

V. Combining standards (e.g. Statistical Data and Metadata eXchange and Data Documentation Initiative)

14. SDMX and Data Documentation Initiative (DDI) take very different views of the data exchange process and data lifecycle regarding the differences in focus and scope of these standards. One of the key differences between DDI and SDMX is the intended use of the standards. DDI, in its early versions, was primarily intended as an archival standard, providing an electronic format for descriptive, human-readable metadata for social sciences researchers. SDMX has a different focus - it is designed to facilitate the automated exchange and processing of data and metadata between organizations.

15. It should be noted here that DDI and SDMX are standards which are related, but which are not in competition with one another. The choice of which standard to use depends on the focus of the organization which is doing a standards based implementation. It may well be that both standards are useful within an organization which deals both with

micro-data and aggregates. But, there are ways in which both standards can be used together or complement each other in data and metadata exchange:

- (a) Using DDI metadata as a link to source data for SDMX aggregates;
- (b) SDMX and DDI as complementary formats for processing and dissemination;
- (c) The SDMX Registry as a DDI metadata repository to support lifecycle.

VI. Developing the necessary tools - coordinated and shared development

16. The existing SDMX tools have been developed (or commissioned) by SDMX sponsors and other organizations actively involved in implementing SDMX and, in general, are made available under open source licences. Although the SDMX users are encouraged to evaluate and test the already existing tools and modules and to contribute to and cooperate in the further development of SDMX, many enthusiastic users found that the tools and modules are not easy to re-use: they require the existence of a certain SDMX framework and have to be accompanied by proper user documentation as well.

17. In the coming years, the NSOs will be requested to deliver data in a more standardized way, to prepare a dissemination database and develop computer systems that will provide data in an automatic way, and to satisfy user needs efficiently. Each of the national or international institutions is free to choose SDMX tools and elements that are more suitable in any particular case, thus reducing the time for implementation of SDMX.

18. The SSO's intention is to use SDMX tools for statistical data and metadata that exist as open source applications and that could be used in accordance with the existing IT solutions.

VII. Implementing and expanding the Statistical Data and Metadata eXchange Content-oriented Guidelines

19. International organizations use many different standards for data and metadata. These standards use statistical concepts or codes that are not standardized between international organizations. Implementing the SDMX Content-oriented Guidelines (COG) should help to overcome those difficulties concerning statistical data and metadata.

20. The usage of SDMX COG should be promoted to further implement them at national and international levels because they focus on the harmonization of specific concepts and terminology that are common to a large number of statistical domains. Such harmonization complements the potential efficiency gains to be achieved when applying the SDMX technical specifications.

21. In particular, Euro SDMX Metadata Structure (ESMS) aims at documenting methodologies, quality and the statistical production processes and it uses 21 high-level concepts, with a limited breakdown of sub-items, strictly derived from the list of cross domain concepts in the SDMX COG. Based on the SDMX 'acquis', Eurostat improved and extended the currently used Special Data Dissemination Standard (SDDS) structure for reference metadata. This improved and extended standard for reference metadata now also includes many standard quality criteria. The application of this new standard will improve the production and dissemination of data quality within the European Statistical System considerably.

VIII. Reconciling the needs of national and international statistical organizations

22. Both national and international statistical organizations are faced with several challenges:

- (a) To reduce costs;
- (b) To improve effectiveness and ability to respond;
- (c) To improve the level of international cooperation.

23. The permanently increasing demands of official statistics can be met only with a new generation of information management systems. In this resource-consuming development task, statistical institutions should not act alone.

24. International statistical organizations, and SDMX sponsors in particular, are showing progress in the improvement of technical standards, guidelines and tools, but obviously it would be advantageous to have a more balanced development between technology and content. There is an urgent need for the international organizations to better harmonize their own requirements for Data Structure Definitions (DSDs). The slow process of achieving agreement on DSDs and Metadata Structure Definitions (MSDs) between organizations is the main reason for not having sufficient DSDs and MSDs to allow national organizations to map their data and metadata to SDMX.

25. Agreement on the content issue brings many benefits to each of the national and international participants. The greatest benefit of deploying the SDMX standards for the national organizations should be the possibility to provide data and metadata to more international organizations once and for all. When this becomes possible, there will be no more discussions about slow progress implementation. Unfortunately, it seems there is still a long way to go.

IX. References

“SDMX user guide”

<http://sdmx.org/wp-content/uploads/2009/02/sdmx-userguide-version2009-1-71.pdf>

“SDMX content-oriented guidelines”

http://sdmx.org/wp-content/uploads/2009/01/00_sdmx_content-oriented_guidelines_2009.pdf

Statistical Commission Forty-second session, Note by the Secretary-General “Progress report of the Task Force to Establish Standards on Data and Metadata Exchange”
<http://unstats.un.org/unsd/statcom/doc11/2011-31-SDMX-E.pdf>

Doc. Eurostat/ITDG/October 2010/3.4.c, “SDMX: How to go ahead faster”
http://circa.europa.eu/Public/irc/dsis/itsteer/library?l=/directors_19-20&vm=detailed&sb=Title

Rune Gløersen, IT Director, Statistics Norway “Interoperability using SDMX Some viewpoints from an NSF”, Paper presented at the OECD SDMX Meeting, Geneva 8 – 9 March 2010

<http://www.oecd.org/dataoecd/15/24/44642613.pdf?contentId=44642614>

Conclusions from the meeting in Geneva 8-9 March 2010 , “Joint OECD/UNECE Expert Group on Statistical Data and Metadata Exchange”

<http://www.oecd.org/dataoecd/16/14/44829878.pdf>

Steven Vale, “Exploring the relationship between DDI, SDMX and the Generic Statistical Business Process Model”

<http://www1.unece.org/stat/platform/download/attachments/55476343/EDDI+paper.pdf?version=1>

Arofan Gregory, Pascal Heus “DDI and SDMX: Complementary, Not Competing, Standards”

http://www.opendatafoundation.org/papers/DDI_and_SDMX.pdf

G rard Salou and Xavier Sosnovsky, “SDMX as the logical foundation of the data and metadata model at the ECB”

<http://www.bis.org/ifc/publ/ifcb33i.pdf>

SDMX Guidelines 2009. “ Mapping of SDMX Cross-Domain Concepts to metadata frameworks at international organisations (IMF-Data Quality Assessment Framework, Eurostat-SDMX Metadata Structure and OECD-Metastore)”

http://sdmx.org/?page_id=11

Marco Pellegrino, Eurostat “SDMX implementation in the statistical practice”

http://www.czso.cz/conference2009/proceedings/data/meta_systems/pellegrino_paper.pdf

Papazoska, H., Ristevska Karajovanovikj, B. and Lipikj, S. 2010. “The value of adopting and implementation of ESMS structure in Macedonian State Statistical Office”, Paper presented at the Work Session on Statistical Metadata (METIS), Geneva, 9-11 March 2010

<http://www.unece.org/stats/documents/2010.03.metis.htm>
