United Nations ECE/CES/2011/31

Economic and Social Council

Distr.: General
14 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: strategic and operational challenges for a national implementation

Note by the Swiss Federal Statistical Office

Summary

The paper provides an overview of the challenges of implementing the Statistical Data and Metadata eXchange standard at national level and how the Swiss Federal Statistical Office meets these challenges. Apart from a standardised format for exchanging information, Statistical Data and Metadata eXchange offers an information technology framework which concerns practically every process of the statistical life-cycle and which implies a harmonisation of the metadata. For a statistical institution, implementing Statistical Data and Metadata eXchange often represents a paradigm shift, which involves adequate training of employees and, in particular, a strong commitment from the top management. The understanding and the acceptance of Statistical Data and Metadata eXchange are key factors for taking full advantage of all benefits that such a standard could offer.

I. Abstract

1. Statistical Data and Metadata eXchange (SDMX) is becoming the new standard for modelling and exchanging statistical information. A number of important issues are at stake: comparability of data and metadata, increased quality, faster dissemination, reuse of information, standard exchange and lower costs for production and for information technology (IT) development.

2. The adoption of SDMX poses several challenges to national and international statistical agencies. Apart from a standardised format for exchanging information, SDMX offers an IT framework which concerns practically every process of the statistical life-cycle and which implies a harmonisation of the metadata. Concerning Switzerland, the harmonisation process also includes the integration of SDMX with other existing national standards.

3. For a statistical institution, implementing SDMX often represents a paradigm shift, which involves adequate training of employees and, in particular, a strong commitment from the top management. The understanding and the acceptance of SDMX are key factors for taking full advantage of all benefits that such a standard could offer. It is therefore important to be able to benefit from the know-how of the national institutions which have already integrated SDMX into their information systems.

4. The main challenge of SDMX is to assure sharing and standardised access to statistical information. A nationwide coordination for the implementation of the available SDMX technologies would allow this information-sharing and would open new perspectives for corporate access to statistical information. In this sense, the Census Hub solution proposed by the EU is already a concrete expression of this future vision.

II. The Statistical Data and Metadata eXchange standard and the challenges it brings to an institution willing to adopt it

5. The existence of a generic statistical metadata model, which could be used by all statistics, has been regarded for a long time as truly utopian. A number of countries have made significant efforts to create and maintain their own centralised metadata systems, but the solutions they came up with were based on classical data base systems and were usually too complex, too heavy to manage and static by design. Although a large number of international references exist in statistical data and metadata modelling, only a few fully functional realisations have been created and used for a longer period of time.

6. One such reference is the relatively new standard SDMX. Mainly, SDMX allows the implementation of low cost solutions which could cover the entire statistical production value-chain and could offer noticeable rapidity in the production process, lower IT costs, simplified user manipulations, and finally, faster delivery of data publications and metadata reports. Despite all the benefits that this standard brings to its users, SDMX also carries plenty of challenges, which should be efficiently overcome in order to exploit its full potential.

7. To consider SDMX just as a standardised format for exchange of statistical data and metadata counting on different tools to do the mapping to and from SDMX would be a too restrictive approach for the adoption of this new standard. Indeed, SDMX is initially conceived to be much more than a simple exchange format and thus, to use it in such a manner would only bring higher workload to the statisticians than any noticeable benefits to the adopting institution.

8. The main challenges of SDMX, in an institutional perspective, could be summarised as follow:

(a) To develop awareness of the potential of SDMX and the benefits that it could bring to an adopting institution;

(b) To secure a firm decision by the top management for the implementation of SDMX;

(c) To define a practical strategy for an implementation of SDMX and to create the necessary conditions for accomplishing it (model of governance);

(d) To establish SDMX compliant architecture and tools that fit into an existing IT infrastructure;

(e) The progressive (step-by-step) deployment.

9. One should note that an exhaustive SDMX implementation would have an impact on the entire statistical life-cycle and thus, on all the actors involved in the statistical production. Concerning the statisticians and the data production units, the adoption of SDMX signifies:

(a) New user-friendly generic tools that would allow the visualisation, navigation and editing of the statistical metadata and the access to the statistical data;

(b) Knowledge-management and documentation tools that guarantee the relevance of the information as well as the transmission of the statistical heritage;

(c) Increased statistical production efficiency, starting from the concept phase until the diffusion of the final statistical results.

10. Concerning the management units, the adoption of SDMX means:

(a) Simple and flexible IT architecture, which offers transversal solutions instead of domain specific implementations;

(b) Faster and modular IT development, which offers better compatibility with the existing applications;

(c) Quicker statistical production with higher data quality and improved documentation.

11. Finally, the successful implementation of SDMX by many central banks has already shown that the challenges are realistic and could serve as an excellent reference to all institutions willing to adopt SDMX.

III. The complexity of implementing Statistical Data and Metadata eXchange

12. The SDMX information model is a relatively complex one and as such, its implementation requires a profound knowledge and experience in the area of the statistical data and metadata. Moreover, an efficient integration of any new standard in an existing IT environment should only be accomplished progressively, where the novelties are introduced painlessly and the awareness of their existence is distributed over time. In this respect, the Swiss Federal Statistical Office (FSO) chose a progressive strategy for the adoption of SDMX, which consists of multiple steps allowing the decomposition of the complexity of the standard into smaller and more-accessible stages.

13. The successive stages of the implementation can be summarised as follows:

(a) Step.1. Translation (mapping) of the currently used metadata models (e.g. the *Neuchâtel group metadata model*) into the SDMX information model. This work is essential, since it requires profound understanding of both standards as well as of the XML representation of SDMX. An SDMX compliant graphical editor provides to the end user a convenient way to create SDMX metadata entities from different sources of metadata, as well as the possibility to optimise certain metadata representations;

(b) Step.2. Migration of the statistical metadata used by one or several statistics (as pilot projects) into the SDMX format. This will allow the examination of the statistical metadata created at different stages during the life-cycle of a statistical activity, as well as the introduction of the SDMX standard to the statisticians, offering at the same time a practical and efficient solution. The transversal metadata created during the process is stored in a central repository (a SDMX registry) and is made accessible via a generic web interface to all users;

(c) Step.3. Developing of adequate interfaces to the existing statistical applications. This would allow the reutilisation of the already created transversal metadata by the statistics without disturbing the production process. Even if a generic SDMX solution would theoretically allow to model and manage the quasi-totality of the available statistical metadata, it would be difficult to cover all kinds of existing particularities. Considering this, the FSO has chosen a practical approach based on a flexible SDMX integration, where the SDMX solution could easily be extended by efficient "plug-ins" (small personalised modules), which would satisfy all sort of specific functional requirements. This would guarantee a standard and generic solution based on the SDMX information model, which satisfies at the same time the specific needs of the statistical users;

(d) Step.4. Integration of SDMX with the statistical units and existing IT applications. The integration could be accomplished during the revision of the statistical production and the modernisation of the existing IT applications;

(e) Step.5. Introduction of the SDMX "Data Hub" architecture and efficient web visualisation tools, which will allow the access and navigation through the available data via the existing metadata. The SDMX "Data Hub" solution could be seen as a generalisation of the *Census Hub* approach proposed by the European Union, which serves as an example of a concrete realisation of the SDMX Hub architecture and is used to visualise and navigate through the European population census data.

14. The harmonisation of the statistical metadata is implicitly affected by all steps mentioned above. That is, the structural harmonisation is accomplished by default once the implementation of SDMX is completed. However, the harmonisation of the content should be accomplished explicitly and has two main aspects to be considered: the definition of non-redundant statistical concepts and the political aspect.

15. The statistical business process is also affected by the adoption of SDMX. In this context, the Generic Statistical Business Process Model (GSBPM) is closely related to SDMX and offers a starting point for an institution to define its own detailed business process in a close relation to the statistical metadata. The GSBPM is a very generic and highly flexible model, which offers to each statistical unit a way to describe its business process needs. That is, if each of the GSBPM sub-processes is considered as a building block that provides a concrete and modular solution, the SDMX standard can be used as an interface between these solutions allowing a standardised information flow from one building block to another.

16. Finally, in order to get the most of SDMX, the harmonisation of the statistical data modelling should also be considered. Harmonised data models could have a particular importance for the statistics which use output data produced by other statistics and which

have complex interpretation processes. Similarly, the diffusion of the statistical data could also benefit from harmonised data models, e.g. using the SDMX "Data Hub" paradigm which would allow better visibility and higher quality of the diffused data and metadata.

IV. Strategies for overcoming the obstacles to a successful Statistical Data and Metadata eXchange implementation

17. Some of the main obstacles to the implementation of SDMX, together with several strategies for dealing with them are presented in the table below.

Main Obstacles	Examples of Strategies
Ignorance – SDMX is often considered only as a simple format for exchange of statistical data and metadata.	**1. Involvement of senior management:** convince the senior management of the necessity of the SDMX implementation, using strong advocacy, precise information and appropriate awareness. **2. Tangible results:** focus on the practical demonstrations of SDMX tools using real productive data, in order to demonstrate the huge potential which the model has. **3. Acknowledged competence:** establish a SDMX competence centre, which should assure the acquisition of know-how and the training of the statisticians.
Harmonisation – most of the harmonisation process becomes natural with the implementation of SDMX, which however requires: sufficient resources, operational IT tools and common transversal solutions.	**1. Policy:** issue a policy concerning the management and the harmonisation of the statistical metadata. **2. Model of governance:** elaborate a governance model, which defines the processes and the organisational structures involved. **3. Guidelines:** establish a set of rules to be used during the management and maintenance of the statistical metadata. **4. Generic tools:** develop generic tools, capable to expand and adapt to the specific needs of the statistical production process.
Change resistance – part of the statistical users and developers will show strong resistance concerning SDMX because of its complexity and the general change of paradigm.	**1. Involvement of senior management:** the involvement and the participation of senior management is a necessity. **2. Competent team members:** confine the SDMX implementation to a well trained group of people who possess the necessary competence and confidence to accomplish it. **3. Short implementation phases:** split the implementation into short phases (2-3 months each), where after each phase a part of the system is functional and could be used by the statistical users. **4. Operational tools:** quickly provide to the users tools and methods to accomplish their work (concerning the data production, analysis and dissemination) and assure

Main Obstacles	Examples of Strategies
	an adequate support.
Complete outsourcing – the outsourcing of the SDMX IT development automatically casts doubt on the relevance of SDMX as a competitor to the classical IT solutions, since the IT companies rarely possess the know-how and, thus, promotes classical approaches.	**1. Acquisition of know-how:** build the know-how within the institution. **2. National and international coordination:** coordinate the implementation with other national and international institutions and exchange tools and accumulated experience. **3. Open source tools:** use open source tools and assure the in-house capacity for adapting these tools for the specific IT needs of the specific statistical institution. **4. Limited outsourcing:** avoid the outsourcing of the metadata development or outsource small well defined IT modules that could be developed by SDMX ignorant personnel.
System usage – a widespread and long term usage of the new SDMX system is one of the main operational challenges.	**1. Integrated system:** provide to the end users an integrated solution which would allow simple usage and flexible creation of metadata during the entire statistical production process. **2. Remove redundancies:** any metadata should be managed in one single system only. **3. Balance generic and internal specific solutions:** the new system should be flexible enough to satisfy the user specific needs and to attract them, but at the same time should remain generic and 100% SDMX compliant.
System compliance – compliance of the new system to the existing in-house norms and solutions. Specifically, the definition and the development of interfaces between the new system and the existing IT applications could be an elaborate and never ending task.	**1. Clear system limits:** the statistical metadata managed by the new system should be clearly defined from the beginning of the implementation, as well as all the systems which would be replaced or interfaced by it. **2. Start with interfaces development:** the SDMX integration within an existing productive infrastructure is a complicated endeavour. **3. Plug-ins:** develop plug-ins for all in-house user-specific and non-generic functionalities.
Quality management – ensure that the different statistical users respect the new standard without altering the SDMX model.	**1. Organisational guidelines:** elaborate simple and unambiguous guidelines concerning the statistical business processes, the user roles and the quality control mechanisms. **2. In-house training:** assure the in-house IT and statistical training for the concerned personnel. **3. Efficient support:** provide efficient user support.

V. Conclusion

18. This paper presents the main advantages and implications of adopting SDMX in a statistical organisation, where a progressive implementation of the standard is defended as a mean to minimise the negative impacts on the ongoing statistical production. In section two, a five-step implementation programme is proposed as a practical solution for adopting SDMX. The main sources of institutional problems are summarised in the form of obstacles in section three. Examples of strategies to overcome each of the presented obstacles are proposed based on the accumulated experience in the Swiss Federal Statistical Office. Finally, the main global challenge of SDMX is found to be the realisation of a standardised access to the entire statistical information (data and metadata) and the coherent management of the statistical metadata considering all its institutional diversities and particularities.