



**Economic and Social
Council**

Distr.
GENERAL

ECE/CES/2008/3
2 April 2008

Original: ENGLISH

ECONOMIC COMMISSION FOR EUROPE

STATISTICAL COMMISSION

CONFERENCE OF EUROPEAN STATISTICIANS

Fifty-fifth plenary session
Paris, 10-12 June 2008
Item 4 of the provisional agenda

STATISTICAL METADATA

Note by the Steering Group on Statistical Metadata¹

Summary

The paper is prepared at the request of the Bureau of the Conference of European Statisticians. At its October 2007 meeting, the Bureau emphasised the need to send a strong message about the fundamental importance of metadata to the Heads of statistical offices and decided to include the topic in the agenda of the 2008 plenary session of the Conference.

This paper explains the role and importance of metadata in the production of official statistics. New technologies are enabling statistics to be rapidly and widely communicated, increasing the number of users and the diversity of their needs. This highlights the necessity for standards and systems to maintain consistent, high-quality metadata across the statistical business process. Metadata are not a purely technical issue, but rather a vital management tool. The role of metadata, its corporate value, and the benefits to internal and external users groups are discussed in this paper. The Conference is invited to formulate a message to senior managers of national statistical offices as outlined in the introduction to the paper (paragraph 3).

¹ The document was submitted one day after the deadline because additional clarifications were needed in a review process by members of the Steering Group on Statistical Metadata.

I. INTRODUCTION

1. Economic globalisation and the rapid growth of information and communication technologies (ICT) have a significant impact on statistical services. While facing a growing number of users at both national and international level, statistical offices must react to changing user needs, provide more integrated statistical information, and respond to calls for higher quality. At the same time, they must increase the efficiency of statistical production and reduce respondents' burden.
2. All this represents new challenges for statistical information systems. These systems must be better integrated across different subject matter areas, as well as support the statistical business process cycle from early specification of needs, through to collection, processing, analysis, dissemination, and archiving of statistics. Statistical metadata represent an indispensable element for integrating statistical information systems and are the key to modern corporate management of statistical activities. The high number of metadata users and their diverse needs highlight the strategic nature of the metadata systems that support statistical business processes.
3. Metadata are not a purely technical issue, but rather a vital management tool. This implies that metadata management should be an integral part of the high-level strategic plans of a statistical office. The Conference is invited to send a message, through the conclusion of its 2008 plenary session, to executive heads and senior managers of statistical offices outlining that:
 - (a) Metadata are important, both for efficient corporate management of statistical offices and for users of statistics;
 - (b) Metadata systems must not only be developed, but also maintained efficiently, through remodelling business processes accordingly and employing automation to the highest extent possible;
 - (c) Metadata and metadata systems provide the necessary infrastructure for effective modern statistical information systems.

II. THE CORPORATE VALUE OF METADATA

A. Role and functions of statistical metadata systems

4. The definition "metadata are information about information" predetermines that the Statistical Metadata System (SMS) contains information about the Statistical Information System (SIS). Integrated and transparent descriptions of information flows inside and outside a statistical office are vital. The use of technology for data collection, interactive communication with users, and dissemination of statistics, calls for a coherent and well-functioning SMS. Processing of statistical data should be driven by metadata stored in the SMS, and the outcome of processes should be documented in the SMS.
5. In this context, the SMS should be a tool enabling statistical offices to effectively perform the following functions:
 - (a) Produce official statistics efficiently across all phases of the statistical business

process (including data collection, storage, evaluation and dissemination);

- (b) Plan, design, implement and evaluate the statistical production process;
- (c) Manage data sources and cooperation with respondents;
- (d) Manage methodological activities;
- (e) Improve discovery and exchange of data between statistical offices and users,
- (f) Improve quality of statistical data by providing a relevant set of metadata for all quality criteria;
- (g) Disseminate and exchange statistical information by facilitating search, navigation, and interpretation, assisting in post-processing of statistical data, and recording user feedback;
- (h) Improve integration of statistical information systems with other national and international information systems, in order to facilitate collection of data from administrative registers and records;
- (i) Improve international integration and comparability of official statistics;
- (j) Manage, unify and standardize workflows inside the statistical office;
- (k) Provide a knowledge base on the processes of the SIS, to facilitate collaboration and mobility of staff between different areas;
- (l) Improve administration of SIS by clarifying responsibilities, legislation, performance, and users' satisfaction;
- (m) Manage the costs and revenues of the statistical office;
- (n) Unify statistical terminology and provide a vehicle for better communication and understanding between managers, designers, subject matter statisticians, methodologists, respondents and users of the SIS.

B. Major users of statistical metadata

6. A primary challenge for the SMS is to cope with the requirements of different metadata users. Improvements in information and communication technologies (ICTs) have created more users of statistics with a diversification of needs. Effort should be made to understand who the users are, as requirements for data and metadata may vary substantially. The major users of statistical metadata may fall into the following groups:

- (a) **Users inside statistical offices** encompass many professions involved in the phases of preparation, production, dissemination and communication of official statistics and the functioning of the SIS. These include the following metadata users:

- (i) Senior management;
- (ii) Methodologists, subject matter statisticians and statistical researchers;
- (iii) Developers and administrators of information systems;
- (iv) Corporate planners, administrative staff and auditor;

(b) **Respondents** – those that supply data to the statistical office. This includes suppliers of administrative data for statistical use.

(c) **End users at the national level** include governmental institutions, political decision makers, researchers, public officials, academics, archivists, librarians, journalists and the general public. Non-government organizations are also important users at the national level.

(d) **International users** - individuals, multinational enterprises, international organizations and others are important users of statistical metadata. Integration of metadata from national statistical organizations with the statistical metadata of international users is becoming increasingly necessary.

C. Benefits that statistical metadata systems provide to users

7. Statistical organizations and other metadata stakeholders benefit when consistent metadata flows from the point of creation to archiving the data, rather than remaining captive to a particular statistical processing system or infrastructure package. The following benefits are applicable for all user groups outlined above:

- (a) Better discoverability, retrieval and exchange of data and metadata between organizations to improve access to statistics;
- (b) Use of common terminology, names and descriptions for standard metadata elements to improve communication and understanding;
- (c) Central metadata repositories organized to facilitate reuse of existing data;
- (d) Increased use of metadata standards;
- (e) Improved knowledge of metadata flows;
- (f) Better quality statistical information and metadata.

8. The specific benefits that the SMS provides to each group of users² are described overleaf.

D. Users inside statistical offices

- (a) Senior management: The SMS facilitates design, planning, decision-making and

² More details on benefits that statistical metadata systems (SMS) provide to specific groups of users are provided on the website of the Common Metadata Framework (CMF): www.unece.org/stats/cmf/PartA3_3.html

evaluation processes of the SIS. The SMS provides a tool for monitoring the use of data and user feedback; accuracy, timelines, availability and coherence of the statistical output; as well as the costs and benefits of individual statistical activities.

(b) Designers and evaluators: These users need access to metadata from similar systems, either within or outside the organization, to inform the design, development and implementation of a new system.

(c) Methodologists: The SMS creates a framework for design and implementation of statistical tasks and surveys in production of official statistics, and for meeting the needs of end users. Furthermore, the SMS facilitates the maintenance, use and further development of statistical classifications and nomenclatures, the use of statistical registers, the promotion and maintenance of statistical standards, and the documentation of knowledge about statistical methods and relevant research methods.

(d) Subject matter statisticians: The creation and provision of associated metadata helps subject matter statisticians to understand users' information requirements, to communicate information to users and to evaluate the statistical business process. The SMS represents a knowledge base about various objects of metadata description (including a history of changes) like statistical classifications, statistical variables, statistical tasks, surveys, statistical production process, including all previous iterations of that process.

(e) Statistical researchers: These users need easy access to metadata to help them to interpret and understand the quality of data, and to provide the basis for data integration.

(f) Administrators of the metadata content: The SMS should ensure smooth and systematic update and maintenance of statistical metadata. Metadata should be updated within the SMS corporate metadata repository, once only and in one place, in order to avoid inconsistencies and unnecessary redundancies. Linked updates to all the dimensions of the corporate metadata repository should be automated, and a user friendly interface should be available for human interaction.

(g) Technical administrators of metadata: IT experts should use SMS tools for technical maintenance of the corporate metadata repository, in cooperation with designers, evaluators and content administrators.

(h) Information technology specialists: Metadata driven statistical production creates favourable conditions for standardization and thus efficiency of IT systems supporting statistical production.

E. Respondents

9. The role of respondents is becoming more important with the growing number of systems and on-line communication options. Bearing in mind the possibility of on-line supply of data directly from respondents' information systems, the SMS plays a key role in facilitation and automation of tasks related to data capture.

F. End users at the national level

10. The SMS should help users to discover, understand, interpret and interrogate statistical data. The proliferation of information has raised the issue of consistency and comparability of data, and these can be addressed by a well-functioning SMS by explaining what the differences are and the reasons for them. The SMS should also assist in conveying the credibility of statistical data and in the recognition of intellectual property rights.

G. International Users

11. There are increasing demands by international users for greater consistency between national statistics. A well functioning SMS should provide a basis for more efficient sharing of statistical data and metadata and decreased burden on national data providers. The Statistical Data and Metadata Exchange initiative (SDMX) is a positive example, in particular SDMX joint hubs and metadata registries³.

II. STRATEGIES AND POLICY FRAMEWORK

A. Preparation of a corporate SMS vision

12. A prerequisite for the successful design, implementation and functioning of the SMS is the development of a corporate vision of SMS in the statistical organization. The functions of the SMS, centred upon metadata and data users, are oriented towards the diverse processes and activities of the SIS. Organizational units within a statistical agency, respondents and end users are all involved in the preparation, implementation and use of the SMS tools. The vision should be developed with the direct involvement of senior management within the statistical agency. When developing it, it is essential to express clearly that the first priority of the SMS is to safeguard the content and methodological integration of statistical data and metadata.

13. The vision should be an integral part of the strategic direction of the statistical organization. It should define the major goals, benefits and functions of the SMS⁴, such as:

- (a) The metadata requirements associated with each element of standard business;
- (b) Ability to analyse existing statistical metadata objects and services;

³ See www.sdmx.org for more information on the Statistical Data and Metadata Initiative (SMDX).

⁴ More in-depth reading on SMS Vision is provided on the website of the Common Metadata Framework (CMF): <http://www.unece.org/stats/cmef/PartA4.html>

(c) Clear specification of what kind of existing metadata should be used in the corporate SMS (statistical variables and values sets, statistical surveys, social-economic classifications and nomenclatures, time series, statistical publications, statistical population, economic subjects, statistical units, aggregation and statistical evaluation methods, output tables and others);

(d) Cost/budget propositions of the SMS project;

(e) A metadata model complying with the functions of the statistical information system.

B. Objects for describing statistical metadata

14. Metadata should be structured according to the objects and the properties of those objects that they inform about. There are three major categories of metadata objects related to the functions defined by the vision:

(a) Statistical data and associated concepts;

(b) Statistical processes and associated procedures;

(c) Tools enabling production and usage processes.

15. Metadata objects related to statistical data and associated concepts are all important tools supporting production processes and the final use of statistical information. The most important objects are statistical concepts, statistical characteristics, statistical variables, population, classifications, registers, statistical surveys, time series, aggregation and statistical methods, micro data, macro data, final outputs, statistical publications, statistical databases and archives. Respondents, end users, statistical websites and other metadata objects related to statistical data, also belong to this group.

16. Inside statistical processes and related procedures, two major groups can be distinguished:

(a) Those associated with statistical production (data collection, data storage, data evaluation, data dissemination);

(b) Those associated with statistical information systems and statistical organization (planning and evaluation processes, supply processes, usage processes, total quality management and other management processes).

17. All processes as metadata objects are associated with several important metadata like costs, performance measures, errors and errors rates, diverse benchmarking indicators, etc. Processes are related to metadata in three ways:

(a) They are objects of metadata, carriers of metadata resources;

(b) They use metadata (about themselves and about other metadata objects);

(c) They produce metadata (about themselves and about other metadata objects).

C. Strategic plan for the statistical metadata system

18 The strategic plan should be developed and approved by all actors involved in the design, implementation and maintenance of the SMS, possibly with several iterations. The plan should give visibility, clarity and stability to the development efforts, but as aspects are likely to change during implementation, which may span several years, the plan should be regularly reviewed and revised.

19. Detailed plans should be developed and approved for the design and implementation phases of the SMS development. Such plans should reflect agreed priorities in the solution of individual components of the SMS. Last but not least, specific plans should be prepared for the phases following implementation: SMS use, maintenance and evaluation.

20. Plans should include the establishment of an organizational framework and management strategy. The recommendations below have been made by experts for establishing the strategic plan.

(a) Quality of data and metadata should be considered a high priority;

(b) External cooperation should be clearly defined; categorization and priority setting for external users should be specified. The plan should take the existing working plans of all external partners into consideration;

(c) The plan should be prepared in such detail that all partners will be able to commit their participation;

(d) External projects to establish data and metadata warehouses, both on the national and international level, should be considered for potential impact on the SMS;

(e) An integral part of the plan should be activities dealing with the development and implementation of international standards;

(f) The plan should also consider activities to promote the SMS and create an atmosphere of cooperation with all participating parties. To this end, prototypes to demonstrate SMS functions could be useful;

(g) Research activities on feasibility studies and analysis of user feedback should also be taken into the consideration when preparing an SMS plan;

(h) Transfer of know-how and training for participants in the SMS development process should be incorporated in the plan;

(i) Attention should be paid to change management in case of re-engineering of the SMS.

D. Corporate management of statistical metadata systems

21. Diverse organizational units of the statistical office (SO) and external bodies participate in metadata projects, including senior and middle management, subject matter statisticians, methodologists and IT experts. The projects should also reflect the feedback from respondents and end users of statistical information. The points below outline some recommended practices to ensure involvement in SMS management across the whole organization.

(a) The roles and responsibilities of all partners should be clearly defined, understood and followed. Where possible, automated workflows should be used to enforce agreed roles and responsibilities;

(b) The integration role played by SMS, both inside and outside the statistical office, and therefore the necessity of senior management involvement, should be clearly recognized when defining the SMS management strategy. Statistical activities are traditionally managed in the framework of individual statistical domains, tasks and/or projects in accordance with the organizational structure of the statistical office, which creates a strong need for integration measures;

(c) Metadata management is part of every project and should be considered alongside resource allocation and accountabilities, in the same way as business processes and data flows are considered;

(d) Establish SMS management boards to take an ultimate, corporate view on all decisions dealing with the SMS development;

(e) The SMS management strategy should be specified in close alliance with the existing managerial structure of the SO. Senior management should take a leading role in the SMS management model, and clear links should also be defined to the middle management level and the experts' level (methodologists, subject matter statisticians, IT experts);

(f) A multidisciplinary team should be the major organizational form for the development of SMS project. The ideal SMS team(s) will include: statistical methodologists; subject matter statisticians; dissemination specialists; end users; specialists in the implementation of statistical standards; researchers; and IT specialists in data modelling, business process design, architecture, and applications development;

(g) Implementation of a SMS management strategy may highlight the need for change in the organization of statistical activities, particularly where a corporate SMS does not exist. It is especially true for subject matter statisticians. Many critical moments could appear. Such moments should be as much as possible foreseen and reflected when progressing from the definition of the Vision goals and activities to the Vision plans.

III. CORE PRINCIPLES FOR METADATA MANAGEMENT

22. The following are core principles and recommendations for effectively managing the design and implementation stages of an SMS project.

- (a) Make metadata-related work an integral part of business processes across the organization;
- (b) Describe metadata flow with the statistical and business processes (alongside the data flow and business logic);
- (c) Ensure that customers are clearly identified for all metadata processes, and that all metadata capturing will create value for stakeholders;
- (d) Metadata presented to the end-users should match the metadata that drove the business process or were created during the business process;
- (e) Develop the SMS as a self-sustainable project, independent of any production systems;
- (f) The SMS is the definitive set of tools, stores and services to support metadata use and further development in the statistical office. If a metadata store, tool or service is not defined by the management to be a part of SMS, then it is not an “approved” metadata facility;
- (g) The diversity of metadata is recognised and there are different views corresponding to the different uses to which the data is being put. Different users require different levels of detail. Metadata appear in different formats depending on the processes and goals for which they are produced and used;
- (h) Make metadata active to the greatest extent possible. Active metadata drive other processes and actions will therefore be accurate and up-to-date;
- (i) Manage metadata with a life-cycle focus (including maintenance and update);
- (j) Preserve history (old versions) of metadata;
- (k) Capture metadata at their natural sources, preferably automatically as a by-product of other processes. Minimize errors by entering only once where possible;
- (l) Exchange metadata and use them for informing both computer based processes and human interpretation. The infrastructure for exchange of data and associated metadata should be based on loosely coupled components, with choice of standard exchange language, such as Extended Markup Language (XML);
- (m) All data and other objects of the SMS should be well supported by accessible metadata that are of appropriate quality;
- (n) Ensure that metadata are readily available and useable in the context of the client's information need (whether the client is internal or external);
- (o) There should be a single, authoritative source ('registration authority') for each metadata element;

(p) A registration process (workflow) should be associated with each metadata element, so that there is a clear identification of ownership, approval status, date of operation etc.;

(q) Reuse metadata where possible for statistical integration as well as efficiency reasons (no new metadata elements are created until the designer/architect has determined that no appropriate element exists and this fact has been agreed by the relevant 'standards area');

(r) A cost/benefit mechanism will help to ensure that the cost to producers of metadata is justified by the benefit to users of metadata;

(s) Variations from standards should be tightly managed/approved, documented and visible;

(t) Ensure systematic training and transfer of know-how for all partners involved. Train trainers.

IV. CORPORATE GOVERNANCE MODELS FOR METADATA MANAGEMENT

A. Lessons for good corporate governance of a SMS

23. It is not sensible to prescribe an ideal model for corporate governance of metadata. This is because every statistical organization works under different legislation, organizational arrangements, workplace culture, business rules, and levels of autonomy with respect to central public sector agencies. Therefore, the statistical community should look for 'good lessons' for governance. These may come from the Eurostat sponsored project Metanet, concluded in 2003, as well as case studies collected within the Common Metadata Framework (CMF).

24. What are some of the lessons for corporate governance of data and metadata management that have come from the experiences at national statistical agencies in the implementation of a metadata management strategy?

(a) The senior management group, including the Chief Statistician, should be very involved in policy formulation, approval of development projects and monitoring of outcome achievement. It is very helpful when your Chief Statistician and other senior executives ask questions about metadata matters;

(b) There is often scepticism in the organization against metadata projects. Moreover, metadata projects are usually strategic projects for the organization. If they should be carried out at all, managers on different levels and in different parts of the organization must be committed to the project;

(c) Clearly understood roles and accountability are needed for all organizational units with respect to metadata. The subject matter areas are responsible for the creation, maintenance, re-use, and approval for dissemination of all the data and metadata content for their statistical domain. A 'corporate data management unit' could be accountable to provide client support to subject matter areas, to develop and maintain infrastructure, to provide training, etc.

(d) The organization should develop an information management culture. That is, all staff should understand that it is their responsibility to work towards achieving the ideals of statistical integration, comparability of statistics across surveys and time, and to reuse statistical metadata as appropriate. These goals are achieved by adherence to the metadata management principles;

(e) Utilise existing governance arrangements to reinforce the metadata messages, that is, do not create new committees. Particular specialist staff e.g. business and systems analysts, IT architects, statistical standards experts, are more likely than others to come across new opportunities for advancing better metadata integration, so a particular focus is needed on working with these staff;

(f) Make sure that your organization has an endorsed metadata strategy, including a global architecture and an implementation plan, and that this strategy is integrated into broader corporate plans and strategies;

(g) Metadata projects are often more abstract, more complex, and more difficult to manage than most other types of projects. These characteristics need to be recognised in project plans, and the importance of communication with the rest of the organization about the project cannot be overstated;

(h) Make sure that your organization also learns from failures and successes in other statistical organizations. Benchmarking and international cooperation are always useful;

(i) Make systematic use of metadata systems for capturing and organizing tacit knowledge of individual persons in order to make it available to the organization as a whole and to external users of statistics;

(j) Either commit yourself to a metadata project – or don't let it happen. Lukewarm enthusiasm is the last thing a metadata project needs.

B. Potential risks and challenges

25. The Metanet and Accompanying Measure to Research and Development in Statistics (AMRADS) projects of the European Union explored adoption issues with respect to statistical metadata systems. Realization that there are potential barriers is an important part of the management and governance of such projects. Consideration of appropriate risk mitigation actions is a significant part of project governance. Potential barriers to the adoption of metadata solutions - technical, organizational and human have been explored in the framework of a special survey conducted for those purposes⁵.

26. National statistical offices were respondents to this survey. The questions aimed to identify in which area each of the potential problems were most important, as well as to go into more detail concerning the different aspects of human related issues.

⁵ See Chapter 7 of the report at http://www.epros.ed.ac.uk/metanet/deliverables/D7/IST-1999-29093_D7.zip

27. According to the results, the greatest challenges in relation to documentation and retrieval of data are considered to be partly organizational and human. On the other side, the technical challenges are the most important in relation to documentation of data for exchange and retrieval by IT systems.

C. Human resources and organizational problems

28. The human factor is fundamental to the successful adoption of metadata systems, yet a number of challenges have been identified. A loss of individual power as a result of providing metadata is one of the barriers. Other issues may be a low priority given to metadata by subject-matter specialists, the time it takes to provide metadata, and a perceived lack of resources.

29. If the barriers to effective metadata provision are to be overcome, the status of metadata activities must be elevated. This demands not only the education and active involvement of would-be providers, but also increased management awareness and support. There is a need to acquire input and feedback from subject-matter specialists from different areas regarding metadata/data concepts and methods in order to develop viable common metadata standards.

V. CASE STUDIES

30. Case studies about current metadata management systems and processes are being collected from statistical offices for publishing within the Common Metadata Framework (CMF). They will be published and maintained using wiki software that allows statistical offices to visit and update their own information and others to discuss it. The currently available case studies can be accessed at www.unece.org/stats/metis/wiki, and are structured under the following headings:

- (a) Introduction, strategy, current situation;
- (b) Statistical metadata systems and the statistical cycle;
- (c) Statistical metadata in each phase of the statistical business process cycle;
- (d) System and design issues;
- (e) Organizational and workplace culture issues;
- (f) Lessons learned;
- (g) Attachments and links.

31. At the time of writing this report the case studies from following countries are available: Australia, Canada, Croatia, Czech Republic, Ireland, New Zealand, Norway, Portugal, South Africa, Sweden, and the United Nations Industrial Development Organization (UNIDO). Case studies by other countries are envisaged.

32. Since there is no blueprint for design and implementation of a national SMS, the case studies represent a valuable source of information on practical SMS approaches and experiences.

The content clearly demonstrates a shift in importance away from the historical focus on ICT systems and design (metadata needed for building of statistical databases, data warehouses etc) to the functions and benefits of metadata across the statistical business process cycle. Furthermore, it demonstrates that statistical offices put more weight on links between the SIS and SMS.

VI. IMPLEMENTATION

33. The Common Metadata Framework (CMF) comprises other examples of implementation of statistical metadata systems by national and international statistical organizations⁶. At the time of writing these were:

- (a) Statline 4 metadata implementation (Netherlands);
- (b) Developing a System for Description of Microdata at Statistics Sweden (Sweden);
- (c) Implementation of MetaStore in OECD (OECD);
- (d) Development of a Metadata System at the Croatian Bureau of Statistics (Croatia);
- (e) Using SDMX standards for rapid dissemination of short-term indicators on the European economy (Eurostat);
- (f) Statistical metadata in Statistics Norway (Norway).

VII. CONCLUSION

34. The Conference of European Statisticians is invited to take note of issues of relevance to the executive heads of statistical offices, to recognise the strategic importance of metadata for corporate management of statistical offices, and to include the points (a) to (c) in paragraph 3 above in the report of the plenary session.

35. The members of the Conference are also invited to consider lessons that may be learned from the SDMX Content Oriented Guidelines for developing their statistical metadata systems. The members of the Conference may wish to consult other metadata related standards, and studies on the relationships between these standards and SDMX, that are available on the website of the Common Metadata Framework (www.unece.org/stats/cmf/).

* * * * *

⁶ <http://www.unece.org/stats/cmf/PartD.html>