

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

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

**ORGANISATION FOR ECONOMIC CO-OPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Work Session on Statistical Metadata (METIS)
(Geneva, Switzerland, 10-12 March 2010)

FINAL REPORT OF THE MEETING

1. The Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata was held in Geneva, Switzerland, from 10 to 12 March 2010. It was attended by participants from: Albania, Australia, Austria, Azerbaijan, Belgium, Bosnia and Herzegovina, Canada, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Italy, Japan, Kazakhstan, Latvia, Mexico, Mongolia, Netherlands, New Zealand, Norway, Philippines, Poland, Portugal, Russian Federation, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, The former Yugoslav Republic of Macedonia, Tunisia, Turkey, Ukraine, United Kingdom and the United States of America. Representatives of the Statistical Office of European Union (Eurostat), United Nations Educational, Scientific and Cultural Organization (UNESCO) Institute for Statistics, United Nations Industrial Development Organization (UNIDO), United Nations Secretariat Office for Information and Communication Technology (UNOICT), United Nations Statistical Institute for Asia and the Pacific (UNSIAP), United Nations Mission in Kosovo (UNMIK), Bank for International Settlements (BIS), International Monetary Fund (IMF), World Bank, Organisation for Economic Co-operation and Development (OECD), World Health Organisation, Paris 21 Secretariat and the United Nations Economic Commission for Europe (UNECE) also attended. Several experts also attended at the invitation of the UNECE Secretariat.

2. The agenda of the work session consisted of the following substantive topics:

- (i) Progress on Part B of the Common Metadata Framework: Metadata Concepts, Standards, Models and Registries;
- (ii) Progress on Part C of the Common Metadata Framework and the Generic Statistical Business Process Model;
- (iii) Progress on Part D of the Common Metadata Framework: Case Studies;
- (iv) Metadata and Quality Reporting;
- (v) 20 years of METIS.

3. The work session was chaired by Mr. Daniel Gillman (United States), and organized by the Steering Group on Statistical Metadata (Alice Born (Canada), Alistair Hamilton (Australia), August Götzfried (Eurostat), Daniel Gillman (United States Bureau of Labor Statistics), Ebbo Petrikovits (Czech Republic), Håkan Linden (Eurostat), Jana Meliskova (on behalf of the UNECE Secretariat), Jenny Linnerud (Norway), Jens Dosse (OECD), Joza Klep (Slovenia), Klas Blomqvist (Sweden), Marco Pellegrino (Eurostat), Matjaz Jug (New Zealand), Max Booleman (Netherlands), Trevor Fletcher (OECD) and Steven Vale (UNECE)). Members of the Steering Group acted as Session Organizers.

4. Ms. Lidia Bratanova, Director of the UNECE Statistical Division opened the work session and welcomed delegates. The expert group on statistical metadata was created 20 years ago following

approval by the Conference of European Statisticians at its 1990 session. She highlighted the very important work undertaken within this group over the past 20 years and which is highly appreciated by the CES and drew attention to the publication of Part A of the Common Metadata Framework which is already available in English and will soon be available in Russian. She also expressed appreciation for the support of partner organizations OECD and Eurostat. She concluded by wishing participants successful deliberations.

5. The conclusions reached during the discussion of the substantive items of the agenda are contained in the Annex. Presentations and papers for the meeting are available on the website of the UNECE Statistical Division (<http://www.unece.org/stats/documents/2010.03.metis.htm>).

6. In his closing address, the chairman highlighted the following main themes and conclusions of the work session:

- Quality has become an important issue in METIS discussions: not only quality of data, but quality of metadata as well. This is an issue that we are certainly going to see again in METIS discussions.
- Another important topic that emerged is the harmonization of standards. There was a particular interest in exploring the relationship between Data Documentation Initiative (DDI) and Statistical Data and Metadata Exchange (SDMX) standards.
- Looking at statistical production in terms of processes is an important development. The METIS Generic Statistical Business Process Model can be used as a reference for this work. It has been adopted by many organizations and now has the necessary degree of maturity.
- The increase in interest in the “semantic web” raises questions of interoperability, harmonization of data, and effective use of web services. These are things that the METIS group will continue working on in the future. It is important for these processes to be driven by people who know statistics well in order to avoid a technological bias that would result in solutions that don’t work in practice.
- Discussion of cases studies helps to increase the visibility of the METIS group and provides one of the best outputs of the METIS group. Case studies provide a useful opportunity to learn about what is working and what is not working in other organizations.
- The METIS community is growing; this will bring new organizational challenges, but is a clear sign of the importance of this work.

7. The participants adopted the present report before the work session adjourned.

ANNEX

SUMMARY OF THE MAIN CONCLUSIONS REACHED DURING THE WORK SESSION ON STATISTICAL METADATA (METIS 2010)**Topic (i): Progress on Part B of the Common Metadata Framework: Metadata Concepts, Standards, Models and Registries**

Session Organizers: August Götzfried (Eurostat) and Daniel Gillman (United States Bureau of Labor Statistics)

Documentation: Invited papers by the Task Force on Part B, Open Data Foundation, and United States Bureau of Labor Statistics; supporting papers by Eurostat, IMF, United States Federal Reserve Board, and OECD.

1. This topic considered the issue of metadata standards, with a specific focus on harmonization and inter-operability between standards. It highlighted recent developments and implementations of various standards, as well as possible synergies between them.
2. The report from the informal task force on Part B of the Common Metadata Framework presented progress done so far and raised questions to help finalize this work in the course of the coming year. This work has been undertaken following a strong demand from national statistical organizations and is carried out under the umbrella of the Conference of European Statisticians. The aim is to create a knowledge base (inventory) about standards relevant to statistical metadata and to offer statistical metadata system designers an overview of existing resources. Part B will be an internet publication, disseminated via the METIS wiki (www.unece.org/stats/cmf).
3. The presentation by Arofan Gregory of the Open Data Foundation discussed approaches to, and benefits of combining metadata standards. The presentation focused specifically on Data Documentation Initiative (DDI), Semantic Web and Statistical Data and Metadata eXchange (SDMX) standards, and how they can be combined to support the data production process. The presentation also proposed a formal endorsement by the official statistical community of the SDMX-DDI mappings that are now being produced and the publication of statistical classifications in standard formats to make them more usable. Proactive engagement with the Semantic Web was proposed to ensure that the requirements and standards of official statistics are not forgotten. Several statistical organizations are interested in applying SDMX and DDI together. The Australian Bureau of Statistics has recently started a project on this.
4. The presentation by the United States Bureau of Labor Statistics on the Semantic Web explored the possible relevance of this technology to official statistics. In this context, it has a clear need for statistical metadata, in particular describing concepts and relationships, though for the Semantic web this is in the form of ontologies, which are difficult to build. The goal to provide metadata for data discovery, closely linked to the data themselves, is not new and a key theme of work in METIS work sessions over many years. The Semantic Web does not, however, seem to offer any benefits in terms of data understanding and decision making.
5. Eurostat presented the on-going review of the SDMX technical specifications. This presentation included the outcomes of a meeting in Luxembourg (February 2010) to determine the requirements for a possible revision from version 2.0 to 2.1. The new version of the SDMX technical standards should fix bugs and inconsistencies, while providing solutions to well-identified use cases that are not currently supported. The SDMX technical standard should not become too complex to avoid blocking its further use and implementation, and causing higher costs.
6. The IMF presented recent developments in metadata in their organization. They have made several investments in systems for metadata collection, management and use. They have developed an SDMX Data Structure Definition for Balance of Payments and are working on one for government finance statistics. They are collaborating with other organizations to harmonize standard metadata,

particularly code-lists, but further content harmonization is needed. The SDMX Content-oriented Guidelines should be expanded.

7. The contribution from the United States Federal Reserve Board highlighted various issues relating to describing data sets from private sources and any associated restrictions on usage. Several organizations are looking at the possibility of applying Creative Commons licensing for statistical outputs, but it is not clear how appropriate this is for statistical data sets.

8. OECD presented information on publishing standards for datasets and data tables. Researchers use various tools to search across millions of articles, books, journals, etc., but it is very difficult for them to find the underlying data. In 2004, OECD launched the StatLinks service that enables readers to get hold of the data used in tables, charts and graphs in articles and book chapters by following a Digital Object Identifier link. Almost 1 million spreadsheet files were downloaded through StatLinks in 2008, a strong proof of the readers' interest in getting hold of the original data. It is therefore increasingly important to provide links to data from articles and to develop international standards for cataloguing datasets. Libraries are starting to establish groups to address this and find better ways to cite and discover data within cataloging systems.

9. The following points were raised during the discussion:

- Work on Part B of the Common Metadata Framework was seen as useful and important, particularly the description of relationships between standards. It was suggested to add a description of the Portuguese model, as other organizations are planning to adopt this approach.
- There are different views on the Semantic Web, with some participants favouring a “wait and see” approach, while others want to be more proactive. Some considered that it does not offer much that is new and there is a risk of reinventing something that already exists. The usefulness of publishing data according to the RDF (Resource Description Framework) data model was questioned. However, if statistical organizations do not get involved now, it will be difficult to ensure our standards and requirements in areas such as data confidentiality are properly taken into account.
- There is a clear need for enlarging and integrating technical standards and to see how far SDMX can be applied within the statistical production process. There is growing interest in using DDI and SDMX together, with DDI seen as more appropriate for micro-data and SDMX for macro-data and dissemination.
- Further integration and expansion of content standards, such as the SDMX Content-oriented Guidelines and Eurostat/IMF quality frameworks, is seen as essential.
- To improve accessibility, multidimensional tables should be user-driven rather than production-driven, as users are increasingly interested in cross-cutting issues such as sustainable development.

Topic (ii): Progress on Part C of the Common Metadata Framework and the Generic Statistical Business Process Model, Session Organizers: Jenny Linnerud (Norway) and Steven Vale (UNECE)
Documentation: Invited papers by Australia and Denmark; supporting papers by Nordic / Netherlands Task Force on Business Architecture.

10. This session considered the development of the Generic Statistical Business Process Model (GSBPM), which is fundamental for discussing the organization of statistical metadata, as well as the implementation of this model in statistical organisations.

11. The presentation by Australia focused on their new Information Management Transformation Programme, which aims to improve the efficiency of statistical production through the application of common standards. Metadata management is a major part of this programme as metadata are seen as key to process improvements. The main standards to be used are SDMX and DDI, within the framework of

the GSBPM. The Australian Bureau of Statistics is seeking a comprehensive, consistent approach to ensure an end-to-end management of the statistical production process.

12. Denmark presented recent work to apply the GSBPM as a framework for statistical production. Standardization has become an important goal to help meet efficiency targets. The GSBPM has been used as a framework to analyse several production processes and has helped to identify where improvements can be made. The vision for the future is to create a single metadata system covering the entire process, alongside a data model based on three data stages (raw data, micro data, macro data). This work has taken approximately two person-years so far.

13. The presentation by the Nordic/Netherlands Task Force on Business Architecture described the goal of facilitating common development and/or sharing of IT solutions and software tools amongst the participating organizations. The GSBPM has been used as a framework to aid the comparison of current business architecture models, and the standard terminology it provides has been particularly helpful. None of the countries have fully adopted the GSBPM, but their business process models are sufficiently close to allow relatively easy mappings. The Task Force suggested that sub-process descriptions could also address general problems rather than particular solutions.

14. The following points were raised during the discussion:

- The GSBPM encourages standardization and better documentation, and can be used to look at existing metadata systems, identify gaps and determine how complete the overall metadata system is.
- There is a need for more focus on phases 1 and 2 of the GSBPM and to provide statisticians with better metadata tools for these phases.
- The ways of working and results of the Nordic/Netherlands cooperation work could be more broadly applicable to other organizations.
- The current level of detail in the GSBPM is seen as sufficient for its role as a reference model. National implementations can add more detail as necessary.
- The GSBPM should not be updated too frequently. More implementation experiences should be gathered, with a possible (minor) update based on these after at least two years.
- If the model diagram can be made to look less linear, this may help understanding in statistical organizations.
- An inventory of implementations, building on work from the CORA (Common Reference Architecture) project, would be useful, perhaps as part of a set of on-line resources to encourage the establishment of a user community.
- The links between the GSBPM and process quality management are also important and could be explored further.

Topic (iii): Progress on Part D of the Common Metadata Framework: Case Studies,

Session Organizers: Alice Born (Canada) and Joza Klep (Slovenia)

Documentation: New case studies from Finland, Latvia and the Netherlands: Updates from Sweden and Germany: Supporting papers from Eurostat, World Health Organization, Canada / Slovenia, and Australia.

15. This session considered national experiences of the implementation of statistical metadata systems, including analyses of the contents and usefulness of the existing case studies. It also considered new developments in statistical metadata systems in international organizations.

16. Statistics Finland presented their new case study. Their statistical metadata system is in a transitional phase from relational databases to an XML-based environment. A new XML database is being introduced as the primary repository for metadata, starting with a variable editor and a classification editor. However, at the moment the relational databases still feature strongly in the metadata architecture.

17. The Central Statistical Bureau of Latvia presented a case study on their integrated metadata driven statistical data management system. The case study presents the development and implementation of the Integrated Metadata Driven Statistical Data Management System (IMD SDMS), which covers the whole statistical production process. This system is process oriented and metadata driven. They have mapped several business statistics processes to the GSBPM and found a close fit, though there may be a need to expand the number of sub-processes in the “Collect” phase. A new project will extend this work to social statistics.
18. The Statistics Netherlands case study showed that although their business architecture is more detailed than the GSBPM, it is fairly easy to map to it. They have introduced the concept of “steady states”, which are data sets (together with the necessary metadata) at different points in the statistical production process. Every process has at least two steady states, and this concept has helped to reinforce the business case for metadata systems.
19. Statistics Sweden presented an update of their case study. Their overall strategy is based on process oriented production, focussing on customers, and increasing efficiency and standardization. There is no separate strategy for metadata; it is included in the overall strategy. The goal is to have an integrated metadata system, and common metadata for the whole production process.
20. An update of the case study on metadata management in the German official statistics system was presented by Destatis. It reviewed the situation in Germany, where the GSBPM has now been adopted as a central standard, facilitating cooperation between the federal and regional statistical organizations. A new classification server is being developed in an IT project led by the Bavarian statistical office. It will serve as a reference database for classifications and as a standard IT tool for semi automatic coding. A new SDMX based standard for the exchange of classifications is needed. A census variable server based on the Neuchâtel model will also be developed.
21. Eurostat presented their Metadata Handler which is intended to be the backbone of the data-driven harmonized statistical business process in the European Statistical System. Harmonised code-lists will be added and it will be possible for countries to upload national data structure definitions. The National Reference Metadata Editor will be an online tool available to national producers, to allow the production and transmission of metadata to Eurostat. These tools are part of the new Eurostat vision to improve statistical production.
22. The WHO presentation raised the issue of varying definitions from multitude of guidelines documents and sources. Standards have been developed for the SDMX-Health Domain, including computer-readable indicator definitions and aggregate data exchange standard for monitoring systems. The WHO Indicator and Metadata Registry (IMR) is an intersection between IT people and data analysts to share methodological information. It is available online on the WHO website.
23. The Australian presentation considered how case studies have proved useful beyond the METIS community. It is interesting to look at value of case studies and how the information is used. Should case studies be regarded as akin to “professional” blog entries, “professional papers” or official statements by NSOs? They can be used to prepare for international visits and also for e-government purposes. It may be necessary to consider holding them also in national languages.
24. Slovenia and Canada presented an analysis of existing metadata case studies, giving an update of what was presented last year. There are currently 17 case studies, increasing in number each year. The analysis focused on metadata strategies, system and design issues, organizational issues and lessons learned. Almost half of the countries are using or planning to use SDMX and 29% indicated that they are using the GSPBM. Countries are encouraged to post further case studies in the METIS wiki.
25. The following points were raised during the discussion:

- Country quality reports sent to Eurostat are not posted automatically on the web, but are discussed with subject-matter experts and relevant working groups.
- The Eurostat SDMX registry may eventually hold mappings between data structure definitions, but this is currently a low priority.
- Case studies should not be regarded as official statements, more as professional blog entries, so that they retain their usefulness, and can be updated more easily.
- The UNECE secretariat will look into the possibility to version the case studies.

Topic (iv): Metadata and Quality Reporting

Session organizers: Jana Meliskova (on behalf of the UNECE Secretariat) and Klas Blomqvist (Sweden)

Documentation: Invited papers by Eurostat, Italy, Czech Republic, and Australia: Supporting papers by Canada, the Former Yugoslav Republic of Macedonia, and Lithuania.

26. This session considered the sharing of ideas and experiences, to help statistical organizations to increase efficiency in monitoring the quality of statistics and preparing their quality reports. It focused on how metadata relating to data quality can be better embedded into international quality reporting systems.

27. Eurostat presented information on the links between quality reporting and metadata within the European Statistical System (ESS). Reporting on statistical data quality exists in many domains, but the new Eurostat vision on improving the production method of European Union statistics requires that this is extended. A new reporting structure for quality related metadata has been created: the ESS Standard for Quality Reports Structure (ESQRS). The break-down of the main quality indicators into attributes that are easily reported was seen as something that could possibly be adopted as a wider standard.

28. Italy presented work on metadata and quality indicators reuse for quality reporting in their information system on survey documentation. This system documents all phases of statistical production using a model similar to the GSBPM, and can be used to produce pre-filled reports, including quality reports. The system is designed to document processes rather than variables, so links to data are limited. Layers of the model specify quality activities and software used for each sub-process.

29. The Czech Republic presented their quality metadata system, which aims to provide support for quality reporting, and improve the management of quality and statistical business processes. The development of the quality metadata system is a high priority for the next two years. The development of quality methodology on both national and international levels will be taken into account.

30. The Australian Bureau of Statistics presented their work on linking quality assessment to development of performance indicators, based on a data quality framework adopted in May 2009. This framework is also used by the Council of Australian Governments. An online tool has been designed to allow users to make quality statements. It provides a series of questions to guide users on the content.

31. The paper on quality indicators and the role of metadata repositories at Statistics Canada was presented with an emphasis on quality management and improving data quality. A number of quality indicators have been developed at the aggregate and survey programme levels. Statistics Canada has updated its quality assurance framework based on the GSBPM, and this is helping to identify the sub-processes with the greatest risk of introducing errors.

32. The former Yugoslav Republic of Macedonia reported on linking quality assessment to development of performance indicators, increasing the availability and reusability of reference metadata. This has been facilitated by the introduction of a central metadata repository based on the Euro-SDMX Metadata Structure (ESMS) standard. Dynamic reports can be generated for quality reporting to international organizations.

33. The paper by Statistics Lithuania was summarized by the session organizer. It discusses the objectives of the Integrated Statistical Information System (ISIS), the structure of metadata in ISIS, and the management of metadata.

34. The following points were raised during the discussion:

- The need to increase the efficiency of the production of quality reports, as the requirements for quality reporting are growing.
- Balancing user needs and producer burden would be facilitated by the convergence of reporting requirements.
- There is a close link between metadata and quality, which could be developed in the context of the biennial “Q” conferences. These are currently held the same year as the biennial METIS work sessions, but it may be preferable to alternate them.
- The aim of quality reporting should be to improve the statistical business process, and the quality of output data. As a first step, quality reports help to ensure transparency, so if they are applied to more statistical domains, they can help to identify which statistical processes work best.
- Quality reporting can be seen as an application of metadata reporting. It therefore provides a strong use case for metadata systems.
- It is important to intensify cooperation between experts who focus on metadata and those who focus on quality.
- The quality of metadata is an increasingly important consideration. Good quality metadata are needed to find, use and understand statistical information, so improving metadata quality is an important consideration for statistical organizations. It would be interesting to discuss further about whether the components of data quality also apply to metadata.
- Publishing metadata is a good way of improving their quality, as this gives subject-matter statisticians more incentive to produce good quality metadata.

Topic (v): 20 years of METIS

Session organizers: Daniel Gillman (United States Bureau of Labor Statistics) and Steven Vale (UNECE)

Documentation: Invited papers by the United States Bureau of Labor Statistics, and the Conference of European Statisticians Steering Group on Statistical Metadata.

35. This session reviewed the achievements over the 20 years of the Work Sessions on Statistical Metadata, and discussed the future work programme of the Conference of European Statisticians Steering Group on Statistical Metadata.

36. The chair of the Steering Group on Statistical Metadata reviewed the work and progress of the group over the last 20 years. METIS work sessions are attracting increasing numbers of participants from both national and international statistical organizations, and now have a global following. Technological changes have had a large impact on both metadata systems and the organization of the work sessions. The importance of harmonized terminology has been a consistent theme, though there has been a gradual shift from formalism and theoretical models to practical implementations.

37. The UNECE Secretariat presented the governance structure and priorities for future work of the Steering Group on Statistical Metadata:

- The METIS wiki should be expanded in scope and features, and made more interactive.
- Maintenance arrangements for the Common Metadata Framework need to be formalised.
- No further work is necessary on Part A in the near future.
- Development of Part B is a clear priority, to support and influence international work on the harmonization of metadata contents, formats, systems and standards.

- The extension of Part C to include information on the practical implementation of the GSBPM and its application to process and quality management is a priority for many delegates. An informal group to progress this work was proposed.
- The Steering Group proposed to alternate METIS work sessions with the “Q” conferences on statistical quality.
- A METIS workshop is proposed for autumn 2011 to consider the practical implementation of the GSBPM, and process-related and quality-related metadata within each phase of the model. This workshop may also include a training course in a topic such as the DDI standards.
- The next METIS Work Session will be in spring 2013. Possible agenda items include; harmonization of metadata standards, classification database systems, quality of metadata. New case studies are invited, and work is proposed to study how to further improve their use and usefulness
- Possible locations for the Workshop and Work Session will be considered by the Steering Group. Offers to host either of these events are very welcome.
- The Steering Group will be reconstituted after the meeting, with new terms of reference.

38. During the discussion it was stressed that it is important to reuse quality metadata already provided elsewhere. Quality is directly connected with documentation, and assessment of quality helps to improve the statistical production process. Metadata should not only be connected to quality reporting, but also to quality assessment. A task force on quality metadata for particular purposes was suggested.