

Distr.
GENERAL

CES/AC.71/2004/2
2 June 2004

Original: ENGLISH

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

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

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

Joint ECE/Eurostat/OECD meeting on the management of statistical information systems
(Geneva, 17-19 May 2004)

**REPORT OF THE MAY 2004 JOINT ECE/EUROSTAT/OECD MEETING ON
THE MANAGEMENT OF STATISTICAL INFORMATION SYSTEMS**

Prepared by the ECE secretariat

1. The Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems was held in Geneva, Switzerland, from 17 to 19 May 2004. It was attended by participants from: Australia, Austria, Belgium, Bulgaria, Canada, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Kazakhstan, Latvia, Lithuania, New Zealand, Norway, Poland, Portugal, Romania, Russian Federation, Slovakia, Slovenia, Spain, Sweden, Switzerland, The former Yugoslav Republic of Macedonia, Turkey, United Kingdom, and United States of America. The European Commission was represented by Eurostat. Representatives from the following international organizations also attended: United Nations Statistics Division (UNSD), United Nations Educational, Scientific and Cultural Organization (UNESCO), Organization for Economic Cooperation and Development (OECD), United Nations Food and Agriculture Organization (FAO), World Trade Organization (WTO), the International Monetary Fund (IMF), the European Central Bank (ECB), and the Bank for International Settlements (BIS).
2. The representatives of ECE, Eurostat and OECD addressed the meeting at the opening session. They summarized the priorities of their organizations as established by their governing bodies. These priorities show that all three organizations attach great importance to the management of statistical information systems, as a necessary pre-condition for improving the efficiency and quality of their statistics.
3. The agenda of the meeting consisted of the following substantive topics:
 - (i) Web technology in statistical information systems;
 - (ii) Development of IT strategies in statistical offices;
 - (iii) Open source in statistics;
 - (iv) Annual review of the website of IT practices in the National Statistical Offices;
 - (v) Review and follow-up to the activities of the Conference of European Statisticians.
4. Mr. Mel Turner (Canada) acted as Chairman of the meeting. The preparation of topics (i) to (v) above was organized respectively by: Mr. Marton Vuicsan (Netherlands); Mr. Dayantha Joshua (United Kingdom); Mr. Richard Swartz (United States); Mr. Lars Thygesen (OECD); and Mr. Juraj Riecan (ECE).

RECOMMENDED FUTURE WORK

5. The participants considered it useful for national and international statistical offices to continue the exchange of experience in the field of management of statistical information systems. Therefore, they recommended that a joint ECE/Eurostat/OECD meeting on the Management of Statistical Information Systems (MSIS) be convened in 2005, subject to the approval of the Bureau. The meeting would continue to review the activities of the Conference of European Statisticians related to statistical information systems, under the leadership of the Steering Group on the Management of Statistical Information Systems.
6. The following substantive topics were suggested for discussion at the next meeting:
 - (a) IT governance in statistical offices (aligning business and IT strategies, value of IT, macro- and micro-level organization, etc.);
 - (b) Development strategies for statistical information systems (metadata-driven systems, information systems architecture, data collection, metadata standards, dissemination, etc.);
 - (c) XML and web services (progress in SDMX standards, statistical tables with complex metadata, archiving aspects, practical experiences with web services, use of XML tools inside national statistical offices, etc.)
7. The participants also identified the following topics of interest for future discussion:
 - (a) Technical aspects of access to microdata;
 - (b) Software ergonomics and accessibility;
 - (c) Geographical mapping for presentation of statistics (thematic maps, GIS, etc.).
8. It was further suggested that the Steering Group consider the inclusion of a discussion on the outcomes of research projects (e.g. 5th framework of the EU) and their applicability to the statistical information systems in statistical offices on the agenda of a future meeting.
9. The participants agreed to discuss the content of the draft questionnaire for the International Survey on the Role of National Statistical Offices in e-Services and to try to fill-out the questionnaire. They will transmit the completed questionnaires and their comments to the ECE secretariat by the end of July 2004.
10. The meeting found that the impact of the website on IT practices in national statistical offices was minimal, because it did not reach the critical mass of information. Therefore, the participants agreed that the present website would be closed down and that its original content and new practices in statistical information systems would be incorporated into the web-based documents library of the UNECE Statistical Division (see: <http://www.unece.org/stats/sis/> and <http://www.unece.org/stats/archive/>).

ORGANIZATIONAL ASPECTS OF THE MEETING

11. The participants discussed the possible hosting of future meetings by national statistical offices. The benefit of this practice would be in providing a better and more in-depth knowledge of visited offices, and in making practical presentations. On the other hand, the necessity of installing the conference equipment permitting simultaneous interpretation in three languages may create difficulties and a financial burden on the member countries. In order to encourage participation from the CIS countries it was suggested to consider a two-language interpretation (English-Russian), which would decrease the complexity while encouraging the participation from the CIS countries.

ADOPTION OF THE REPORT

12. The participants adopted this report and the summary of the main conclusions of the Meeting (as presented in the Annex) at its closing session.

Annex

SUMMARY OF THE MAIN CONCLUSIONS REACHED BY THE PARTICIPANTS

Topic (i): Web technology in statistical information systems

Session Organizer: Marton Vuicsan (Netherlands); Discussant: Juraj Riecan (UNECE)

Documentation: Invited papers by Austria, Canada, and OECD; supporting papers by Germany, Italy, Norway, Slovenia, and United States

1. The participants discussed the use of web technologies in collection and dissemination of statistical data. The use of web technologies for data sharing among international organizations, aimed at decreasing the reporting burden on member countries, was also discussed.
2. The use of web technologies is more typical for the collection of data for enterprises than for households, but examples of the 2002 Canadian Census and the United States school survey were presented at the meeting. In both cases, the use of web data collection cannot be imposed on the respondents, but is one of the choices. While in the case of enterprises, data collection may be recurrent and the use of security certificates and encryption software does not seem to be a burden, in the case of population censuses, the present certification methods seem to be too onerous. Therefore, new ways of providing high security are being considered, while keeping the certification process transparent to households and private individuals (e.g. through the re-use of certificates in the public key infrastructure (PKI)).
3. Statistical data confidentiality is a very sensitive issue. It is important for the statistical agencies to maintain the trust of respondents. Therefore, the security should be greater than the standards for electronic business. This is primarily important for the collection and exchange of microdata using web technologies. Disclosure protection is also important for the dissemination of macrodata, because the use of web technologies provides greater possibilities to "hackers". The IT architecture should, therefore, provide a high level of protection for internal source databases and for sensitive information.
4. The use of web technologies provides a new dimension to application development, as practically each web form, needs its own individual processing. Therefore, approaches are being developed to create generic tools, generating the processing scripts, web forms and other components needed for surveys (ST.AT in Austria, Teleform in Italy, GENESIS in Germany). These are usually metadata driven. Formats for metadata and data, description of multidimensional arrays and time series, were also discussed. A significant achievement was made through SDMX-ML, which is a standard that will allow efficient data sharing based on a generic data and metadata model and using XML as syntax. SDMX-ML is intended for web services when data are available in a machine-readable form for re-use by other partners of the interchange network.
5. Data sharing seems to be more typical for international statistical work, and some participants had doubts about its importance at the national level. Other participants saw the potential in data sharing at the national level, for example within the research community.
6. The user friendliness, and simplicity of use is also important, to attract more respondents and users of the web-based data collection and dissemination. Both active user/respondent interaction and file transfer are used in statistics. Concerning the IT techniques, the use of Java, asp.net, cgi and XML were reported at the meeting, but more traditional formats (CSV, etc.) are also still in use. The majority of examples seemed to prefer server-side processing, which avoids the need of downloading special software, and which permits the use of a Java-enabled standard browser.
7. Pre-filling of the web questionnaires may be one of the motivations for the respondents to make the use of the web option attractive. However, in some cases, possibilities of pre-filling may be restricted. Another motivation may be the growing popularity of the Internet. Some offices estimate that when providing the web option in the population census, some 30% of respondents may choose this option. In the case of businesses,

the programmatic interfaces may be more appreciated than simple web forms. Advertising of web services was also discussed, and ebXML and UDDI were cited as examples.

8. In some data dissemination approaches, the data are published as a free service, in other cases a subscription is needed. There was not a sole opinion on what was better. In the case of a subscription service, it is easier to have statistics of the use of the subscription service, and the subscription does not necessarily imply a paid service. On the other hand, for official statistics, the tendency is towards a free unrestricted access.

9. When discussing dissemination, the issues related to metadata repositories and dissemination databases were considered. PC-AXIS, as a family of products providing the necessary capabilities, was mentioned. While PC-AXIS is used by several countries, other approaches were also presented, namely GENESIS and ConIstat. The management of metadata was discussed in connection with all of the dissemination activities.

10. Web technologies have created a sort of a revolution in our lives in the past decade. However, some participants stressed that the future development provide greater opportunities for statistics. The opinion was expressed that the web represents another, new modality for statistical data collection and dissemination. At present the costs savings are not so visible, but may grow in the future. Another opinion cited decreased costs for inquiries.

Topic (ii): Development of IT strategies in statistical offices

Discussant: Dayantha Joshua (United Kingdom)

Documentation: invited papers by Latvia, and OECD; supporting papers by Armenia, Croatia, Italy, Russian Federation and Sweden.

11. The discussion covered a range of topics related to the reorganization of information systems in statistical offices. These included metadata driven systems, lessons learned from migration from mainframes to micro computers, implemented new IT architecture and integrated statistical information systems.

12. Metadata may be considered as a key to statistical information, and therefore, are in the centre of strategies for national and international statistical agencies. The metadata systems help to describe the data, and underlying methodologies, as well as the processes applied to the data owned by statistical offices. Metadata also represent an integrating element within statistical information systems. The central focus should be on the organization and management of metadata. The following principles may be considered, along with other considerations: (i) Statistical metadata must be consistent; (ii) Metadata must be created only once; (iii) For each dataset there must be someone responsible for the metadata.

13. Statistical production can be decentralized to several substantive departments, while the statistical agency's aim is to provide a coherent corporate statistical system. This requires a more systematic approach to application development. Formal approaches were presented at the meeting, aiming at better applications development, project management, test management and test improvement processes. The systematic approach would help to ensure that individual modules could properly interoperate, and that redundant developments are avoided.

14. The present statistical information systems encompass all aspects of the Organization's statistical processes: data and metadata collection, validation, processing, storage, discovery and retrieval, and dissemination. Three working layers may be considered: a production layer, a storage layer and a dissemination layer. These layers are obviously interoperating through a workflow system. The integration and consolidation of the information system brings with it the necessity to manage large databases.

15. The importance of using standards may improve the coherence of the system, as well as facilitate cooperation with external partners. There are typical technology standards, such as XML, Web services etc. There are also standards related to the statistics, like SDMX, various classifications and nomenclatures, standard tabulations, etc.

16. The motivation of statisticians to move from the present to a new data processing environment is essential when designing and implementing a new IT strategy. On the other hand, it is important to involve all IT staff in discussions leading to the development of a new IT strategy. There are numerous challenges related to the change of management with regard to skills, making the staff move forward, people management, relationship with trade unions, etc. There are no easy answers to the question of how to manage the change.

17. One possibility is in the outsourcing part of the operations. However, when outsourcing part of IT operations, it is important to identify what activities are to be outsourced, and what are to be provided by internal staff. There are significant differences - some offices turn towards outsourcing, while others manage mostly with the internal resources. The opinion was expressed that the core business should not be outsourced. Examples were given when the information systems management is provided by another government institution, which is not true outsourcing but permits concentrating the necessary IT skills in a specialized institution.

18. The IT practitioners in national statistical office face the following challenges: to keep pace with technological changes; to provide technological leadership to our statistical colleagues; and to standardize and consolidate the information systems processes. The statistical information systems should be based on strong metadata and classification disciplines. Keeping pace with the web developments, such as content management systems and their links to consolidated systems, is also an important part of present IT strategy in official statistics.

19. In concluding the discussion, the participants emphasized that information management should: be responsive to demands for modernization, acquire the support staff with the right skills, create the technological basis for business change and manage the standardization without compromising flexibility.

Topic (iii): Open source in statistics

Discussant: Richard Swartz (United States of America)

Documentation: invited papers by Norway, Sweden and Eurostat; supporting papers by Azerbaijan and United States of America.

20. There are many examples in the IT world showing the growing use of open source software (OSS). For example Linux has gained the interest of users. The use of open source software is subject to policies set up by the national statistical office or more general policies of the government administration. The participants agreed that most of the OSS examples are quite general, and that the future considerations at MSIS should be focused on the applications where the statistical community is unique.

21. The Statistical Open Source Standards Group (SOS) was created in the late 1990s with the aim of establishing a forum of common interest in software development close to the needs and strategies of national statistical offices. It is based on simple principles facilitating the exchange of experience, re-use of components and minimizing the responsibilities, but it does not make a provision for common funding.

22. The shared development like PC-AXIS, a family of data dissemination software products, is a good example, which brings benefits to participating statistical offices, even if it is not an open source product in the strict understanding. The PC-AXIS Group represents a software consortium rather than an example of typical open source software. Members of the consortium share the development costs according to a price list taking into account the size, but also financial possibilities of the country.

23. Another shared statistical dissemination system is used for Census and FedStats in the United States. Applications such as "Census QuickFacts" and "FedStats MapStats" were cited. Open source provided opportunities for the two organizations to share a single application, while maintaining different hardware platforms.

24. The open source software does not imply free software. The value of using open source is not in cost saving but in the possibility of focusing, reusing and switching the components. Therefore, the potential can be

found in the wider sharing of the source codes between statistical offices. Participants also gave some examples of the use of open source software for integrating non-open source software.

25. The discussion suggested that open source principles (recommended leading practices) for statistical purposes may need to be spelled out and agreed upon. While there is a long-lasting tradition of attempts to share statistical products, there is still a need to have closer cooperation within the international statistical community. One of the obstacles seems to be using conceptually different models, and therefore, the software sharing may have to begin with an agreement on common concepts and models.

26. In concluding the discussion, the participants listed some of the issues that need to be clarified with regard to open software:

- Production of open source software by national statistical offices;
- Use of Linux as a server side system;
- Use of Linux as a desktop operating system;
- General purpose open source software useful in statistical applications;
- Support issues.

Topic (iv): Annual review of the website on IT practices in national statistical offices

Discussant: Lars Thygesen (OECD)

Documentation: invited paper by OECD

27. The use of the website is minimal. One of the main reasons for this situation may be that the website did not reach the critical mass of information. A possible way out of this situation may be to merge the website's content with other similar collections of methodological materials and leading practices. The AMRADS project was cited as an example in this connection.

28. Maintaining such an ambitious project would require putting in place a management structure comprising website managers and theme managers for individual themes. Animating the website would require a commitment, as the animation of such website is similar to publishing a scientific journal.

29. In this connection the participants mentioned the example of the UNECE website, which preserves a large collection of documents. They expressed their appreciation of the stability of the website as a repository of substantive documents arising from the UNECE sponsored (and joint) meetings. The participants also asked the UNECE secretariat to continue maintenance of this repository, and keeping the addresses (URLs) stable. In the last year few thematic pages were added to the subject classification of the UNECE's documents library on the Internet. These thematic pages do not correspond to a concrete meeting, but they represent a collection of technical papers produced by various working groups and task forces. This may allow continuing the website in a modest and simple form. This was the conclusion accepted by the participants after the discussion.

Topic (v): Review and follow-up to the activities of the Conference of European Statisticians

Discussant: Juraj Riečan (UNECE)

Documentation: background paper by the UNECE and reports of meetings related to statistical information systems.

30. The participants agreed that the number of meetings (groups) with recurrent activities, and which are related to statistical information systems and data processing and computing methodologies, exceeds the possibilities of available resources. These activities cover the following topics: (i) management of statistical information systems; (ii) technical aspects of statistical confidentiality and disclosure protection; (iii) statistical data editing; (iii) electronic data reporting; (iv) statistical metadata and (v) geographical information systems. Therefore, streamlining of the activities is needed. However, the participants pointed out that each of these topics deserves the attention of experts from national statistical offices, and instead of abandoning some of the topics, these may be included on the agendas of other related meetings.

31. In the field of metadata the work will concentrate, in the near future, on creating a common framework for statistical metadata with a focus on management, functions and use of metadata in statistical offices. A Task Force is being put in place to coordinate this work. In the field of statistical data editing a new publication is under preparation on the impact of data editing and imputation on data quality.

32. The last meeting on technical aspects of confidentiality and disclosure control took place in 2003 (jointly organized by UNECE and Eurostat), and Eurostat has recently published the proceedings of this meeting. The participants recommended that these issues might be taken up also by one of the future MSIS meetings (see recommendations in paragraph 7 of the main report).

33. The participants pointed out that geographical information systems represent an important methodological topic related to statistical information systems. This is treated somehow separately because of numerous **specificities** (specifications?) of GIS databases, geo-referencing, remote sensing, cartographic methods, etc. However, they would prefer incorporating GIS-related topics into other activities of the Conference rather than abandoning them. In this respect the UNECE and Eurostat Secretariats were asked to clarify the difference between the joint meetings and the upcoming Eurostat's meeting. The participants stressed that advanced GIS techniques from countries outside the European Union are of interest for the exchange of experience.

34. In concluding their discussion, the participants suggested streamlining the activities without leaving important issues out, and therefore suggested combining related topics within fewer meetings.

International Survey on the role of national statistical offices in e-Sevices

Discussant: Mel Turner (Canada)

Documentation: draft questionnaire prepared by Statistics Canada

35. The participants discussed the proposed questionnaire presented by Statistics Canada. The questionnaire is primarily designed for national statistical offices and some of its content does not apply to international organizations. The suggestions made at the meeting concerned aspects of metadata, censuses, etc.

36. The participants agreed to discuss the contents of the questionnaire within their organizations and try to fill-out the questionnaire. They will address the completed questionnaire and their comments to the UNECE Secretariat by the end of July 2004.

- - - - -