

Distr.
RESTRICTED

CES/SEM.47/2
10 May 2002

Original: ENGLISH

**STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

**COMMISSION OF THE
EUROPEAN COMMUNITIES**

CONFERENCE OF EUROPEAN STATISTICIANS

EUROSTAT

**Joint UNECE/Eurostat Seminar on Integrated Statistical
Information Systems and Related Matters (ISIS 2002)**

(17-19 April 2002, Geneva, Switzerland)

**REPORT OF THE APRIL 2002 SEMINAR ON INTEGRATED
STATISTICAL INFORMATION SYSTEMS**

1. The Joint UNECE/Eurostat Seminar on Integrated Statistical Information Systems and Related Matters was held in Geneva, Switzerland, from 17 to 19 April 2002. It was attended by participants from: Armenia, Austria, Azerbaijan, Belgium, Bulgaria, Canada, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Kazakhstan, Latvia, Lithuania, Netherlands, Norway, Poland, Romania, Russian Federation, Slovakia, Slovenia, Sweden, Switzerland, The former Yugoslav Republic of Macedonia, and United States of America. The Republic of Korea participated under Article 11 of the Terms of Reference of the Economic Commission for Europe. The European Commission was represented by Eurostat and the Enterprise Directorate General. Representatives from the following international organizations also attended: United Nations Statistical Institute for Asia and the Pacific (SIAP), Organization for Economic Cooperation and Development (OECD), Food and Agriculture Organization (FAO), the World Bank, the World Trade Organization (WTO) and the European Central Bank (ECB). Representatives of the University of Athens also participated at the invitation of Eurostat.

ORGANIZATION OF THE SEMINAR

2. The programme of the seminar consisted of the following substantive topics:

- (i) Application of web technology to integrate statistics;
- (ii) Secure communications and data confidentiality;
- (iii) Object-oriented technologies, component architecture;
- (iv) Ways of making statistical information systems more responsive to users.

3. Mr. Mel TURNER (Canada) acted as Chairman and Mr. Wolfgang KNUEPPEL (Eurostat) as Co-Chairman. The discussion on topics (i) to (iv) above was organized by: Mr. Jan BYFUGLIEN (Norway); Mr. Richard SWARTZ (United States); Mr. Georges PONGAS (Eurostat) and Mr. Marton VUCSAN (Netherlands).

4. Recommendations for future work are given below. Other conclusions that the participants reached at the seminar on the above topics are presented in the Annex to this report.

RECOMMENDED FUTURE WORK

5. The meeting recommended that it would be useful for countries to exchange experiences on developments and best practices in this field also in the future. However, participants recognized that two similar series of meetings were regularly convened under the work programme of the Conference – Meetings on Management of Statistical Information Technology (biannual - odd years) and Seminars on Integrated Statistical Information Systems (biannual – even years). In this regard, the meeting recommended merging the two series into one convened jointly by the ECE and Eurostat. It also recommended creating a steering group, which would define a framework for the new series comprising: (i) objectives of the series; (ii) target audience; (iii) partnership; and (iv) expected achievements.

6. The following participants volunteered to serve on the Steering Group: Mr. Mel Turner (Canada); Mr Jan Byfuglien (Norway); Mr. Marton Vuicsan (Netherlands); Mr. Richard Swartz (United States); Mr. Wolfgang Knüppel, Mr. Daniel Defays (Eurostat), Gerard Salou (OECD) and Mr. Karlis Zeila (Latvia), and agreed to progress the work using means of electronic communications by November 2002.

7. The meeting recommended that a Website on the Best “IT” Practices in Statistical Offices be created as proposed in document CES/SEM.47/CRP.3. OECD offered to host the Website, with the terms of reference specified in the paper. The authors were asked to prepare a revision of the proposal, which would reflect various interest groups emerging in the Programme Activity 2 of the Conference. A coordination group was created to manage the work on the website. Messrs. Markku Huttunen (Finland), Richard Swartz (United States), Peter Lubkert and Gerard Salou (OECD) agreed to serve on the coordination group.

ANNEX

Summary of the main conclusions reached at the April 2002 Seminar on Integrated Statistical Information Systems

Topic (i): Application of Web technology to integrate statistics

Documentation: Invited papers by Canada, United States and OECD. Supporting papers by Israel, Kazakhstan and Sweden.

Discussant: Mr. Jan Byfuglien, Norway

1. The emphasis for this topic was placed on the integration of statistical definitions, collection, processing and dissemination functions and how Web technologies can be applied to accomplish this. There was general agreement that Web-based solutions may provide a unified approach across various statistical programme areas. Originally, the Web was seen as a dissemination tool, but the potential area of application is rather wide from input to output.
2. The application of Web technologies requires a good coordination of the IT specialists and statisticians. The role of IT specialist in the area of creation and implementation of the Internet software, security systems/firewalls, database administration and backup and other traditional IT tasks was discussed. On the technological side, the XML and derived languages/technologies were considered highly relevant to statistics. Other Web-specific issues for further discussion were highlighted: browser inconsistencies and non-compliance with W3C standards; user authentication; encryption and data security. Data security is discussed in detail under Topic (ii) below.
3. As far as the role of statisticians is concerned, most of the presentations pointed out the importance of integrating the metadata, and encouraged further consideration of this issue. Participants agreed that the integration of metadata and harmonization and standardization of statistical concepts is a natural part of Web-based statistical activities, and preferably at the international level. It was agreed that the Web is a good motivation rather than a direct tool for harmonization and integration. An example of metadata harmonization within the OLIS.NET was presented at the Seminar, along with implementing a common classification of statistical activities, and developing common data catalogues and a glossary of statistical terms. Integration also requires considering issues of data, business rules, user identification and common database-related functions.
4. The budgetary aspects of the introduction of Web technologies were also discussed. The opinion was expressed that while new technologies allow reducing the number of printed forms, saving on mailing and data entry, these savings should be invested in the security and acquisition of a new type of expertise. It means that the achievement is not made in cost savings (costs are just shifted to other areas), but in improving the timeliness, reliability, consistency and comparability of data and in providing more timely services to end-users.
5. User centricity (coherence, common concepts), simplicity, currency/timeliness (up-to-date information and real-time interface) and utility (a clear set of task-oriented services) were spelled out as the basic principles for designing Web-based statistical services. Issues to be resolved are related more to the integration of existing statistical applications rather than to the technology itself, as the Web technology is quite simple.

6. As a possible architectural framework for Web-based integration, participants discussed the potential of a "Web portal" controlling all transactions when accessing an existing application, preferably in read-only mode. The Web portal represents an integrated interface to the existing applications; however, these may need partial redesigning in order to ensure coherency and to simplify the technical solutions. It was stressed, that the main functions of the existing components would need to be preserved, while new services and processing are provided.

7. An example of the application of the Web to the 2000 round of population and housing censuses was presented at the seminar, emphasizing the necessity for good coordination and harmonization between various parts of the statistical office. Web technology was considered as a possible alternative to Optical Character Recognition (OCR), but some participants stressed, that the OCR must still prove its efficiency. Another example of data collection assisted by Web technology related to Computer Assisted Personal Interviewing (CAPI).

8. The Discussant, in his concluding remarks emphasized the importance of further studies of data capture potential of the Web, which requires resolving several methodological issues as well as considering the quality. He also stressed that the Web helped to find the inconsistencies. The discussant also concluded that not too many examples exist yet on application of the Web in the data processing phase.

Topic (ii): Secure communications and data confidentiality

Documentation: Invited papers by France, Netherlands and United States. Supporting papers by Norway, Sweden, United States, UN International Computing Centre and European Commission.

Discussant: Mr. Richard Swartz, United States

9. The discussion under this agenda item dealt with the problems of confidentiality of statistical data and protection of respondents bearing in mind networking, the growing demand for more detailed data and microdata. The issues discussed under this topic included the use of the Public Keys Infrastructure (PKI); confidentiality aspects of direct user access to databases, software tools for data disclosure control and safe data interchange.

10. The necessity of a strategic approach to PKI implementation was emphasized. The importance of the organizational impact of PKI use was highlighted, in particular the need to align the PKI model with the way business is carried out. Another challenge discussed at the meeting was the interoperability between the PKI of the organization and PKIs of the business partners. A practical experience with using the Public Key Infrastructure (PKI) within a statistical agency was presented to the participants. It is planned that the PKI be used internally as well as externally. The issues concerned the attribution of certificates, accreditation of externally created certificates, and validity of PKI certificates across various public services. Another example related to implementation of PKI by the European Commission (IDA PKI). Some participants stressed that in their countries PKI can be quite expensive, and thus its use for statistical purposes was not successful.

11. A separation of the metadata identifying the reporting units and the statistical microdata (e.g. assigning random numbers while removing personal identity numbers and geographical information) was discussed as a possibility for disclosure prevention. While this method is currently used by several statistical agencies, it may increase the complexity of data processing. Some participants stressed that the data themselves can often lead to disclosure, even if the identification metadata were removed, so some data cells should also be suppressed and some linked cells may need to be replaced by intervals in the statistical publications

12. Methodology/mathematical methods for the protection of microdata stored in multidimensional statistical tables from being used for identification of the reporting unit were discussed at the Seminar. The risk of disclosure originates from the possibility of linking multidimensional arrays and longitudinal

series. Strategies for releasing statistical models in lieu of original data or tabulations have been proposed to address these concerns. Based on simple evidence gained by examining linear regression and spatial prediction models, it is clear that the advantages and limitations of doing so need to be carefully addressed.

13. The project Computational Aspects of Statistical Confidentiality (CASC) was presented to participants. This project aims at the development of practical tools (ARGUS software) for statistical disclosure control, and on research topics in the field of protection of microdata and tabular data.

14. During the discussion, some participants stressed that the decentralized management of data servers in statistical agencies may make disclosure protection more difficult. In such cases, the regular and coordinated updating of security patches is required. Another security concern was expressed at the Seminar, that often the widespread commercial Internet and communication software may represent a threat to security, and the open source application may better address the security issue.

15. Viruses, widespread in the PC environment, were also mentioned at the Seminar. A concern was expressed mainly in connection to data collection and interviewing, when the virus can copy the sensitive files or capture the keystrokes and transmit to an unauthorized address.

16. While some participants stressed the importance of internal security, others considered that external security is of a much higher importance. It was mentioned that temporary staff might represent a higher risk for internal security.

Topic (iii): Object-oriented technologies, component architecture

Documentation: Invited papers by Latvia, Netherlands and Eurostat. Supporting papers by Azerbaijan, Germany, Sweden and European Central Bank.

Discussant: Mr. Georges Pongas, Eurostat

17. The emphasis of this topic was placed on architectures that promote reuse and common use of processing components. Object-oriented approaches hold the promise of a more granular approach to reuse, building applications from small components and even sharing these components over the network or via object-oriented databases. The discussion covered issues of interoperability, new models of IT architecture for statistical services and impact on the organization.

18. A project of major reengineering of the statistical information systems of a national statistical agency was presented. This programme, which comprises complete replacement of the software, is based on an ideal schema outlined at the February 1999 Meeting on Management of Statistical information Technology (Bo Sundgren, Statistics Sweden: "An information systems architecture for national and international statistical organisations"). Another presentation focused on combining functionally rich components Argus, Slice, Bascula, Blaise Cristal). Eurostat presented its approach for software interoperability applied in the context of the STATOBJECT IST project, where components base their interactions on the semantics of the business they support. Other presentations covered the network based on the Linux operating system, statistical information system GENESIS, UML Model of the message for interchange of time series GESMES/CB and development of components.

19. It was highlighted during the discussion that using object-oriented technology requires a conceptual approach. While some participants stressed that ordinary developers may experience difficulties in practicing well these techniques, others mentioned that object-oriented technology is a potentially natural choice for statistical applications. Participants also discussed the constraints on the component software and the limits of the applications made up of loosely linked components.

20. The management of complex projects was discussed at the Seminar. This often involves independent teams working on individual components. In this connection the issues of outsourcing and the relationship between the IT architecture and the organizational structure were discussed. The need for coordination was highlighted in the discussion, when the importing and creating of new components have

to be cleared before being allowed to be included in the system. Various sources of components (e.g. imported and self-made components, successful applications, etc.) were considered.

21. The discussion revealed that it is difficult to predict which platform will prevail in future, as past experience has brought some surprises. Therefore, we can expect a wide variety of platforms from Microsoft to Linux also in the future. The choice of platform has an impact on other software for XML support, UML modelling, database querying, etc.

22. In concluding the discussion, the participants agreed that: (i) the component architecture is a very promising technique, but has to be used with care; and (ii) the development of new products has a tendency towards standardization (e.g. XML, standards data exchange formats, etc.)

Topic (iv): Ways of making statistical information systems more responsive to users

Documentation: Invited papers by Canada, Greece and OECD. Supporting papers by Netherlands, United States and SIAP. Proposal for a website on best practices by Armenia, Canada, Netherlands and OECD.

Discussant: Mr. Marton Vuksan, Netherlands

23. The participants welcomed the proposal by Canada, Netherlands and OECD to create a website on best IT practices in statistical offices. They agreed that the website will be hosted by the OECD and conform to the standards for OECD websites, while the ownership and responsibility for content would reside with the contributing countries. It was also agreed that each participating national statistical office would nominate a focal point to ensure communication with the OECD. The heads of delegations present at the Seminar will initially assume the function of focal point. A coordination group was created to manage the work on the website. Messrs. Markku Huttunen (Finland), Richard Swartz (United States), Peter Lubkert and Gerard Salou (OECD) agreed to serve on the coordination group.

24. Modelling issues and realizations were discussed in connection to metadata. It was also pointed out that metadata represent a prerequisite for a good statistical website. Warehousing, as mentioned in the discussion, gives power to the statisticians. Some strategic issues (e.g. using warehousing as a main data repository) as well as some technical issues were considered in this connection. There was general agreement that the Web has proven its great dissemination capability, on the condition that coherent metadata and warehouse architecture is applied.

25. At least three types of metadata were identified: (i) semantic – informal descriptions, free text; (ii) operational – process information mostly, data quality, number of imputes, estimations made, etc.; (iii) technical – field names, data types, sizes, etc. (for computer processing, not for human eyes).

26. A project aiming at the establishment of an integrated metadatabase in a national statistical office was discussed at the Seminar. It was noted that the collection and update of metadata are enormous tasks requiring major effort and commitment by the organization. In some cases multilingualism may be required, for example the demonstrated project has to cover equally two national languages. Another presentation related to a semantically rich metadata model aimed at supporting metadata-guided processing.

27. The innovative use of information and communication technology allows the provision of a common access to the organisation's databases. In this connection, OECD presented an improvement of access to its data using a common portal. Some basic conditions were mentioned for easy user access: access based on the Internet (no need to install any software on the users' computers); availability of a search tool, easy user interaction and data manipulation, availability of or easy linking to analytical tools, dynamic access and ODBC/OLEDB connectivity.

- - - - -