



**Economic and Social
Council**

Distr.
GENERAL

CES/AC.71/1999/31
3 March 1999

Original: ENGLISH

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Meeting on the Management of Statistical Information Technology
(Geneva, Switzerland, 15-17 February 1999)

REPORT OF THE FEBRUARY 1999 MEETING

1. The Meeting on the Management of Statistical Information Technology was held in Geneva, Switzerland, from 15-17 February 1999. It was attended by participants from: Armenia, Austria, Belgium, Canada, Croatia, Cyprus, Czech Republic, Estonia, Finland, France, Germany, Hungary, Ireland, Israel, Latvia, Lithuania, the Netherlands, Norway, Poland, Portugal, Romania, Russian Federation, Slovakia, Slovenia, Spain, Sweden, Switzerland, The former Yugoslav Republic of Macedonia, Turkey, United Kingdom, and the United States. Representatives of the Organization for Economic Cooperation and Development (OECD), the United Nations Food and Agriculture Organization (FAO), the United Nations Population Fund (UNFPA) and the United Nations Statistics Division were also present.

2. The provisional agenda was adopted.

3. Mrs. Barbara Slater (Canada) was elected Chair and Mr. Karlis Zeila (Latvia) was elected Vice-Chair.

4. The following substantive topics were discussed at the meeting:

(i) The impact of Internet on the statistical production and dissemination process;

- (ii) Economic issues associated with the implementation of modern IT and management of human resources as driving forces to improve the timeliness and quality of statistical data;
- (iii) Integration of statistical activities at the national and international levels, including data modelling strategies and standards needed for statistical data integration;
- (iv) The technologies and methods in the implementation of the 2000 round of Censuses;
- (v) Year 2000 problem

5. The following participants acted as Discussants: Mr. Christophe ALVISET (France) for topic (i); Mr. Jan BYFUGLIEN (Norway) for topic (ii); Mr. Heinrich BRÜNGGER (Switzerland) for topic (iii); Mr. Aidan PUNCH (Ireland) for topic (iv); and Mrs. Barbara SLATER (Canada) for topic (v).

6. The Meeting recommended that a further Meeting on the Management of Information Technology be convened in 2000/2001. It recommended, therefore, that the following text be included in the 2000/2001 Integrated Presentation of the Programme of Work of the Conference of European Statisticians:

2.1. Management of information technology infrastructure

Activities of the ECE

The Meeting on the Management of Statistical Information Technology in 2000/2001 to consider the following (those countries who expressed their interest in the issue in question are noted in parentheses):

- (i) The impact of data warehousing on the management of statistical offices (the Netherlands, Switzerland);
- (ii) Challenges and opportunities for statistical offices working in a network environment (Sweden, France);
- (iii) Resource management in statistical offices and the role of the IT departments (United Kingdom, United States, Canada);
- (iv) Integration of statistical (survey) data with registers (administrative) data (Sweden, Croatia).

7. The participants expressed high appreciation for the invited paper "An information systems architecture for national and international statistical organizations" which Mr. Bo Sundgren of Statistics Sweden had prepared for this session. They agreed that the paper provided an excellent basis for the consideration of the topic on "Integration of statistical activities at the national and international levels, including data modelling strategies and standards needed for statistical data integration". It was pointed out that such

material could significantly assist senior management of NSOs in their decision-making. The secretariat was asked to ascertain whether the author, in cooperation with a group of interested experts (Canada, the Netherlands, Switzerland and United States), could revise this material with the aim of preparing draft guidelines for statistical offices which could be submitted to the Conference of European Statisticians for adoption and publication in the Conference's Statistical Standards and Studies series. It was also considered important that the finalized material be made available in all three ECE working languages.

8. The main conclusions the participants reached in their discussions are presented in the Annex.

ANNEX

**Main conclusions reached at the Meeting
on the substantive items of the agenda**

A. The impact of Internet on the statistical production and dissemination process

1. The meeting discussed different methods of Internet use and its development in statistical offices. The development in a particular country depends not only on the statistical office but also on the general IT environment in the country. The Internet should be used along with numerous other technological tools. The advanced Internet tools for data collection and dissemination can be used only when the communication networks in both the country and the statistical organization concerned have reached the relevant technical and security levels.
2. The impact of Internet on management of the statistical office is much broader than the simple facility of data dissemination. In many cases, the Internet has proven to be a driving force for modernising the organization of work in the whole statistical office. For example, Internet could push statistical offices to redesign their organizational models from isolated subject-matter statistical units towards integrated production.
3. The impact of new technology to re-engineer these processes is most strongly felt in statistical data dissemination and collection; it has not yet brought about significant changes in the way statistical offices operate. It was also pointed out that re-engineering involves substantial changes in the processes of the organization. Thus, not every change can be considered as re-engineering (e.g. transition from statistical production of one computer platform to another).
4. In **data collection**, the Internet provides an opportunity to enhance public cooperation by providing an additional response mode; Internet offers possibilities to reduce response burden and to improve data quality and processing efficiency. The importance of administrative registers and other sources for data collection is increasing. Combining this secondary information with statistical data is a methodological challenge. Stringent data quality procedures must be implemented to verify the accuracy of information before it is entered into the database.
5. With regard to statistical work at the international level, the requirements relating to comparable results become indispensable. The work on harmonized definitions and methods of data collection should be intensified and extended to other statistical areas. The role of international organizations in this regard was emphasized.
6. The contributions also demonstrated the increasing use of Internet as a

tool for **data dissemination**. Furthermore, Internet is used to enhance the image of the statistical office and to advertise statistical data that is available. It is also used for commercial marketing and supplying data against payment. In many agencies, Internet is considered as the principal dissemination channel of the future. Some countries reported, however, that Internet would not seriously compete with conventional publications in the near future.

7. Several participants underlined the importance of finding solutions concerning **pricing** of the data disseminated over the Internet. Making data available for free often increases interest in statistical services, and thus generates additional revenue. However, while the marginal costs of information disseminated through the Internet could be very low, there are significant costs in operating an Internet site and in developing and updating its content. These costs need to be provided for by the government and/or recovered by the statistical office from the user.

8. The lower costs in Internet dissemination can only be achieved by automating as many production steps as possible. **Database publishing** methods have been developed through which HTML pages are created and updated automatically. The main concept of database publishing is to separate the maintenance of the underlying information from the representation of its contents as HTML pages.

9. Database publishing requires expert resources for the development of the necessary databases, systems and procedures. For a less frequent publishing program it is simpler and cheaper to create HTML pages manually. The trade-off between the manual process and the automated database publishing system needs to be evaluated for each case.

10. A **data warehouse** of published and publishable statistics will play an increasingly important role in data dissemination over the Internet. The data warehouse must accommodate both the data and the related metadata. Such a data warehouse could then serve as the primary source for publishing on the Internet. Often, the main prerequisite for an efficient data warehouse is the availability of an electronic database with interfaces to the Internet and Intranet.

11. Many participants emphasised the increasing importance of data security and confidentiality when using the Internet. Different approaches were reported to solve the security problem. Some countries mentioned, for example, an approach based on the maintenance of two separate networks: one for internal and one for external users. It was pointed out, however, that this approach could be quite expensive. There was all-round agreement that there is no standard solution to this problem, as the conditions influencing this issue vary from country to country and different institutions are very often responsible for individual issues (e.g. legal aspects, questions of privacy, and interests of individual external users).

12. Internet/Intranet may efficiently serve both external and internal users. HTML creates a common platform and standard for communication; no special effort is required for the transformation of data and its structure for either internal or external use.

13. Internet has an **impact on the quality of statistical data**, in particular with respect to coherence of concepts and consistency of data. Internet highlights the importance of statistical metainformation. It also offers some new, more efficient opportunities for easier tools to analyse and manage metadata, including in natural language.

14. A positive feature of the Internet is the close relationship between data suppliers and users, and the possibility to improve statistical service via user feedback. Statistics on access to different parts of the Internet service enable the data suppliers to analyse individual demand for the different subject-matter areas, and to adjust future plans. Therefore, continuing market surveys and user feedback analysis should be included in the annual programmes of all statistical agencies.

15. Development towards the Internet requires skilled IT personnel. Therefore, proper training and movement towards 'standards' and 'best practices' for Web development work is necessary.

16. **Further trends in Internet application** depend on a wide range of factors, such as the development of more user-friendly and self-explanatory software, the capacity of transmission networks, the development of access speed, and prices, etc. Keeping up with the latest Internet developments requires an immense effort on the part of the statistical office. Also, the balance should be kept between a forward-looking design and realistic assumptions of the skills and infrastructure of the clients' premises.

B. Economic issues associated with the implementation of modern IT and management of human resources as driving forces to improve the timeliness and quality of statistical data

17. It was noted that significant new features in IT development are a greater computer capacity, lower prices (but also a shorter lifecycle) for hardware and software, internal and external networking, distributed processing and the decentralized use of data, and more flexible and user-friendly tools.

18. Rapid IT development calls for regular reconsideration of development strategies and could have a significant influence on the related economic considerations. There is less potential for large investments which are presumed to be stable for many years; it is more important to develop "model" solutions and a common strategy for the organization. More flexibility in planning is needed. It is most important to maintain documentation on and evidence of historical developments. Plans for IT development and its implementation are in many cases not realistic. The crucial question is how

to improve planning and its follow up.

19. The main aim is to implement IT in a **cost-efficient** way. Several examples of how to analyse cost-efficiency were provided. The discussion revealed that the estimation of cost-efficiency of IT in a statistical office is a complex task.

20. One reason for this is that to specify the benefits by producing a "public good" such as statistics, where there is no tangible market and competition, is very difficult. It is particularly complicated to examine cost-effectiveness over time when the cost accounting systems have changed. Furthermore, new IT often involves intangible benefits and costs that are difficult to measure (e.g. integration of data from different sources, and enhancement of the statistical office's image).

21. Cost-effectiveness and cost-benefit analyses are merely tools to guide IT investments. Ultimately, the goals and strategic objectives of a statistical office must determine what investments are made.

22. The meeting considered it useful to prepare a draft set of international guidelines for evaluating IT investment decisions in statistical offices and requested the secretariat to organize this work. The U.S. Bureau of the Census offered to prepare a first version of this material. A group of experts from Canada, France, Latvia and Norway will cooperate in the review and refinement of this material. All other countries are encouraged to provide input to this document.

23. The question of **human resources** is of major concern to IT management. To recruit and maintain high level IT staff is an indispensable prerequisite for the effective and efficient use of IT. Governmental organizations often cannot compete with private businesses or research institutions, particularly with respect to salaries. Hence, other factors such as interesting and challenging tasks, a good working environment, possibilities for training and personal development, and more flexible working conditions must be emphasised. The existence of an organization-wide policy for maintaining and developing staff skills is very important in this regard.

24. New technology very often implies new organization of the work processes, which should be performed with close cooperation between all parties involved in the organization. Poorly defined responsibilities and links to top management may negatively influence the whole implementation process. Planning and implementing new IT requires clear coordination, which can be a challenge, especially in a decentralized organization.

25. Implementation of new IT is accompanied by increased training needs. The meeting highlighted the usefulness of sharing experience in new IT implementation. The importance of international cooperation in this respect was emphasized.

26. The meeting discussed the relationship between modern IT and statistical **data quality**. There was common agreement that modern IT permits improvement in the accessibility, usability and cost-efficiency of statistics. It also provides tools to improve the **timeliness** and documentation of the data and metadata. At the same time, it raises problems in ensuring coherence and comparability across time and different subject-matter areas.

27. Decisions on technological change must be endorsed by top management, followed by the allocation of the necessary resources for implementation. IT strategy should support the overall statistical office's development strategy.

28. The meeting looked at some ways to facilitate the implementation of modern IT in a statistical office. The discussion reconfirmed that the collection and handling of statistical data differs from that of other administrative and commercial data. Tabulation and statistical analysis as well as documentation sometimes require clarification of a type that is not provided in general database and software solutions, meaning that in-house, or perhaps joint, development may be necessary. General commercial solutions, however, should be sought rather than specific in-house development whenever possible.

C. Integration of statistical activities at the national and international levels, including data modelling strategies and standards needed for statistical data integration

29. With the orientation of statistical offices moving towards output and customers demands, it becomes increasingly important to integrate individual output into a consistent system. Decentralization and freedom in selecting approaches has been the general trend in statistical offices. To maintain and further develop an integrated and consistent system is now a vital task in this context.

30. An important component of the information systems architecture is a **corporate data warehouse**. The discussion focused on the objectives and difficulties in the implementation of a statistical data warehouse. It was pointed out that changes in corporate culture and in work processes to emphasize sharing of information are very important for accomplishing this, and that the driving force in establishing the warehouse is the customers' needs.

31. More flexibility in combining statistical data across different subject-matter areas is a substantial characteristic of a statistical data warehouse. Such a data depository could be the main tool for implementing the clearing-house function, i.e. matching the needs of the customers with the data provided by the statistical office. The data warehouse could function internally as a repository for all corporate data. Externally the data warehouse, or a selected subset of it, could function as an output database for external users.

32. A data warehouse is a strong tool for internal coordination. The critical success factor for its implementation is the consistency of the statistical data across subject-matter areas. Often the need to harmonise the concepts requires a strong driving force and a coordinator or coordinating body at a senior level within the office.

33. The strategies and difficulties in the implementation of data warehouses were discussed in more detail, e.g. how to migrate to the data warehouse approach, how to choose which parts/functions of the data warehouse could be implemented first. Another important problem is how to best structure microdata (final observation registers), and whether there are any standards for the multidimensional model for macrodata. A difficult question when maintaining a data warehouse is how to ensure coherence among the data. It is also important to compare the benefits experienced by countries which have set up a corporate data warehouse with respect to the investment and running costs.

34. A data warehouse approach requires generalized software supporting the standardization of data and metadata. In addition, there could be generalized software tools supporting all important processes and sub-processes in survey processing systems and analytical processing systems.

35. Flexibility is an important consideration for all the technological components (hardware, software and data components). A modification of the contents, structure or storage of data should not necessitate modifications of technological tools.

36. It is recommended to use modelling tools as a starting point for the design and re-engineering of a statistical system. Several participants considered it important to use an iterative implementation approach. The technological changes should be modelled and designed in close co-operation with the subject-matter statisticians in order to meet their requirements. Prototyping and direct user involvement help to ensure that a maximum of user requirements are satisfied.

37. The overall strategy in the area of IT is to minimise the number of platforms and operating systems, in order to minimise costs and to maximise compatibility among products and services. In addition, working with the standards makes it easier to rapidly benefit from progress in technology. In general, it is better to standardize in terms of interfaces between components rather than in components themselves.

38. The role of top management in decision-making with regard to the content and architecture of data warehouses was emphasised. Some countries reported on their experiences with respect to the coordination needed for the efficient functioning of the statistical data warehouse (e.g. a Steering Committee responsible for the running and maintenance of the statistical data warehouse). Customer feedback was highlighted as an important factor for reasonable data warehouse functioning and its further development.

39. OECD reported recent steps taken in integrating data, metadata and calculation methods together in one database as much as possible. This project is based on the approaches developed in the invited paper.

D. The technologies and methods in the implementation of the 2000 round of Censuses

40. The meeting considered implications of modern IT on the technologies and methods for the Census 2000. Most of the processes benefit enormously from the application of modern technology. In terms of scale and cost, the Population and Housing Census represents the most important operation which most statistical offices undertake. It also provides basic data for many other statistical projects.

41. Censuses also often provide statistical offices with the opportunity to obtain considerable funding for improving their IT systems. This is optimized when the relevant software and/or hardware has been purchased and when the applications can be used elsewhere in the organization.

42. The decision as to whether to adopt an evolutionary or revolutionary approach for the introduction of modern technology in census operations depends on a number of factors, the principal one being the frequency of census-taking in the country. The longer interval may make approaches taken in previous censuses outdated. In many cases, the census may turn out to be a trial base for new technology. The risks with the implementation of new IT are higher in a census than in most other statistical surveys. Very up-to-date solutions can add an additional layer of complexity and risk into the census organization.

43. Several offices use outsourcing for individual census procedures. This need may arise because the organization does not itself have the necessary expertise, or because the resources are required for short-term, non-repeated activities. With outsourcing comes an added layer of risk which has to be carefully managed. The ultimate responsibility and accountability for census operations resides with the statistical office.

44. Another important task is the selection, appointment and training of census enumerators. The training should ensure both the subject-matter and technological skills. Establishing the required number of competent enumerators and administration of the census personnel requires special attention.

45. In data **collection**, a large-scale move towards scanning and automatic recognition techniques is taking place. The questionnaire should therefore be designed with this in mind. The use of tick-boxes can ease the burden on respondents and at the same time simplify processing.

46. Some participants reported the provision of automated telephone assistance to aid the public in completing the questionnaires. The Internet

is also being used for census responses in a small number of countries. In those cases, the need to avoid double-counting is very important. The U.S. Bureau of the Census reported that it uses a computerized matching software system to identify and eliminate multiple responses for the same household. The software can also be used to implement other census follow-up coverage measurement operations.

47. In those statistical offices where national address databases exist, the latter have a significant influence on the census organization, especially where they are both up-to-date and comprehensive. The mail-back option is particularly suited where confidentiality concerns are a major issue (e.g. in the case of censuses which have potentially sensitive questions such as income).

48. A discussion took place on whether the census organization differs considerably in those countries using register-based statistics. The large availability of population registers, where they exist, can be a good alternative to traditional censuses. In order to implement a register-based census successfully, harmonization of different registers may be required.

49. Automatic scanning and optical reading for data capture means that the usual bottleneck which occurs when entering data into a computer can be avoided. In many cases the scanning and optical reading will be performed by external companies, as this is a one-time exercise requiring extensive computer capacities and know-how.

50. Considerable progress has been achieved in the area of automatic coding. The possibility of recognizing written responses coupled with automatic coding techniques enables 100% coding and processing of variables to take place.

51. It was pointed out that in designing data editing rules, there is a danger of over-editing. Data editing is a costly and time-consuming process. A balance therefore has to be struck in order to produce a sufficiently clean output within reasonable time and cost. A special problem is editing register-based data, and linking register data with statistical indicators.

52. On the operational side, the individual systems should work in harmony. Therefore, good operational procedures are needed for utilizing the powerful new technology.

53. The technology gains will be most apparent in the **data dissemination** area. The major challenge for statistical organizations is the timely provision of accurate results. Users ask that the results of censuses be made available more quickly and in a more flexible fashion than before. To respond to these requests, greater use will have to be made of technological means of dissemination than in the past. The balance will have to switch from the printed medium to electronic and other forms of publication. In several countries, the use of the Internet will be the main innovation in disseminating the results of the 2000 round of censuses.

54. It was pointed out that it would be useful to transfer the experiences and know-how gathered in organizing population censuses to censuses in other statistical areas (e.g. agricultural census). The need to make a post-census evaluation, i.e. to consider how planned actions have turned out, was also mentioned.

E. Year 2000 problem

55. The discussion focused on the impact of the Year 2000 (Y2K) on statistical agencies, and on the managerial responsibility for finding solutions to the related problems. Within statistical agencies the Y2K problem can affect both the statistical and non-statistical information systems, such as administrative, personnel, procurement and finance systems. Special attention should be paid to this problem in regional statistical offices.

56. Furthermore, the Y2K problem should be addressed in a wider statistical environment involving data suppliers and data users at both the national and international levels. It is important to discuss their preparation with other involved government departments, and with data providers regarding their continued provision of accurate and reliable data.

57. Some countries reported that they have been dealing with the Y2K problem for some time, e.g. concerning population registers. Similar kinds of issues have been addressed in several countries in connection with the introduction of the euro currency.

58. Finding solutions to these problems in a statistical agency is not limited to the technology. The Year 2000 issue is a business continuity problem rather than an IT problem. It needs to be addressed organization-wide at the level of top management, in order to plan and monitor the corrective actions, the human and financial resources, the availability and cost of external expertise, and the extent to which the organization is dependent on external computer systems (telecommunications, financial institutions, electricity supply, etc.). Checking is required even when the technology has been recently purchased.

59. It is recommended that statistical offices make an inventory of the custom computer programs that they use, and to develop a plan for the remediation or replacement of each application that may be non-compliant with the Year 2000 requirements. The mission critical programs that provide key current socio-economic indicators, and the systems related to internal pay and cash handling, require special attention. The discussion revealed that analysis and preparation for solving the Y2K problem could positively influence the management and project planning of statistical offices in general.