STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Joint ECE/EUROSTAT Meeting on the Management of Statistical Information Technology
(Geneva, Switzerland, 14-16 February 2001)

REPORT OF THE FEBRUARY 2001 MEETING

1.   The Meeting on the Management of Statistical Information Technology was held in Geneva, Switzerland, from 14-16 February 2001. It was attended by participants from: Albania, Armenia, Bosnia and Herzegovina, Canada, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Israel, Italy, Latvia, Lithuania, the Netherlands, Norway, Poland, Portugal, Republic of Moldova, Romania, Slovakia, Slovenia, Sweden, The former Yugoslav Republic of Macedonia, United Kingdom and the United States. The European Commission was represented by Eurostat. Japan also participated in the meeting under the terms of Article XI of the Terms of Reference of the ECE. Representatives of the Food and Agriculture Organization (FAO), the International Monetary Fund (IMF), the Organization for Economic Cooperation and Development (OECD), the United Nations Educational, Scientific and Cultural Organization (UNESCO), the United Nations Industrial Development Organization (UNIDO), the World Trade Organization (WTO) and the Bank for International Settlements (BIS) were also present.

2.   The provisional agenda was adopted.
GE.01-

3.   Mr. Mel TURNER (Canada) was elected Chair, Messrs. Karlis ZEILA (Latvia) and Enrico GIOVANNINI (OECD) were elected Vice-Chairs.

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

(a)   Challenges and opportunities for statistical offices working in a network environment;

(b)   The impact of data warehousing on the management of statistical offices;

(c)   Integration of statistical (survey) data with registers (administrative) data ;

(d)   Resource management in statistical offices and the role of the IT departments.

5.   The following participants acted as Discussants: Mr. Wolfgang KNUEPPEL (Eurostat) for topic (a); Mr. Lars RAUCH (Sweden) for topic (b); Mr. Christophe ALVISET (France) for topic (c); and Mr. Richard SWARTZ (U.S. Census Bureau) for topic (d).

6.   The participants recommended to develop a Website containing best practices on applications of statistical information and communication technologies, and transfer of know-how in this area. The secretariat was requested to ascertain the possibilities of solving this task..

7.   The Meeting recommended that a further Meeting on the Management of Information Technology be convened in February 2003. It recommended, therefore, that the following text be included in the 2002/2003 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 2002/2003 to consider the following:

(i)   Measures for the improvement of quality on IT management level;

(ii)   The impact of technical measures and standards on data quality;

(iii)   Efficient management of increasing technical complexity.

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.   Challenges and opportunities for statistical offices working in a network environment

1.   As the development in statistical production is to a great extent driven by IT, the statistical offices are facing very similar challenges in this area. The session considered the current state of networking in statistical offices and identified areas where some (positive or negative) experiences can be shared and areas where the problems hindering further development.

2.   The work in a network environment has an impact on all three phases of statistical production: data collection, processing and dissemination. In addition, there are some horizontal issues, such as management, training and support, that are cross-cutting. IT management needs to assess the risks and opportunities that derive from technological evolution and how these affect the information production process. The desire to speed up innovation can lead to proceeding prematurely, while overlooking critical aspects such as the careful design of the systems, continuity of existing practices, etc..

3.   The network environment creates many opportunities to improve the work of the statistical offices. At the same time, new problems arise which the managers must solve, such as remote management and monitoring of operating systems, data protection and the right to privacy, data integration, the need for real time support, the need for training in the latest technologies, the constant rise in costs for the new software and hardware, and the changing role of IT managers.

4.   Electronic collection of raw data (e-CoRD) using Internet technologies is a new development that has the potential to improve timeliness and reduce administrative costs. To keep pace with the emerging new tools in e-business, new approaches are necessary for electronic data collection. However, in practice the use of such solutions is relatively low at present. The meeting discussed the factors that are hindering exploitation of e-data collection. Several countries pointed out the need for international standards and guidance for the development directions of electronic data collection. Solutions are also needed for practical technical questions like questionnaires that need to be filled in by multiple respondents or during multiple sessions.

5.   There is often pressure from business respondents for electronic means of data provision but in some countries the legislation is not yet at a level permitting full implementation. Also, the business enterprises have not realised the importance of security in this process and are ready to sacrifice security for cheaper or simpler solutions. The tighter security mechanisms are often considered too cumbersome, and their wider use can not be expected until the problems with technology and user-interfaces are solved in a user-friendly manner.

6.   Electronic collection requires new skills and instruments which will increase costs in the development phase but which can decrease expenses in the long run. The classical EDIFACT-based solutions work properly but are too expensive to implement for medium- and small-sized enterprises. Several offices are using electronic questionnaires for data collection. The expected reduction in Internet costs will make it possible to move from completing and e-mailing pre-defined questionnaires to an interactive mode. Hungary described a newly launched pilot project for electronic data collection via Internet using a secure environment. The Web-based applications for data collection should be accompanied by online help functions and the network service must be reliable to allow the required data transfer.

7.   Several offices shared their experience in guaranteeing network security. In many cases, the available technology (e.g. SSL, Public Key Infrastructure, VPN) seems relatively immature or difficult to implement. It is not possible to construct a system with unbreakable security in all areas; some degree of risk must be accepted in practice. It can be helpful to outsource security services to companies specialising in implementing security systems.

8.   The World Wide Web has a great potential for improving dissemination of statistical products. At the same time, it increases users' expectations with regard to timeliness, availability, accessibility, intelligibility and usability. The number and variety of users has significantly expanded. Among these are highly qualified expert users, as well as the general public, school students, etc.. Users' understanding of economics, statistics and computers, and their needs for data, vary significantly. It can be observed that the importance of the Internet as a dissemination vehicle is constantly increasing. It is likely that in the future the quality of services of the statistical offices will often be judged by the quality of their Web pages. Designing Web pages is a similar kind of craft to that of designing a survey, and appropriate attention has to be paid to the skills and staff required for this.

9.   Methodologies, tools and techniques exist for evaluating and improving the design of

dissemination web-sites and web pages. Lately, a lot of attention is being paid to usability, i.e. the degree to which the Website helps the user to accomplish his/her specific task. Usability testing implies measuring the characteristics of human-computer interaction while using a Web page. Usability engineering simplifies back-end activities (maintenance of site, fewer inquiries and complaints, reduced load on servers and networks) but adds complexity and cost to the development. Statistical offices can implement usability testing gradually. A simple kind of testing can be implemented without any special technology and can take only a few hours of staff's working time. At the other extreme, it can be built up to furnish full usability testing laboratories.

10.   Where services are provided over the Internet for data collection or dissemination, the users expect support to be provided continuously and to obtain a rapid response. Together with a dissemination policy, there has to be a policy for user support in the statistical office. The support needs to be both technical and functional. There are different levels of support (e.g. technical, conceptual), and it has to be made clear what level of service can be expected for each of these areas. Some offices experiment with automatic replies to simple queries. An important point is that user feedback can give valuable information for improving the whole production process. Furthermore, the increasing Web-based dissemination raises the question of necessity for statistical offices to protect their e-publishing and use techniques to guarantee the copyright of their publications.

11.   Although the use of the Internet will increase significantly and may ultimately become dominant, the take-up rate will be slow. Therefore, existing channels must continue to be supplied in parallel with these new opportunities and overall costs are unlikely to be significantly reduced in the short term.

12.   The Internet can make data available from anywhere – the problem for the user is to find it. More integration of data and metadata between international organisations has been suggested, with links inserted from the websites of international organisations to the more detailed metadata on national networks. A possible single portal for this kind of access could be based on the IMF Dissemination Standards Bulletin Board (DSBB). The meeting was informed about a similar project started at the November 2000 Work Session on Statistical Metadata. The key problems to be solved in order to guarantee the success of this approach are language, harmonisation of terms, variation of structure and resources.

13.   A network environment can provide statistical offices with an opportunity to share data and metadata and to eliminate data redundancy. It requires the harmonisation of concepts, classifications,

etc.. International organisations should play an important role in setting up standards in this area. Practically, harmonisation of concepts can be very difficult and sometimes even impossible. It is often more efficient to have tools for negotiating mapping of concepts in particular areas instead of trying to agree on common standard concepts. It can be recommended to follow the standards to the highest degree possible and then to document the differences. This raises the question of having sufficiently detailed data definitions and efficient techniques to collect and store this metadata. Solutions from the area of artificial intelligence might be needed to capture the meaning of concepts.

14.   The importance of networking can be observed not only at the technical but also at the organisational level, within and between countries. New developments in IT, and the need to use these efficiently, encourage countries to increase cooperation regarding statistical information technologies. There are common drivers for development in the area of IT: common user requirements, concerns of data suppliers (confidentiality), administrative concerns (outsourcing, use of administrative resources), cooperation among NSIs (sharing of experience, software, manuals), European integration. This has invoked Eurostat to take a role in drawing up and implementing a community policy on IT for statistics. Common approaches can be seen in finding solutions for data dissemination, concerns about the protection of privacy, joint approaches in combining the information systems in statistical offices, making greater use of administrative sources and encouraging the use of standards in data exchange. The research should play an important role in solving these issues.

15.   The new Internet technologies also have the potential to facilitate groupworking through the creation of Extranets on the public network. Examples of a range of different projects that exploit this were given. Extranets can provide facilities for group-oriented work, contribute to the statistics production process, help to coordinate the statistical system and can provide information to the main statistics systems' users. The potential for future developments in this area was discussed and its possible applications in the European Statistical System. Often there are no ready-made solutions for implementing Extranets and they have to be tailor-made. One of the promising new developments is to use the XML format for real time messages delivery with administrative partners.

16.   Extranet is especially useful in countries with the distributed statistical system allowing people in different ministries' statistics departments to work together. Some countries have already functioning or are building up state administration communication infrastructure using an Extranet. This will make it possible to mutually use administrative sources within public administration and the basic nation-wide registers such as population register, business register, etc..

17.   XML was often mentioned in several contexts as a technology with significant promise to statistical agencies. It is favoured as a means to describe data structures and as an interchange format. It has the potential to unify multiple existing standards for interfacing disparate systems and data sources.

B.   The impact of data warehousing on the management of statistical offices

18.   The development of output databases and data warehouses may change the paradigm of the statistics production process that is traditionally subject-area oriented. The consequences for management caused by the move from a stovepipe organised production to an integrated approach for the statistical production system were considered.

19.   Although the implementation of data warehouses is in its initial stages and there is not yet much experience gathered, some statistical offices provided examples of the kind of organisational changes which have been necessary. In addition to re-engineering individual production processes, a cultural change in statisticians' approach towards the production of statistics is required. A strategy is needed for spreading the integration culture in the office and for solving the associated personnel problems. It can be helpful to rotate staff involved in integration or to create a central pool of staff who can be distributed to major statistical units. At the same time, subject-matter expertise should be maintained. Alongside integrated production systems, there will remain a need for teams focused on subject-matter.

20.   Contrary to the stovepipe organised statistics process, the development and maintenance of data warehouses and centralised databases is not the responsibility of a particular subject-matter department. Harmonisation and unification of processes that are distributed in the organisation will not work without a strong and efficient management. Administration is an important success factor and an elaborate data warehouse administration structure with a hierarchy up to agency level is needed.

21.   Several offices have chosen a gradual approach to the implementation of data warehouses and integrated statistical information systems. They build on the existing systems and provide interfaces to the new ones because of the necessity to keep the production process running. It can be recommended to adopt solutions which have been tried and tested in other environments. It would be good to be able to measure the degree of integration of different production streams. Such a measure could help to monitor the process and to obtain attention and support from the management and staff.

22.   Cumulating events in the data warehouse is time-critical. The processing structure should be oriented towards a system of continuous processing. It means moving from a system with preliminary and final figures to a system with a continuous margin of error. The margin of error for the result figures decreases over time when more precise data becomes available. Quality will be achieved in a continuous fashion and not at discrete intervals. It would allow users to make their own decisions between different aspects of data quality such as timeliness and precision.

23.   One of the main problems in data warehouse implementation is data consistency. Sophisticated metadata systems and harmonised classifications are needed to support the data consistency. Although statistical metadata are increasingly used in the production process, there is a lack of unification and office-wide harmonisation. The use of relevant metadata should cover the whole production process, starting with data collection. Setting up a centrally managed metadata system is the responsibility of management, taking into account the different interests in a large organisation such as a NSO. Documenting discrepancies from the existing standards is often not enough; tools are required to estimate the bias introduced by inconsistent definitions. Also, it is sometimes necessary to maintain different sets of data over time in order to be able to analyse phenomena from the viewpoint of the situation at a given time period. A good solution would be to create automatic tools based on metadata description in order to make data consistent. However, the development of such tools is not yet feasible at the current level of metadata management systems.

24.   The high volume of confidential data available in data warehouses and integrated statistical information systems makes it necessary to review the production environment from a confidentiality perspective. The dramatically extended possibilities to combine data from different sources, drill down functionalities, cross-tabulation features, etc. raise the need for tools able to protect confidentiality in these conditions. In particular, public use of anonymous microdata for analytical work should be avoided because of the very high risk of disclosure. The database management systems do not include tools for confidentiality protection which is a great disadvantage in their use in statistics. Also, the current techniques to treat disclosure are not powerful enough to handle all the confidentiality problems in large data warehouses.

25.   The responsibility for solving these problems can not be left only to subject-matter experts and IT staff. It is a management task to create the necessary awareness of confidentiality problems and to manage the preconditions for efficient solutions on organisational, methodological and technical levels. It was pointed out that there is a need to make a distinction in the confidentiality policy concerning data on enterprises and data on natural persons. The problem can not be solved by technical means. Often the rules for enterprise data need to be redefined at the legislative level.

26.   The need for a data warehouse approach also extends to international organisations. Decentralised management often reaches its limits. The linear model of the statistical process does not always fit into the shared production and is focused on the needs of production, not dissemination. Eurostat demonstrated its future information systems' architecture designed to encourage the interoperability and re-use of different systems and applications. It aims to overcome the difficulty of following information through the different stages of processing and the lack of links between different systems.

27.   The use of datawarehouses in many statistical offices and international organisations in a network environment might cause the exchange of data between national offices and international organisations to be re-engineered. The format of delivering data and the associated metadata requires efficient solutions. On the international level, the role of metadata in data dissemination is even more important than in the national environment. It might be necessary to develop a standard for the exchange of statistical data and metadata in a network environment. A possible solution could be XML in connection with data warehouse and network technologies. It might be necessary to develop a standard XML-based language for the exchange of statistical data. An initiative to develop such a language was launched at the Work Session on statistical metadata in November 2000.

28.   Using data warehouses as a basis for data access via Internet will require new functionalities and methods. On the one hand, the data warehouse creates demands for new functionalities to be provided on the Website. On the other hand, the fast development of the Internet technology offers new possibilities that can be used by statistical Internet websites. From the management view, it is necessary to observe this development, and to draw conclusions for the work. Several participants pointed out that Internet search engines are not compatible with statistical queries. Special mechanisms need to be developed for searching statistical data over databases on Internet.

29.   The main focus of the activities of a NSO should be the user. The development of data warehouses must lead to significant advantages for the users. The data warehouse technology applied in an efficient way will provide a faster and more flexible way to access data. It should also make the data access much easier. The presentation methods should include flexible graphical and geographical visualisation which can lead to a new quality of statistics presentation. User satisfaction and user behaviour requires continuous attention. User groups have different requirements and their behaviour changes over time.

C.   Integration of statistical (survey) data with registers (administrative) data

30.   Many statistical offices are using a mixture of administrative and survey sources. Surveys may be needed to assess or complement administrative sources concerning data availability, data quality, consistency over time, undercoverage and differences in concepts. In addition to administrative registers collected by public authorities, the data collected inside an organisation can be used for its own administrative purposes, e.g. book-keeping or administration of wages. A common identification number (e.g. the Person Number) is used in several countries to link different registers.

31.   The trend of increasing use of administrative registers for statistical purposes can be observed in different countries. Tools and practices are evolving in the same direction, and more cooperation and moving towards standard solutions would make the initial developing phase much easier for countries starting with similar systems. An international agreement on common definitions and concepts to be measured for comparability would be helpful. Many statistical offices are considering data collection from the information systems of enterprises. Therefore, it would be useful if binding international standards for book-keeping systems could be agreed upon. Also, harmonisation of the rules for processing data in different fields of statistics within statistical offices would enhance further development in this area.

32.   The development of register-based statistics requires a lot of investment in the development phase and can be extremely demanding on resources. However, once the initial investment has been made, the NSI can save considerably in the daily operations. Resources will also be saved on the respondent side. There is no doubt that both administrative and statistical sources will have to be used in future for the production of statistics. However, there are many practical questions to be solved to make this feasible.

33.   The great advantage of administrative sources is that they have almost full coverage, possess a high degree of consistency and include longitudinal data. They are generally highly acceptable to reporting units. Advances in computer technology and record linkage accuracy have significantly increased the feasibility of the expanded use of administrative records for statistical purposes.

34.   The problems range from data quality and comparability to feasibility and legal aspects to resource considerations. Considerable work is required to turn administrative data into statistical data. This work includes reprocessing the administrative data, providing overall consistency and enriching the data through specific surveys. Huge volumes of data make data management cumbersome, take too long to process, and it is complicated to make any changes in the questionnaires. The quality of register-based data depends to a great extent on the quality of the underlying registers.

35.   Management issues include setting up cooperation with respondents or agencies, cost recovery and handling of changes in administrative sources. They also include metadata management and the ability to compare and adhere to national and international standards. It is important not to forget the associated human resources side and the culture and mindset that go together with changing the orientation of statistics towards more use of administrative data.

36.   One of the most important problems with administrative data is their validity and relevance. Very often the data are not in harmony with statistical concepts. In some (few) favourable cases, the statisticians might have some influence over the content of registers. Another problem is that changes in the legal and administrative system change the data content of administrative registers. Registers are kept for specific administrative purposes and they can not therefore statisfy the needs and wishes that serve statistical purposes. Also, the concepts or person groups who do not exist in the registry might also not show up in statistics. It is important that there be some basic registers that would record all units. When registers can not give a complete picture of society, it is necessary to collect additional data via surveys. Concerning timeliness, the situation differs from one register to another. Some can provide rapidly up-to-date data, others have a time-lag of several months (e.g. a tax register).

37.   The legislation can pose problems for the use of administrative data from the confidentiality viewpoint. In some cases, the use of registers may be forbidden by law because of the danger of disclosure. An interesting example was provided by Sweden, where by law it is a crime to try to disclose personal data. This approach makes it much easier for the statistical office to release anonymised microdata files since, in the case of breach of confidentiality, the person who has divulged the data will be

held responsible, not the office releasing the data.

38.   The importance of solving the technical aspects for register-based statistics was highlighted. These include securing electronic transfers over the Internet, managing versions of data formats and metadata, proper identification of respondents and statistical units and the management of this identification. Efficient tools are needed to derive, calculate and document data from different sources. Often, no commercial tools are available and the offices develop their own tailor-made solutions.

39.   Technical opportunities could be the use of XML, setting up XML schemas servers, business to business software availability such as e-marketplaces, enterprise application integration software to manage data transformation and management, and certificates for identification over the Internet. Some initiatives in this area, such as ebXML, UDDI and XML/EDI, provide an environment in which electronic business can be conducted (registry services).

40.   Among the best practices in register-based statistics can be mentioned deriving data from commercial off-the-shelf software, setting up groupware and Extranets to facilitate projects covering different organisations, a central repository for metadata, and a central repository for handling all incoming and edited data.

41.   Many participants highlighted the need to obtain more information on how the register-based statistics are prepared in practice in more advanced countries in this area. It would greatly facilitate the task for countries that are in the initial stages of setting up their register-based statistical systems. In some countries (e.g. USA, Denmark), best practices in this area are collected and documented. The meeting expressed willingness to prepare a methodological material on best practices for register-based statistics. The Nordic countries (e.g. Denmark, Sweden and Finland) and France could contribute to this material. It was agreed that such a methodological material could be of great help to countries. An efficient and visible tool for this could be a Website

D.   Resource management in statistical offices and the role of it departments

42.   The statistical agencies must re-examine their management practices in the light of changes in information technology and the increasing use of the Internet. Advances in technology are enabling statistical offices to move towards organising the production process around functions such as data collection, processing and dissemination. The initiatives for e-government and developments in electronic

reporting also have an impact on the tasks of the IT managers. The meeting considered how these changes affect the resource management in statistical offices and the role of IT departments. Some solutions to resolve the new management challenges concern internal cost recovery systems, teamwork and distributed processing.

43.   Statistical offices are using different cost recovery strategies, for example the central IT unit charging a fee for its services or a flat rate per employee, the agency charging outside customers for data disseminated over the internet, or obtaining direct funding from the government. An example of an internal cost recovery (ICR) system was presented by Canada. The IT services are managed as a business unit that recovers full costs by charging fees for internal services, including funding for capital acquisition and replacement. There is no separate IT funding, so the program manager can choose the mix and quantity of services needed to achieve his program objectives, and to make economic choices among internal and external suppliers. The advantage of the ICR mechanism is that business requirements drive the provision of central services. The costs are associated directly with projects which allow tracking the human resource expenses and to make strategic choices when budget cuts have to be made or funds reallocated. The ICR approach has had a fundamental effect on the culture and behaviour of the organisation. It has proven itself flexible in accommodating new requirements in changing situations. The transparency of fund allocation engenders trust between the service areas and the programs.

44.   The organisation of teamwork promotes interdisciplinary teams and permits using human capital more efficiently. A possible teamwork solution replacing traditional hierarchical organisation with a flat, self-directed organisation was demonstrated by Finland. The advantages of the teamwork approach mean that the division of work is more flexible, substitutes are more easily available, cooperation and interactive working methods are developed and job satisfaction is increased.

45.   The role and influence of the IT department and its manager depend on the organisational structure of the agency. It can vary greatly in centralised or decentralised environments and each environment requires a unique approach. Opposite trends can be observed in this regard in statistical offices. On the one hand, the increasing use of networking entails the move to distributed processing. On the other hand, the data warehousing and integration of production processes calls for the centralisation of IT services.

46.   Distributed computing in a large statistical organisation has many advantages, such as broadening access to human resources, expanding research, creativity, and increasing flexibility of financial resources. Risks include difficulty of leading and maintaining a distributed IT environment without direct centralised control, and an increased need for co-ordination and for software and hardware standardisation. Certain enterprise-wide IT resources (e.g. core infrastructure, telecommunications) and processes, like data management, metadata, research and development, demand strong and central control by the IT manager. This can affect strategic and operational planning throughout the organisation. Methods are needed to promote common standards, practices, and quality control.

47.   The IT manager (Chief Information Officer) must actively lead IT resource management in both centralised and decentralised environments. His role becomes especially critical in a decentralised environment where some sort of a "friendly dictatorship" might be needed. The management approach of IT managers in a distributed environment requires a new set of skills, like interpersonal skills, team building, corporate orientation, vision, salesmanship and leadership. These go well beyond the traditional technical excellence qualification associated with the position. At the same time, an important issue is educating the IT managers to be knowledgeable and competitive concerning the latest technical developments.

48.   The challenges are not only in managing but also in acquiring the IT resources. Hiring and retaining top IT professionals is difficult due to intense competition for high quality human resources. As statistical offices can not always compete with salaries in the private sector, it is important to retain a proper mix of IT resources between outsourced and internally provided services. Another continuous challenge is to acquire financial resources to retain the level of innovation and improvement in IT. The technology is changing quickly and prices for advanced technology are increasing while budgeting remains conservative. The issue of data security and confidentiality raises the need for additional investment in firewalls and other security measures. Often, it is not easy to justify the cost of infrastructure. The funds are typically distributed according to the number of staff, not taking into account the relative complexity of IT areas. Special indicators and the comparison of the situation in different countries would be helpful in this respect.