UNITED
NATIONS

E

Economic and Social
Council

Distr.
GENERAL

CES/2000/25
8 May 2000

ORIGINAL : ENGLISH

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

Forty-eighth plenary session
(Paris, 13-15 June 2000)

REPORT OF THE APRIL 2000 WORK SESSION ON METHODOLOGICAL ISSUES INVOLVING THE INTEGRATION OF STATISTICS AND GEOGRAPHY

Note prepared by the Secretariat

1. The meeting was held from 10-12 April 2000 in Neuchâtel, Switzerland. It was hosted by the Swiss Federal Statistical Office. It was attended by participants from Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Israel, Latvia, the Netherlands, Norway, Poland, Slovenia, Spain, Sweden, Switzerland, The former Yugoslav Republic of Macedonia, United Kingdom, and United States. The European Commission was represented by Eurostat.

2. The meeting was opened by Mr. Carlo MALAGUERRA, Director of the Swiss Federal Statistical Office, who welcomed the participants and wished them success in their work and an enjoyable stay in Neuchâtel.

3. The provisional agenda was adopted.

4. Mr. Dick Meuldijk (the Netherlands) was elected Chair and Mr. Rainer Humbel (Switzerland) was elected Vice-Chair.

ORGANIZATION OF THE SESSION

5. The following substantive topics were discussed at the meeting:
(i) Policy and organizational aspects in GIS and statistics;
(ii) Spatial database management and (geo-)data warehousing;
(iii) Internet and Intranet solutions;
(iv) Spatial analysis in a statistical context and disclosure control procedures;
(v) Guidelines on Cartography for Statistical Purposes.

6. The following participants acted as Session Organizers: Mrs. Margaret WAGGET (United Kingdom) for topic (i); Mrs. Randy FUSARO (U.S. Bureau of the Census) for topic (ii); Mr. Danny WALL (Canada) for topic (iii); Mr. Gilles DECAND (EUROSTAT) for topic (iv); and Mr. Alistair CALDER (United Kingdom) for topic (v).

7. The topics were discussed on the basis of papers and demonstrations prepared by Canada, Estonia, Eurostat, Finland, France, Kazakhstan, Norway, Poland, Slovenia, Sweden, Switzerland, United Kingdom and United States.

8. The invited papers were prepared by the following countries:

- Topic (i): Eurostat, Switzerland, United Kingdom and United States;
- Topic (ii): United States;
- Topic (iii): Canada and Finland;
- Topic (iv): France and United Kingdom
- Topic (v): Finland and United Kingdom.

9. The Work Session discussed and approved the Guidelines on Cartography for Statistical Purposes prepared by the United Kingdom and Finland. It recommended that the Conference approve them so that they can be made available on the ECE's Website. It was agreed that the United Kingdom and Finland should further administer and maintain the Guidelines. The Work Session encouraged the countries to use the Guidelines and to inform the secretariat, United Kingdom and Finland on the feedback. It was also recommended that the Guidelines be revised in two years' time.

FUTURE WORK

10. The Work Session recommended organising the next meeting in September 2001. Estonia offered to host that meeting.

11. The Work Session recommended organising a half-day training session on GIS implementation for interested countries back-to-back with the Work Session.

12. The following substantive topics were proposed for discussion at the 2001 meeting:

(i) New opportunities created by cooperation and partnership;
(ii) New technological solutions, including those based on on-line data access;
(iii) Spatial analysis;
(iv) Standards and metadata.

OTHER BUSINESS

13. The participants expressed appreciation for the hospitality and excellent organisation that the Swiss Federal Statistical Office had provided for the meeting, and expressed their warm thanks to the organisers.

14. The participants adopted the report of the meeting at its closing session.

15. The more detailed summary of the discussion that took place at the meeting on the five substantive agenda items will be prepared before the end of April 2000, and distributed (in English only) to meeting participants and other interested parties, upon request.

SUMMARY OF THE MAIN CONCLUSIONS REACHED AT THE WORK SESSION ON METHODOLOGICAL ISSUES INVOLVING THE INTEGRATION OF STATISTICS AND GEOGRAPHY[1] (10-12 April 2000, Neuchâtel, Switzerland)

Note by the secretariat

A. Survey on GIS use in statistical offices

1. This agenda item was considered based on the analysis of the results of the *Survey on the implementation of GIS in Statistics* completed by countries in 2000. It was a continuation of a Survey on GIS Use in Statistical Offices conducted in 1993, 1997 and 1998. The 2000 survey included more detailed questions on GIS software and applications in different statistical subject-matter areas.

2. The survey proved the increasing interest of countries in GIS technology. Applications were reported in many different areas of statistics, such as population census, social and demographic statistics (health, justice, education, labour), economic statistics (business surveys, trade, transport, tourism, agriculture, etc.) and environment statistics. GIS is used in all the different phases of statistical production, i.e. in planning, data collection, data processing and data dissemination activities (including ensuring confidentiality). The number of cross-sectoral and inter-agency projects is also increasing.

3. The situation regarding the use of diverse GIS software is similar in all the responding countries, which facilitates the exchange of experience. Taking into account the scale of use of GIS in the area of statistics, statistical offices could form a reasonably strong interest group to influence the software producers to implement GIS software functionalities required for statistical purposes.

4. The challenges in GIS implementation involve dealing with rapid changes in technology offering new solutions. The main problems mentioned in GIS implementation were data creation and management, personnel issues, system acquisition and development. New concerns are related to the upcoming population census and Internet GIS solutions. In future, the questionnaire could concentrate more on the Web-based GIS as this is an increasing area of GIS applications.

5. The Work Session recommended continuing with the survey in future and comparing the results of different GIS surveys conducted over time

[1] See the Report of the April 2000 Work Session on Methodological Issues Involving the Integration of Statistics and Geography, No. CES/2000/25, submitted to the 2000 Plenary Session of the Conference of European Statisticians.

permitting analysis of changes and identification of trends in GIS use in statistical offices. It was mentioned, however, that it is difficult to construct time series because of significant changes in the content of the questionnaire from survey to survey.

B. Policy and organisational aspects in GIS and statistics

6. The work session considered this topic from two aspects: the use of GIS as a tool for integrating data to provide better information, and the management of geographic and statistical data so that they can be integrated in a such a way as to release more information. The presented papers clearly indicate the benefits of GIS as a tool in integrating data.

7. Although robust GIS tools for data integration exist, there is not yet a breakthrough in this use of GIS in many countries. There was general agreement that political will is a necessary precondition to improve the GIS role in data integration.

8. A number of issues were considered, e.g.:

- the need for a common geographic framework;
- the importance of standard approaches and harmonisation;
- the ability to rapidly identify and access the right data;
- the need for good quality statistical and other non-graphical data which can be easily integrated into the common framework;
- the importance of rapidly disseminating the appropriate information to the decision-maker concerned.

9. There was common agreement that the main obstacles to data integration are not technical but managerial and organisational. There is often a lack of awareness of the opportunities offered by GIS applications and a reluctance to share data. Among the problems to be addressed at the policy level are the pricing and copyright issues of geographic and geo-referenced statistical data, questions of data ownership, intellectual property rights and fear of loss of control over data. It was also highlighted that a legal framework is needed to regulate these issues to encourage the sharing and integration of data.

10.Different solutions were demonstrated for coordinating this work at government level in different countries (e.g. Working Groups, Geographic Alliance). It was pointed out that if there is political will to integrate data and improve GIS coordination, there is a much better opportunity of achieving it (e.g. the demonstrated case of hurricane Mitch). However, there is a need for a continuous political support to keep the momentum when these initial factors cease to exist. On the other hand, better opportunities for data dissemination that arise from the spatially integrated data can generate interest and customer demand.

11.Several hindering factors are associated with data: its availability, the lack of awareness of what data are available, different data models used, different standards and formats used, different quality levels, the need for a clearer spatial referencing framework, absence of a data dictionary and different terminology.

12.Although the technology does not pose the main problem, there are some technical issues that still need to be considered. The meeting identified the need for exchange of experience in writing and using statistical programs in connection with GIS, and the further assistance in integrating statistical analysis features into GIS software. Among the major impediments in using GIS as an efficient tool for data integration were mentioned the lack of a common spatial referencing system, the lack of metadata and its insufficient quality, standardisation issues, etc..

13.The main issues concerned the data aspects and the obstacles of data sharing. It was considered important to make a distinction between short-term and long-term solutions. The short-term solutions are often influenced by the existing data collection and processing environments in the statistical offices. The starting point could be to make an inventory of the available datasets, to describe them (metadata), and to advertise these datasets (catalogs). It would also be important to identify a unified core set of information. The long-term solutions could include the development of common data models and standards, and the gradual adoption of the object-oriented models.

14.The meeting agreed that the Work Session presented an opportunity to assist in helping statistical offices in the creation of metadata by providing examples of experiences gained by those countries with a good metadata source.

15.Eurostat reported on the Commission initiatives to establish a European Geographical Information Infrastructure (EGII). The Commission has a key role to play in this context:

- to facilitate the creation of the administration and the mechanisms of EGII;
- to stimulate the broader geographical information (GI)-debate and increase awareness;
- to act as a catalyst initiating and promoting direct action in priority areas.

In its Green Paper entitled “Public sector information: a key resource for Europe”, published in January 1999, the Commission expresses concern about the low level of dissemination of public sector information by EU countries. GI is no exception to this rule. European diversity in data policy and its interpretation, in data specification, in pricing and access rules, in private/public sector relationships, is holding back the development of a single market and the growth of the European economy. In the near future, in Europe, the governments should freely access and exploit GI in decision-making, and in solving the pressing political bottlenecks in society such as

social exclusion, security and health. The public sector should work in full partnership with, and encourage, the private sector to provide information via electronic systems. European citizens should use GI for participation in the public decision-making processes, and use many services. GI use should be far more entrenched in education programs. The research community should build up a knowledge infrastructure to be exploited through many knowledge centres.

16.Furthermore, the Work Session was informed about the European Umbrella Organisation for Geographic Information (EUROGI), whose aim is to promote the definition and adoption of a European policy on geographical information and the development of a European geographical information infrastructure. EUROGI also promotes the establishment of national geographical infrastructures.

C. Guidelines on Cartography for Statistical Purposes

17. The updated version of the methodological material “Guidelines on Cartography for Statistical Purposes” was demonstrated. Some issues related to the further development of this material were discussed. For example, information about hardware requirements for GIS software, good and bad examples of maps, tips on what to avoid when producing maps, etc. were some suggestions made for inclusion in the material comments on using specific software.

18. It was agreed to include some good and bad examples of maps in the existing version of the material and to make it available on the UN/ECE Statistical Division’s Internet site as soon as possible. The secretariat will investigate possibilities to translate the material into other UN/ECE official languages. National statistical offices were encouraged to include a link to this material on their homepages. The secretariat will also keep evidence on statistics of users of this methodological material.

D. Spatial database management and (geo-)data warehousing

19.Several important issues concerning spatial database management were discussed. These included standards, metadata, spatial data quality and common tools.

20.The impediments to spatial database/data warehouse development were considered. It was found that the main obstacles are not technical but organisational: arriving at an agreement on standards, definitions, core data set, etc.. When including data from multiple sources, it has to be either taken as it is or transformed. Some participants mentioned, however, that rapid technological development could have a significant impact on spatial data warehousing in the future.

21.Metadata standards are recurring essential items for geodata warehousing. Metadata includes information on how data are collected, defined, processed and checked for quality. There was general agreement that metadata related to the content of statistical information are developed and maintained by national statistical offices already. Statistical classifications and nomenclatures developed internationally create an important framework for those content-oriented statistical metadata. Bearing this in mind regarding metadata in relation to spatially referenced statistical information, special geographic characteristics (e.g. related to boundaries) should be defined and elaborated to complete the description of statistical information in a geographical context.

22.Spatial data quality issues are important in this respect. Different standards (e.g. the Spatial Data Transfer Standard, ICA Commission on Spatial Data Quality) specify the major metadata elements of lineage, positional accuracy, attribute accuracy, completeness, logical consistency, semantic accuracy and temporal information to comprehensively describe the quality of a data set. However, the quality elements are described on a conceptual level without specifying concrete measures for individual quality parameters. Therefore, subsequent research has focused on the development of such measures. The necessity of information on processing for full comprehension of how the data were derived was emphasized.

23.Some activities in research on formalised description of data quality and some attributes describing the content and other characteristics of data have been reported. Standards such as the "Content Standard for Digital Geospatial Metadata" (CSDGM, Federal Geographic Data Committee, 1997) provide a common set of terminology and definitions for the documentation of digital geospatial data. Since about 200 different data elements are defined, the interpretation and assessment of a data set concerning its suitability for a specific application is quite difficult.

24.The development and integration of tools that enable fast access to metadata is therefore important. The tools consist of "de-facto" standard and freely accessible software applications which mainly allow for graphical assessment of data suitability. The main characteristic of the tools is their interoperability, which may be defined as their capability to communicate with other software tools and their exploitability from different platforms and operating systems.

25.The clearinghouse concept for spatial metadata is very promising for documentation and keeping track of spatial data. Development of a well-fitting metainformation system based on relevant international standards is an important prerequisite for the implementation of these projects. In this respect, the Work Session noted with great interest the information on the implementation of a geospatial data clearinghouse by the U.S. Federal Geographic Data Committee.

26.A clearinghouse provides a virtual information space across which searches may be conducted through a single query. The variety of user interfaces are accomplished through use of a common descriptive vocabulary

(metadata), a common search and retrieval protocol, and a registration system for servers of metadata collections. The clearinghouse includes data, ordering mechanisms and map graphics for data browsing. Metadata fulfills three roles, documenting i) the location, ii) the content and structure, and iii) the use of information.

27.Special attention was drawn to the implementation of standards developed under the ISO Technical Committee 211. It was the first practical presentation of these standards presented in the framework of the Work Session. The opinion was expressed that it would be most desirable to prepare practical guidelines aimed at advising national statistical offices on how to use ISO standards for building clearinghouses of statistical spatial data.

28.In future, it would be important to develop common standards and practices in building spatial data infrastructures. The Global Spatial Data Infrastructure (GSDI, 1999) activity is an international effort to raise awareness of the supportive policies and technical standards that promote compatible geospatial data access. Rather than invent new standards, the GSDI participants identify the best practices and standards being developed in regional and international settings that are relevant at all scales of application.

29.Many countries reported new developments in the design and implementation of corporate statistical data warehouses based on distributed data processing and networking. An important new feature is in many cases a globalisation and integration of information on the government horizontal level. The geographical database should be an implicit part of these corporate database projects. They should be developed in close cooperation with national mapping agencies.

30.Geospatial and numerical data stored in a relational database management system (RDBMS) allows multiple clients to access data and use the data with platform diversification. This integrated system promotes increased data security, data consistency and data integrity.

E. Internet and Intranet solutions

31. The considered applications demonstrated that Internet/Intranet is clearly the medium of choice for disseminating information. The main benefits of using Internet/Intranet for the dissemination of geo-referenced data mentioned were: making spatial data widely available, increasing awareness about the possible uses of spatial data, enhancing the homogenisation of data, capacity for spatial analysis, integration of datasets, etc..

32.Internet can be an ideal platform to provide maximum use of (corporate) spatial databases. Internet bridges the complex nature of GIS with a user-friendly interface that provides GIS capabilities to users in a clear and simple manner. Most users want to be able to use GIS functionality to access

and display information. Other sophisticated users want to have advanced spatial analysis capabilities. These GIS functions can be provided to all users via a Web browser using new Internet GIS tools which have been developed by software vendors.

33. Everyone senses that there are tremendous opportunities associated with these applications. Most of the statistical offices have plans to deploy, or have already deployed, Internet solutions for the dissemination of statistical data. There was considerable discussion surrounding possibilities to present geographic displays over Internet/Intranet.

34.The Internet environment further allows for efficiencies in a distributed server model. Subject-matter data can be housed in remote data servers and can be related to a spatial data server by a common geographic key. This permits the integration of subject-matter data that is crucial for the spatial analysis needs of a statistical agency.

35.At the same time, it is clear that there are some obstacles to overcome. Among these:

- lack of data to drive applications remains an issue;
- it is clear that building Internet solutions requires more than just geographic information knowledge, there is also a requirement for widely based information technology knowledge;
- current software solutions lack solid statistical classification routines.

36.Furthermore, pricing and pricing methods remain problematic for the advancement of Internet data applications.

37.The shift from traditional cartographic methods to digital GIS techniques requires a significant investment. There are very high initial short-term costs and ongoing long-term expenses. Short-term costs include the conversion of analogue information to digital data, the purchasing of computers, GIS software and other spatial data, and the hiring and training of staff. Long-term costs include ongoing operational expenses such as system maintenance and upgrades, spatial database updating and staff training. A GIS infrastructure must be viewed as a long-term investment in order to maximise its benefits.

38.Overall, there is a general feeling that Internet/Intranet applications will evolve over the next few years. But, as with any new technology, it is important to proceed with caution.

F. Spatial analysis in a statistical context and disclosure control procedures

39. Spatial analysis was considered from different viewpoints: its benefits, factors that hinder its development, the importance of defining user requirements, and the relation between spatial analysis and disclosure control. Since spatial analysis, particularly its use in statistics, is a relatively new area, there are often no ready-made solutions and the problems have to be considered starting from a rather basic level.

40. Several spatial analysis applications, spatial transformation methods and attempts to standardise the spatial analysis methods were demonstrated.

41. Spatial analysis adds a new dimension to data, thus making data analysis and integration easier, and takes into account the spatial relationship between data which can sometimes have an influence on their characteristics. Statistical organisations spend a lot of money on collecting data; spatial analysis can add value to the data with this (relatively inexpensive) analysis. Spatial analysis could be particularly suitable for describing the modern mobile society where people are no longer tied to a certain (administrative) area.

42. The following factors hindering the use and development of spatial analysis were mentioned: it is (alleged to be) complicated, statisticians are not aware of its possibilities, the tools are not yet sufficiently developed and are not included in the GIS software. Also, statistical offices' requirements for spatial analysis are not well-defined. As the statistical offices form a rather limited market, they often need to develop their own solutions for spatial analysis and use the available tools as best as possible. In this area, international cooperation between statistical offices is of great value.

43. The difficulty in explaining the results of spatial analysis to citizens and politicians was another important issue mentioned in connection with its complexity. In some cases, it might prove efficient to run pilot projects in selected administrative areas that could make up a showcase. Also, users need to be aware of the kind of decision-making in which spatial analysis can be appropriate.

44. An increase of geo-referenced statistical data at the local level raises problems of confidentiality. The procedures for ensuromg disclosure avoidance are normally based on established census geography. It is not known how the disclosure procedures should be applied when users merge datasets and tabulate data for user-defined geographic areas. An example might be a case where overlaying data is reported for different geographies and simple arithmetic can be used to provide estimates of census data for small areas of a size less than thresholds. Specific tools need to be developed for ensuring the confidentiality of spatially referenced statistical data.

45.There was general agreement that the spatial analysis topic requires further attention and consideration at future work sessions. The need to

share experience, expertise and ideas in this area was especially highlighted. An appropriate method could be the proposed Website on spatial analysis linked to the methodological Website on GIS and statistics hosted by the UN/ECE Statistical Division, as described above (para 17-18).

G. Future work

46. The Work Session recommended to organise the next meeting in September 2001 with the following agenda:

(i) New opportunities created by cooperation and partnership

This agenda item will focus on the "soft" side of matters, and the expected contributions can deal with:

- cooperation with National Mapping Organisations;
- outcome from work with value added partners;
- experiences on "how to sell an idea" to other parties, and to demonstrate its benefits;
- intellectual property rights, copyrights, pricing, licensing, etc.;
- sharing costs and profits;
- how to make partnership sustainable in the long run, value chains;
- partnership in terms of in-house, interagency, international;
- policy issues;
- data integration from the organizational point of view.

(ii) New technological solutions, including those based on on-line data access

This topic emphasizes the technical side of matters, and is expected to focus on:

- solutions fueled by new developments in technology (wireless, WAP, RDBMS);
- changes in organization of work processes due to the introduction of on-line systems (Internet, Intranet, etc.);
- database management, modelling etc. issues for enabling on-line applications;
- data integration;
- application of spatial data presentation techniques.

(iii) Spatial analysis

This session could consider the following topics:

- spatial analysis, examples of its use in statistics;
- data mining using spatial analysis;
- human perception and communication issues;
- error propagation and handling;
- understanding data being analyzed;
- modelling statistical phenomena in geographical space.

(iv) Standards and metadata.

This topic focuses on standardization as an enabler of cooperation and data deployment:

- the work of ISO TC/211;
- open-GIS solutions and their benefits;
- metadata: does it come before or after data?
- open interfaces as a key element to enable cooperation;
- standards enabling data integration;
- spatial data quality standards and measuring quality.

47. The Work Session also discussed the possibility to organise an invited lecture from a key figure in the geo-statistical industry (1-2 hours). The aim of this lecture would be to report on important developments in diverse fields in this area, to explore the potential impact of the geo-statistical industry on the future development of statistical production in national statistical offices and to initiate discussions with the geo-statistical community for the benefit of both sides.

48. The participants agreed that it would be desirable to organise a one-day or a half-day workshop back-to-back with the Work Session on practical methodological issues (problems and opportunities) concerning the integration of statistics and geography. Training in concrete problems of data capture, database management systems, spatial analysis, mapping, dissemination and customer relations were mentioned as potential fields of interest.