

UNITED  
NATIONS



**Economic and Social  
Council**

Distr.  
GENERAL

CES/2003/24  
15 January 2003

Original: ENGLISH

---

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

Fifty-first plenary session  
(Geneva, 10-12 June 2003)

**REPORT OF THE DECEMBER 2002 JOINT UNECE-EUROSTAT WORK SESSION ON  
REGISTERS AND ADMINISTRATIVE RECORDS  
FOR SOCIAL AND DEMOGRAPHIC STATISTICS**

Note prepared by the Secretariat

1. The Joint UNECE-EUROSTAT Work Session on Registers and Administrative Records for Social and Demographic Statistics was held in Geneva from 9 to 11 December 2002. It was attended by delegates from Austria, Cyprus, Czech Republic, Denmark, Estonia, Finland, Hungary, Ireland, Israel, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom and the United States. The European Commission was represented by Eurostat. A representative of the International Labour Office (ILO) was also present.
2. The Provisional Agenda was adopted.
3. Mr. Petrus Everaers (the Netherlands) served as Chairman, and Mr. Paul van der Laan (the Netherlands) served as Vice-Chairman.

**Organisation of the meeting**

4. The following substantive topics were considered in separate sessions at the meeting on the basis of 14 invited and supporting papers:

Session I: Quality considerations in the use of administrative data and registers for statistical purposes (Organiser: Mr. Alex Clark, ONS, UK)

Session II: New developments in methodology and innovative applications of information technology in using registers in social statistics (Organiser: Mr. Paul van der Laan, Statistics Netherlands)

Session III (information session): Progress reports by Eurostat on harmonisation and integration, and specific research programmes under the “Fifth Framework” (Report by: Mr. Aarno Laihonon, Eurostat)

Session IV: Use of business register in social statistics (Organiser: Mr. Folke Carlsson, Statistics Sweden)

5. Conclusions reached by the participants during the discussion of the substantive items on the agenda will be presented in the annex to the report to be prepared after the meeting and distributed to participants before 15 January 2003 (in English only).

### **Recommendations for future work**

6. The participants were informed that the Bureau of the Conference of European Statisticians, at its October 2002 meeting, recommended that future meetings in the CES work programme have a specific purpose and/or product in mind, and that they should be prepared by a Steering Group composed of national experts.

7. A small group of participants composed of the Chairman, the session organisers and the representatives of Eurostat and the UNECE discussed the proposals on future work submitted by the delegates, taking into account the recommendations of the CES Bureau.

8. The group proposed that a Steering Group be appointed to elaborate a concrete proposal for future work to be submitted to the CES Bureau at its February 2003 meeting, describing the proposed output, the process for achieving it and the time framework. The group proposed that the Steering Group be composed of representatives of Israel, the Netherlands and the United Kingdom.

9. The Steering Group should approach a small number of countries with experience in this field and ask them to prepare methodological material on the use of registers and administrative records for social and demographic statistics. The material should draw on relevant work being undertaken in other fora, including for instance the Website on IT Practices in Statistical Offices (<http://cs3-hq.oecd.org/scripts/stats/itp/index.asp>). A draft version of the material could be presented at a future joint meeting on registers and administrative records for social and demographic statistics in the CES work programme.

10. The meeting approved the proposal of the group on future work.

### **Adoption of the report**

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

## ANNEX

### SUMMARY OF THE MAIN CONCLUSIONS REACHED AT THE WORK SESSION

#### **Session 1: Quality considerations in the use of administrative data and registers for statistical purposes**

Documentation: Invited papers by United Kingdom (2) and the United States; Supporting papers by Italy, Hungary, Slovakia and United Kingdom.

1. Session organised by Mr Alex Clark (Office for National Statistics, UK).
2. The meeting discussed the experiences of countries that are developing or exploiting the use of administrative records for the production of small area statistics to enhance or to possibly replace decennial censuses and for information about small areas in the period between censuses. Administrative records have been used for some time to support census related tasks and for National Accounts, but there is now growing interest in using the data at small area level, NUTS IV and V. It was recognised that administrative records offer access to a wide range of data and are a major resource to be exploited to provide local and central governments, the public and researchers with a wealth of information as well as reducing the cost of collecting data and respondent burden. In addition to this with the rising cost of decennial censuses the move to an administrative record census, as a full or partial alternative to carrying out a Census, was seen to be inevitable for some countries, but that there are a number of obstacles to be overcome before doing so. The meeting considered key issues relating to harmonisation and standards, the quality of data, confidentiality and disclosure and data protection. Administrative records were also discussed as a potential source for improving data in those countries with population registers.
3. Quality: The presentations and discussion highlighted the importance of measuring the quality of administrative datasets, coping with change over time and conceptual changes in the underpinning administrative systems. It was agreed that it is essential to provide quality reports and comprehensive metadata for each administrative record dataset in order to keep the data user informed of any limitations of the data, give a clear description of how data was collected and maintained, and also to provide information about the impact of edit and imputation procedures to cope with missing data.
4. Quality measurement processes based on the Eurostat Quality Model in the European Statistical System were seen to be essential for this process. The issue of how administrative record datasets can be accessed and maintained without the legal provision of a Statistical Law was raised by participants in regards to one of the UK presentations. This also raised a question about how to maintain the commitment from data owners to produce data to quality standards without legal compunction. With or without a Statistical Law co-operation between National Statistics Institutions and the agencies providing and maintaining administrative records is essential and it would be good practice to set out the arrangements for acquiring and maintaining in a formal agreement between both parties. .
5. Consistency: Several concerns were raised by the participants about consistency between different administrative sources and, in particular, how this issue was likely to be more significant for locally held datasets. The problem of harmonisation of geography data definitions and classifications is an issue that needs to be recognised and a programme needs to be introduced to set standards such as a

standard statistical unit for geography and the use of international classifications and definitions. The compulsory use of international classifications and systematic harmonisation of different administrative sources was discussed in the presentation by the Statistical Office of the Slovak Republic. Classification and conceptual changes are also relevant to the construction of time series. Double coding of variables to bridge the changes introduced by new classification systems was discussed in the context of the presentation on the UK Longitudinal Study.

6. **Confidentiality:** All presentations and participants addressed the issues of confidentiality and the importance of protecting privacy of the individual data subject. Overlapping geographies and retabulation on different axes produce disclosure problems and are issues that need to be faced. The use of a standard geo-statistical unit as a building brick to aggregate statistics helps to overcome this issue, but only at the expense of introducing a degree of approximation for aggregates that straddle the standard units. A further issue raised related to the split in responsibility between the data owner and the NSI on what information can be released and how it should be presented. Countries without population registers faced the problem of linking data without a single personal identifier that is common to all sources. It was recognised that both geographic and statistical key variables would be required for matching and that decisions about the level of proxy matching acceptability would be required. A number of participants queried whether it was possible to match records at all without a unique identifier.

7. **Coverage:** All of the papers presented raised issues over the coverage of administrative registers. In the context of surveys and censuses the hard to reach groups that were commonly highlighted were young ages – babies and young adults – and old people. Many administrative record sets suffer from the same deficiencies. In spite of this, it was agreed that Administrative records can play a major role in support of decennial census activities by highlighting areas at risk of enumeration difficulties and also have the potential to be used to identify and potentially compensate for non-response.

8. Some participants raised the point that there are sound statistical methods for making estimates from poor quality data, and that these estimates can be checked and that this is a possible way forward. The discussion then moved to the issue of producing a traditional count versus statistical estimation. Most countries still see the need for a basic count, but recognise that measuring the coverage and quality of the count is not easy and that statistical estimation has an important role to play in making the best use of a mix of data sources – administrative records, surveys, registers, census counts. However, the level of acceptance of estimates by users needs to be considered and NSIs would need to develop high quality processes that would substantiate the methodology and provide the basis for supporting a culture change amongst users.

## **Session 2: New developments in methodology and innovative applications of information technology in using registers in social statistics**

Documentation: Invited papers by the United Kingdom and the Netherlands (2). Supporting paper by Denmark.

9. The session was organised by Mr. Paul van der Laan (Statistics Netherlands).

10. **Methods and tools:** Using registers and administrative sources in social statistics inevitably means that data sources have to be combined. To provide good quality statistics the maximum value added needs to be squeezed out from administrative sources. This means that tools are needed that enable to match

data in an efficient way and also enable to check and edit the data without too much human effort. Otherwise the benefits from using administrative data above collecting survey data will be offset by the efforts needed to edit the administrative data. On the output side tools are needed that will enable users to quickly find the information they are looking for, but that will also alarm the statisticians when they are publishing conflicting or confusing data.

11. Neighbourhood statistics: The systems being developed to collect, process, disseminate and use neighbourhood statistics in the United Kingdom was presented to the Work Session. The Neighbourhood Statistics project in the UK elucidates how many users have a need for local data rather than national or coarse regional data. For them the local data have higher relevance. Build upon a political motivation to improve developments at the neighbourhood level, this project has an unique opportunity to bring official statistics closer to the public. A comprehensive user-access and user-support and user-feedback facility is provided, that can find application even beyond this project, as other large statistical projects could benefit from equal dissemination facilities.

12. The meeting particularly discussed two aspects from this wide-ranging project: (a) sample size and (b) changes in neighbourhoods over one or two decades. Even though the Census is based upon tens of millions of observations and is strictly speaking no “sample”, the relevant sample size is limited by the smallest “tabulation unit”, say a few hundred or a few thousand persons. As for many person-based characteristics (like education, income, health) it is indispensable to compensate for effects of ages, sex and ethnicity. Furthermore some variables are person based, others household based or family based, so there will be only limited room for adding more variables in the model or table. Not only will this lead to disclosure risk, but also the “noise level” will rise, i.e. the variance increases if more variables are used in a sample with a fixed size.

13. The neighbourhood statistics project is intended not only to measure difference but also to measure change in neighbourhood characteristics. This is a considerably more complex task, given the fact that a neighbourhood has dynamics (“life cycle”) of its own. For example, “grey” neighbourhoods will turn over to ones with a young population where children are again born; and newly built neighbourhoods with mostly “starters” will become “middle aged” within a decade or two. These natural developments must be measured at the neighbourhood level, but must be studied at the city level, as on a larger scale they are communicating vessels.

14. Harmonisation and integration: The development of an integrated system for social statistics in the Netherlands was presented to the Work Session. Based on the concept of “society as a crystal” and a “life cycle model” relevant social statistics can be defined. The basic format of these statistics in every domain (e.g. health, education, labour, poverty etc.) is statistical units x variables x time. Using this basic format, possibilities are created to analyse the situation of persons, the transitions from one situation to another and the duration of a certain situation. This information is used by government and the society at large to monitor and to evaluate the various domains. The analysis can be domain specific, focus on transitions and durations within one domain, relations between domains, as well as relations between transitions and durations among domains.

15. All data needed in an integrated system of social statistics can be stored in a system of “baselines”. In the statistical production process all data subsequently go through these baselines: raw data (per source), matched data (combined sources), edited data (per domain), integrated data (across domains), output data (data “cubes” in an output database). In the Netherlands, the system of baselines is an instrument for consistent statistical output (the so-called “one number policy”). The desire to publish

per reference period only one figure on each statistical item creates an output inspired harmonisation effort across all social statistics.

16. In the discussion it was stressed that the use of harmonised concepts and classifications in both surveys and registers is a prerequisite in order to combine all statistical output in a consistent, user-friendly output database.

17. Data warehouses: An application of data warehouse techniques for the analysis of longitudinal population data in the Netherlands was presented to the Work Session. The Dutch population registers produce a manifold of data on vital events and changes of residence. Much of this information needs to be kept for future use and cannot simply be overwritten with the most recent information. To solve this problem of keeping snapshots as well as continuous updates, new storage solutions have been developed in business information. The creation of data marts is one of these solutions combined with OLAP engines. The OLAP engine pre-aggregates the information in such a way that it can be retrieved much more efficiently for analytical purposes.

18. The use of OLAP and Warehouse techniques will speed up statistical analysis considerably as compared with sequential files. All variables are available and linked over time. The Warehouse integrates all metadata with the figures. Rapid exploration of the data will enable analysts to build up a much better and more complete intuition about the phenomena they are studying. Besides, population registers can open a wealth of longitudinal insights, provided that the history of all records is maintained in a Warehouse.

19. In the discussion it was pointed out that during the initial stage of defining the Warehouse, it is extremely important to carefully reflect the exact operationalisation of concepts and classifications (the “ontology” of the system). This aspect is often ignored in the developing process.

20. System of registers: Some recent developments in the Danish system of register-based personal statistics were also discussed at the Work Session. The present system of registers has a lot of data redundancy. Furthermore, some data processing is repeated more than once and the system has not always the same definition of the same concept. The improvements to the register system will be: a certain variable shall always be picked up at the same place; a register will be defined as a set of rights to access certain variables; the number of physical register units will be severely limited; variables shall not be passed on to third parties. The working of the new system of statistical registers in Denmark was illustrated on the basis of statistics on personal income.

### **Session 3 (information session): Progress reports by Eurostat on harmonisation and integration**

#### Documentation: Power-point presentation by Eurostat.

21. Reports were given by Mr. Aarno Laihonon (Eurostat).

22. The objective of harmonising a set of core variables to create a backbone for integration of the European harmonised sources of social statistics and ‘best’ national sources has unfortunately suffered a “slowdown” in Eurostat because of changes in personnel and lack of resources. The methodological topics of using administrative records and registers in the field of social statistics, however, have still retained their importance. Increasing use of administrative sources at national level has been observed e.g.

in the recent population and housing censuses. One important indication also is replacement of a central statistical survey instrument “European Panel Survey” ECHP with the EU-SILC system, which recognises flexible use of national data sources (e.g. use of administrative data).

23. Even if the systematic work carried out in 1996 – 2000 on harmonisation and core variables has suffered a slowdown in Eurostat, partial progress has been made in some areas. For instance, development work has been made in the field of metadata and in documentation of national data sources (including use of administrative data). Eurostat has established a CODED database (which is available on the Eurostat web site) which contains descriptions of definitions and classifications used in statistics compiled by Eurostat.

24. The Census programme 2001 of the European Community has reached its data collection phase. Data collection covers EU member states, EFTA countries and 13 candidate countries. The table programme, which contains 40 tables with detailed annotations, forms in itself a concrete tool of harmonising census data in the European region. It has been available for countries concerned from the summer 1999.

25. Eurostat has two important documentation projects, which will improve coverage and quality of meta data in the field of population statistics: Documentation of national data sources of annual demographic statistics (covering 36 European countries) and Documentation of 2001 round national population and housing censuses (covering 32 European countries).

#### **Session 4: Use of business registers in social statistics**

##### Documentation: Invited papers by Sweden and Norway. Supporting paper by Finland.

26. The session was organised by Mr. Folke Carlsson (Statistics Sweden).

27. The invited papers described the state of art, ongoing activities and future plans on the use of business registers. The difficult task of identifying and linking employment to the correct establishment was discussed in all three papers. The papers also pointed out the entirely new possibilities that combining person data and enterprise data provide to describe labour force dynamics. In the discussion some question marks were raised concerning the basic register status given in the Swedish paper to their new register of activities. According to traditional concepts, there are three basic registers on persons and organisations (extended business register), which contains actors (decision making units) carrying out different socio-economic activities and a register on real estates and buildings containing physical spatial units (media in which actors carry out their activities).

28. The original Danish approach to describe the dynamics of enterprises and establishment by using labor force movements was discussed by both invited papers. It means that by following the changes in the composition by employees it is possible to detect merging of enterprises and establishments as well as their division into two or more units.

29. The Finnish paper described some difficulties existing in using the business register. Errors in business register may cause excessive errors in regional employment statistics (e.g. if the address of a big establishment is in a wrong municipality, all the employees of that establishment are associated to a wrong municipality). Handling of employees without a fixed work place (truck drivers etc.) is often

problematic. For instance, if they are located to the principal establishment of the enterprise (Head Quarter), they increase artificially the employment of the municipality where the HQ is.

30. Beside the presentation and discussion of the papers, short oral presentations of what is going on in this area were given in the session by Austria, Slovenia and Switzerland. Business registers were mainly used in these countries in connection with population censuses in order to determine location and branch of persons work places.

- - - - -