UNITED NATIONS SECRETARIAT
Department of Economic and Social Affairs
Statistics Division

ESA/STAT/AC.97/6
26 August 2004

English only

United Nations Symposium on
Population and Housing Censuses
13-14 September 2004
New York

Statement from the UNECE Statistical Division

By

Heinrich Bruengger
Director, Statistical Division
Economic Commission for Europe

* This document is being reproduced without formal editing.

THE RELATIONSHIP BETWEEN THE FUNDAMENTAL PRINCIPLE ON CONFIDENTIALITY AND POPULATION CENSUSES

A. Introduction

1. The Fundamental Principles of Official Statistics, adopted in the ECE region at both the statistical and political levels in 1992 (see Annex), and by the UN as a whole at the statistical level in 1994, are at the basis for all activities that are carried out under the umbrella of official statistics at national level and lead to outputs that are disseminated as authoritative statistical results. Population censuses, as the statistical activity at national level that has the greatest number of respondents and the highest cost, do not differ in this respect from all other statistical activities.

2. The fact that in many countries population censuses are regulated by special laws rather than based on a general statistical law should not be interpreted at all as opening the door for any deviations from the fundamental principles. It is rather the opposite: given the high visibility and the encompassing nature of the population census, some stricter regulations than for other statistical operations may be necessary to retain the trust of both respondents and users in the operation. Any separate census law, and any secondary legislation based on such a law, should therefore be examined closely before adoption whether it guarantees at least the same standard of compliance to all fundamental principles as enshrined in the general statistical law. If the standard is the same, a simple cross-reference to the general statistical law might be sufficient.

3. The present contribution will not go through all 10 principles (the first principle contains three principles in one: relevance, equality of access, and impartiality), but concentrates on a principle, which, from the point of view of both statistical offices, respondents and users, is a particularly important and controversial issue in the context of the various forms of population censuses that exist across the world. This principle is the sixth principle, addressing the issue of confidentiality, which states that "Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes". Links between this principle and other principles will be discussed as well.

4. For the discussion of this principle, the implicit definition of what is a population census should be made explicit, without pre-empting the definition that will emerge for the 2010 recommendations as an outcome of the process that is about to start at world level. For the ECE region, the Conference of European Statisticians has decided in June 2003 that a process leading to recommendations for the 2010 round of population and housing censuses 2010 should be organised in such a way that adoption by the Conference is possible in 2006, so this process in the ECE region is already well on the way. In the present paper, a population census is defined as a purely statistical operation aimed at an exhaustive coverage of the population resident (or present) in a given territory[1]. Given recent developments, simultaneity, i.e. coverage of all units at the same moment, will not be a condition, without going into the details of what the additional problems of a staggered population census are. However, in the context of the confidentiality issue, the author does

[1] Complete enumeration can be limited to some key characteristics, with some other characteristics collected only for a sample of units.

not see any difference between a non-simultaneous census as compared to a simultaneous one.

B. Prevention of Non-Statistical Use

5. Whereas other principles are formulated in a more general way, the sixth principle on confidentiality is unique in using terms such as "strictly" or "exclusively". This indicates how important it is for all data collection of official statistics. The willingness of respondents to participate in data collections of official statistics, whether they are censuses or surveys, and to give unbiased answers, will crucially depend on their perception that their answers will never be used in any individual decisions by the government (or other state bodies such as courts) concerning their personal rights, entitlements or obligations. It follows from this that not only the organisational measures have to be designed, for all phases of the census from data collection to dissemination and storage, in such a way that this principle is strictly adhered to, but that the compliance with this principle by all authorities and persons involved in the census is trusted by respondents. This trust does not only depend on the census legislation per se and its implementation, but also on the capital of confidence accumulated by the statistical office over a period of time. The capital of trust is dependent on the respect of all fundamental principles, but impartiality, professional independence, and confidentiality are key. Where statistical offices are perceived either as instruments of propaganda for governments, or to be linked to administrative decisions about individual natural or legal persons either directly or indirectly through providing the necessary data to another authority, they will not benefit from the necessary trust essential to soliciting unbiased information from the population.

6. One implication of the confidentiality principle (the "exclusively for statistical purposes" part) is what is sometimes described as the one-way rule between administrative data sources and purely statistical data collections such as censuses or surveys. Like in a one-way street, individual data can only flow from administrative sources and the authorities responsible for such records to statistical offices, but never the other way round. The input from administrative sources is greatest with so-called register censuses, but some information from such sources is used even in traditional censuses, e.g. to check for omissions. The fundamental principles encourage use of administrative data for official statistics (see the fifth principle), but it cannot be a symmetrical process leading to the updating of administrative registers through census information. Administrative registers[2] are to be used for individual decisions on eligibility, obligations etc. by the competent authorities, and any flow back of data kept in the statistical office, whatever their origin, to such registers is not in line with the obligation of exclusively statistical use.[3]

---

[2] Administrative registers have to be distinguished from statistical registers managed by statistical offices according to their own criteria. Statistical registers can be used only for statistical purposes (which includes the use as address list and sampling frame for statistical surveys). For statistical registers, the one-way principle is fully applicable, i.e. administrative registers (or other administrative sources) can be important inputs into statistical registers, but cannot receive back any information. The use of population censuses to set up/update statistical registers is an important indirect use of censuses that is fully compatible with the confidentiality principle if the same principle is fully applicable to the statistical register itself.

[3] The author is aware of the fact that some countries have organised a population census where at least some of the information collected about persons was used for updating administrative (central or local) population registers, with the main argument of cost-efficiency. In this case, the census ceases to be a purely statistical operation as defined above, and the legal, institutional and organisational issues around such a census become very complicated. Before such a decision is made, the possible negative implications for the overall trust into the statistical system, and into the statistical office in particular, which might be of a long-term nature, should

7. If some individual data for the census are imported from administrative files, why can they not be transmitted back? The answer is that statisticians have the right to alter such data, in order to bring them in line with the statistical concepts, definitions and classifications used. These alterations (or editing) may be based on comparing with other sources at individual level (of statistical or administrative origin), the matching of files, or on direct feedback from respondents. All these activities take place under the umbrella of exclusively statistical purpose, and any disclosure of individual data after such processing has taken place back to an administrative authority would not be in line with this boundary, even if the majority of source information used were of administrative nature. It should be pointed out that in many countries, legislation on data protection does not allow matching of files held by different administrations for different administrative purposes unless there is a specific legal basis allowing such combinations for administrative use; the usual exemption is for official statistics, backed by the general statistics or census laws, and has to be limited to purely statistical purposes and not for non-statistical use. Otherwise, statistical offices would become the place where data protection laws would be circumvented.

8. The issue of non-statistical use of the data collected in censuses is also of paramount importance for the whole chain of operations between the respondents' disclosure of the information for the census and the data arriving at the statistical office, i.e. for the enumerators or any local and regional non-statistical authorities that might be (and in many countries are) involved in this process.[4] The concrete measures to ensure that any persons who has access to individual census information (which contain identifiers in this phase) will not use it for any purposes other than those listed as their tasks in the census, will have to differ from one country to another; however, they have to be specified in the legislation, and any non-compliance has to be punishable in such a way that it acts as a deterrent.

9. An often neglected and delicate issue that is linked to both confidentiality and impartiality is the relationship of persons between households in providing information required by the census. When households are visited by enumerators, and information is obtained through interviews, it is important that each member of the household present be encouraged to give the information about himself/herself to the enumerators directly, rather than in an indirect way through the "head of household" only. In cases where this kind of "proxy" collection of information cannot be avoided (children, disabled persons, persons who are not present) should be explicitly mentioned in the legislation. Enumerators have to be selected accordingly, and be given clear instructions, on how to act in such situations. If information is collected without interviews, i.e. through some way of form filling by each respondent, the instructions given should also encourage direct reporting of information by each person. The same problem arises also for persons living in institutional living arrangements.

---

be carefully weighted against any cost saving for non-statistical authorities. The cost saving argument is often overstated, since administrative registers have to be updated continuously, and the one-time input from the census will soon become outdated for an increasing number of individual records. Respondents have to be told about the dual nature of any such census, and the willingness to participate, and to give unbiased information, may very well be affected negatively among certain population groups. For these reasons, and for their clear break with the confidentiality principle, statisticians should oppose dual purpose (statistical and non-statistical data collection combined) population censuses.

[4] Together with budgetary reasons, the assignment of important tasks in the census to local authorities that are not involved in other statistical data collection operations is one of the main reasons why special legislation is necessary for population censuses in many countries.

10. A major approach in reducing the risk of non-statistical use is the exclusion (and possible destruction) of names (and any numbers referring to lists of names) from the content variables at as early a stage in the processing as possible. This moment should be identified clearly, in view of all direct and indirect statistical uses of the census data within the statistical system. It has to be stated that even without such identifiers, microdata sets from the census with the complete set of content characteristics, whether on individuals or households, have to be protected as confidential, because of the possibility of indirect identification. (For most practical purposes, and unless detailed geographical information is removed, the protection obligation is also valid for individual information about dwellings and buildings, because there is a possibility of indirect identification of persons/households either living in them or owning them).

C. Disclosure: Issues for Statistical Use

11. As for other statistical operations, disclosure of protected census microdata without identifiers, e.g. to researchers, for statistical use is not in conflict with the fundamental principles if:

a) there is a clear legal basis in the relevant legislation, and
b) there is a clear way of ensuring that any recipient of such data will respect the necessary restrictions, e.g. not to make accessible these microdata to any third party, and not to reidentify the units, and
c) the recipient has no conflict of interest that may tempt him to use the data for non-statistical purposes; for this reason, it is not advisable to give access to census microdata for researchers in ministries.[5]

Most statistical legislations have a clause permitting a statistical office to disclose protected data for individual statistical activities outside official statistics under certain conditions. The fulfilment of the criteria has to be examined for each individual request, so that the principle of equality of access is only valid within certain criteria (or with other words, in the case of protected microdata, the principle of confidentiality is given priority to the full application of the principle of equality of access because of the wording "strictly" in the sixth principle).

12. Because of the census being an exhaustive operation, the risk of indirect identification, especially through detailed geographical information, is much higher than for individual data from household sample surveys. Some countries have therefore excluded this form of disclosure to researchers for census data even if they do for survey data. If release of census microdata to researchers is not possible on legal grounds, the statistical office must ensure a sufficient capacity for tailor-made tabulations that researchers may wish to obtain.

13. There are various ways by which statistical offices can provide access to confidential microdata without identifiers to bona fide researchers. In the framework of the Conference of European Statisticians, a group under the chair of Dennis Trewin (Australia) is about to set up a list of good practices in this respect. This work is fully documented on the ECE website.[6]

[5] Through their ownership of many administrative records, ministries have also a better possibility to reidentify individual units.
[6] http://www.unece.org/stats/documents/tfcm.htm

14. Further complications arise if the census includes certain characteristics that national data protection legislation defines as particularly sensitive such as health, ethnicity or religion. The first question to ask is whether such characteristics should be included in the census at all, because some persons might object against the collection of such data, even if it is for purely statistical purposes, as being too intrusive into their privacy (burden for respondents is mentioned in the fifth principle, and a question considered to be intrusive is perceived as more of a burden). Other, less burdensome options (collection of such data in sample surveys without any response obligation) should be carefully considered as alternatives that may produce even better results at the aggregate level. If countries, mainly because of the need of small area results, decide to include sensitive characteristics in the census, the special character of such variables imply that:

a) Non-response to the question should be an option for respondents without sanctions;
b) "Proxy" information from other persons other than the respondent should be discouraged compared to the normal case;
c) Such variables might be excluded from microdata release to researchers;
d) The threshold below which small aggregates are not published or disseminated should be higher than for other characteristics.

Even where there are no legal rules about such sensitive characteristics, it may be advisable for statistical offices to treat them in a more restrictive way.

15. The threshold for making accessible small aggregates, especially for small geographical areas, is a particularly relevant issue for population censuses. The principle to be respected is again the principle of confidentiality. Aggregates cannot be disclosed if they permit the indirect identification of individual persons or households, thereby providing new information on these units that has not been available to the user(s) before. Unlike indirect identification of economic units (which is very easy in some cases), indirect identification of persons/households normally presupposes additional detailed knowledge about a narrowly defined universe such as a small area. Therefore, this risk is limited to statistics about small areas. Tables from censuses without sensitive variables in the above sense that refer to a larger universe can therefore be disseminated without suppressing so-called small aggregates. Furthermore, tables without cross-classifications, but referring to smaller geographical areas, do not offer the risk of disclosing additional information or the risk of non-statistical use, if in addition sensitive values of certain characteristics are suppressed or grouped together with non-sensitive values. A general rule of thumb that all aggregates beyond a certain threshold have to be suppressed (plus appropriate neighbouring cells to exclude indirect compilation of suppressed small aggregates by deducting non-suppressed cells from marginal totals) has the advantage of being simple, but it will restrict considerably and, from the point of view of the confidentiality principle unnecessarily, the amount of small area information which can be released.

16. In this context, the link to the principle of equality of access to statistical results (as opposed to microdata) has to be stressed. If the release of certain small aggregates is excluded, this exclusion is valid for all users, including government users. On the other hand, if certain small aggregates are produced by special compilation at the request of a specific user, and found to be in line with the dissemination criteria, they have to be released to other users on request as well. Any preferential treatment of government units that are not part of the system of official statistics in this respect would not be in line with

the principle of having equal access to statistical results.

17. The seventh principle, by which all laws, regulations and measures under which the statistical systems operate are to be made public, is also linked to the confidentiality issue. The criteria by which it is determined whether small aggregates can be disseminated or not are such measures, and therefore they have to be made public, irrespective of their legal status. In addition, the statistical office has to define a process by which borderline cases are to be decided.

D. Public Use Files

18. An increasing number of countries have started to disseminate census microdata in the form of public use files that can be obtained by everybody. In addition to any exclusion of names or codes referring to names of persons, such files have also to make indirect identification almost impossible, by omitting totally, or including only at level above the finest detail, some characteristics that are particularly "helpful" for indirect identification through allowing comparisons or even matching with other sources at individual level. Such variables are the date of birth (or other dates) and the geographical location of residence (or place of work).

19. The methods to ensure that indirect identification is excluded for all practical purposes vary from country to country. Any release of a public use file should be preceded by a number of thorough tests for indirect identification. In very small countries, public use files may not be a viable option at all, even if all geographical information is excluded. In any case, it cannot be expected that the release of public use files will make redundant all request by researchers for microdata, since most of the interest in census data is linked to precise geographical information, which cannot be made publicly available as public use file with a full set of characteristics about persons and households.

20. As a method to diminish the risk of indirect identification, the deliberate introduction of "noise" through perturbations of some variables, while keeping intact the main aggregates, has been proposed. It is doubtful whether this method is an appropriate one in the context of census microdata and their use for small area statistics. Perturbations can lead to statistical artefacts at this small geographical level, which may discredit the overall accuracy of the census results in general.

21. The risk of indirect identification diminishes somewhat with time, given that the probability that the situation of a defined individual has changed since the census took place increases with time. A possible strategy for countries where complete elimination of indirect identification is almost impossible is therefore to release public use files only after some time since the census has elapsed. Before this is deemed possible, all requests would have to be examined individually against a set of requirements as described in paragraph 11.

E. Transparency of procedures

22. All decisions about the disclosure policies in the context of confidential or potentially confidential census data do involve contradicting objectives. On the one side, the guarantee given to respondents on confidentiality must be strictly adhered to. On the other side, data in official statistics has to be made available for all possible statistical uses by many different users. It is therefore in the interest of the statistical office to fix the standards of implementing the legal requirements with too much compromising on the possible

statistical use of the data in a participatory process that includes various types of users, but also representatives of respondents and, most importantly, authorities in charge of data protection that may exist at national or sub-national level.

23. Confidentiality being a census issue that is if interest to the public at large and to the media, it is in the interest of statistical offices to inform the public about these measures. Ideally, statistical offices would obtain a certificate, e.g. from the data protection authority, that their practice is fully up to the legal and professional standards. This policy is not only important during the collection phase, but should be continued throughout the whole process, beyond the dissemination of the first results. In this way, the census can become a major element in maintaining and strengthening the trust of the society that data in statistical offices are safe from disclosure and non-statistical use, and this credibility is essential for all on-going and future statistical data collection with households or economic units carried out by the statistical office.

ANNEX

FUNDAMENTAL PRINCIPLES OF OFFICIAL STATISTICS

IN THE REGION OF THE ECONOMIC COMMISSION FOR EUROPE

At its meeting on 15 April 1992, the Commission adopted its decision C(47) on the fundamental principles of official statistics in the region of the Economic Commission for Europe:

C (47) THE FUNDAMENTAL PRINCIPLES OF OFFICIAL STATISTICS IN THE REGION OF THE ECONOMIC COMMISSION FOR EUROPE

The Economic Commission for Europe,

Bearing in mind that official statistical information is an essential basis for development in the economic, demographic, social and environmental fields and for mutual knowledge and trade among the States and peoples of the region,

Bearing in mind that the essential trust of the public in official statistical information depends to a large extent on respect for the fundamental values and principles which are the basis of any democratic society which seeks to understand itself and to respect the rights of its members,

Bearing in mind that the quality of official statistics, and thus the quality of the information available to the Government, the economy and the public depends largely on the cooperation of citizens, enterprises and other respondents in providing appropriate data needed for necessary statistical compilations,

Recalling the general provisions and standards adopted to this end by the European Convention on Human Rights, the Convention of the Council of Europe of 28 January 1991 for the Protection of Individuals with regard to automatic processing of personal data, the Final Act of the Helsinki Conference on Security and Cooperation in Europe, the Final Declaration of the Bonn Conference on Economic Cooperation in Europe and the Charter of Paris for a New Europe,

Recalling the efforts of governmental and non-governmental organizations active in statistics to establish standards and concepts to allow comparisons among countries,

Recalling the efforts of governmental and non-governmental organizations active in statistics to establish standards and concepts to allow comparisons among countries,

Recalling also the International Statistical Institute Declaration of Professional Ethics,

Having taken cognizance of the consensus reached within the Conference of European Statisticians on the need to define the principles

governing the activities of the official statistical agencies in the region and in the member States,

Adopts the present resolution:

1. Official statistics provide an indispensable element in the information system of a democratic society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.

2. To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.

3. To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.

4. The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.

5. Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.

6. Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.

7. The laws, regulations and measures under which the statistical systems operate are to be made public.

8. Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.

9. The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.

10. Bilateral and multilateral cooperation in statistics in statistics contributes to the improvement of systems of official statistics in all countries.

The Conference of European Statisticians, at intervals of not more than three years, will discuss these principles, consider ways to contribute to their application and report to the Commission.