



# Economic and Social Council

Distr.: General  
15 July 2014

Original: English

---

## Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Sixteenth Meeting

Geneva, 23-26 September 2014

Item 3 of the provisional agenda

Census methodology

### **Census methodology - Draft text for the Conference of European Statisticians Recommendations for the 2020 census round**

**Note by the UNECE Task Force on Census methodology**

#### *Summary*

This document presents the draft text on census methodology for the new Conference of European Statisticians (CES) Recommendations for the 2020 Round of Population and Housing Censuses. It was prepared by the UNECE Task Force on Census methodology based on the first proposal discussed at the September 2013 meeting of the UNECE-Eurostat Group of Experts on Population and Housing Censuses, and on further discussion within the Task Force and with the UNECE Steering Group on Population and Housing Censuses.

## **I. Introduction to “Part one: Census methodology, technology and operational aspects”**

1. [NOTE: This paragraph is an introduction to Part One, and will be presented before the chapter on methodology] The chapters in Part One present a broad description as to how countries conduct censuses. While these chapters are not designed to provide recommendations, they do give some guidance. Chapter I provides a broad overview of the methodology suggested for the 2020 round of population and housing censuses in the ECE region. It looks at the aims and objectives of a census as well as methodological issues

associated with the different phases of census. Additional information can be found in the “UN Principles and Recommendations for Population and Housing Censuses, Revision 3”<sup>1</sup> [Note: the new revision is expected to be published before the new CES Recommendations]. Chapter I also covers some aspects of census data confidentiality and security. Chapter II on technology looks at some of the technologies that have not been traditionally used in population censuses, but which are now being seriously investigated by some member countries. This chapter also looks at various aspects related to the outsourcing of different census operations. Chapter III covers some operational aspects of the census, including legislation, communications and publicity, dissemination, and documentation. Chapter IV is dedicated to census quality and contains two major subchapters. The first one is on quality assurance and quality management. In the text on quality management the notion of risk is more visible than in the previous recommendations. The second is on quality evaluation. Major elements include the evaluation of the quality of the results from the perspectives of coverage, sampling error, item non-response, processing error, imputation, and the effects of disclosure control, with a view to learning lessons for the future.

## II. Aims and objectives of a census

### A. Role in national statistical systems

2. The objectives of a census are specific to individual countries and differ according to the local circumstances. Its unique role depends on the demand of statistics existing in a country and by the content and structure of its existing statistical system.

3. The population and housing census represents one of the pillars for the data collection on the number and characteristics of the population of a country. The population and housing census is part of an integrated national statistical system, which may include other censuses (for example an agricultural census or a census of production), surveys, registers and administrative files. It provides at regular intervals the benchmark for the population estimates program of the National Statistical Office. For small geographical areas or sub-populations it may represent the only source of information for certain social, demographic and economic characteristics. For many countries the census also provides a unique source for a solid framework to develop sampling frames.

4. In July 2005 the United Nations Economic and Social Council (ECOSOC) adopted a resolution urging “Member States to carry out a population and housing census and to disseminate census results as an essential source of information for small-area, national, regional and international planning and development; and to provide census results to national stakeholders as well as the United Nations and other appropriate intergovernmental organizations to assist in studies on population, environment, and socio-economic development issues and programmes”. [Note: This refers to the 2005-14 programme of World Censuses, and will thus need to be revised to refer to the corresponding resolution for the 2015-2024 programme. However, it will probably not be until the 2015 Statistical Commission Meeting that a resolution for the 2020 round is adopted.]

5. An increasing number of countries now rely on registers to produce some or all of their population and housing statistics. In these countries there is an opportunity to provide

---

<sup>1</sup> NOTE: Add here updated reference to the new Principles and Recommendations, part II Planning and organization.

an integrated view of the country where social, demographic and economic characteristics are linked together.

## **B. Non-statistical functions of a census (implications and risks)**

6. One of the Fundamental Principles of Official Statistics states that “individual data collected by statistical agencies...are to be used exclusively for statistical purposes”<sup>2</sup>. While the use of census data for administrative purposes would violate this Fundamental Principle some countries use the census operational infrastructure not only to collect statistical information for the census but also to collect information on individuals or households for the creation or updating of population registers. Countries who use the census operations in this way may want to consider:

- a) Clearly explaining to respondents the dual purpose of the census operations and that the census information collected will remain confidential and used only for statistical purposes;
- b) Using two separate forms for the two purposes;
- c) Ensuring there is a separate legislative framework for each of the operations; and
- d) Assigning to a different agency (that is not a National Statistical Office) responsibility for updating information required for administrative purposes.

In some countries, in recognition of their sociological and historical value, census records are made available to the public after a period of closure. In doing so, a number of important issues arise; these are noted below in paragraphs 118-120. [Note: these paragraph numbers could change]

## **III. Definitions, essential features and phases of a census**

### **A. Background**

7. Traditionally, the definition of a census has been based on the basic enumeration features of individual enumeration, simultaneity, universality, and defined periodicity. In the last twenty years however different methods have emerged in the ECE region where the census has assumed a wider concept. In some countries the traditional methods based on the field enumeration of all individuals has moved to the use of data recorded in administrative registers. Furthermore, the priority of universal enumeration of individuals and their characteristics shifted toward the need for more frequent and relevant data for the total population and the smallest local areas. Consequently, a common definition of a population and housing census for the ECE region is now based more on the output produced rather than on the methodology used.

### **B. Definition**

8. The population census is defined as the operation that produces at regular intervals the official counting (or benchmark) of the population in the territory of a country and in its

---

<sup>2</sup> See the text on “The Fundamental Principles of Official Statistics in the context of Population and Housing Censuses” (doc. ECE/CES/GE.41/2014/14)

smallest geographical sub-territories together with information on a selected number of demographic and social characteristics of the total population. This operation includes the process of collecting (through a field enumeration or the use of registered-based information), processing and aggregating individual information, and the evaluation, dissemination, measuring the precision and analysis of demographic, economic and social data. In order to plan for, and implement, economic and social development policies, administrative activity or scientific research, it is necessary to have reliable and detailed data on the size, distribution and composition of the population. The population census is a primary source of these basic benchmark statistics, often covering not only the settled population but homeless persons and nomadic groups as well. Data from population censuses may at times be presented and analysed in terms of statistics for a wide variety of geographical units ranging from the country as a whole to individual small localities or city blocks.

9. The housing census is defined as the operation that produces at regular intervals the official counting (or benchmark) of all housing stock in the territory of a country and in its smallest geographical sub-territories together with information on a selected number of characteristics of housing. This operation includes the similar processes of data collection, processing and the aggregation of information related to housing, and the evaluation, dissemination and analysis of data related to the living quarters. The census must provide information on the stock of housing units together with information on the structural characteristics and facilities that have a bearing upon the development of normal family living conditions.

10. The population and housing census is thus the process that produces at the same time inter-related information on the population and housing stock as described above. This operation has the advantage of obtaining information on two universes (population and housing). The outputs of a census process related to the total population and housing stock are indispensable for providing statistics on the population, family, household and housing situation on a uniform basis for small areas and population sub-groups. The characteristics of the population include geographic, demographic, social, economic, and household and family characteristics. For many countries, the outputs obtained through a census process are vital for providing such information since the census is the only source available and there are no other viable alternatives.

### **C. Essential features of a Population and Housing Census**

11. The set of essential features that makes a population and housing census unique is the following:

(a) Individual enumeration

Information on each enumerated person is obtained so that their characteristics can be separately recorded. This also applies to the housing census. This allows cross-classifying the various characteristics and obtaining data by more than one characteristic.

(b) Simultaneity

Information obtained on individuals and housing in a census should refer to a well-defined and unique reference period (or point in time, e.g. the 1st of January at the beginning of the day). Ideally data on all individuals and living quarters should be collected simultaneously. However, if data are not collected simultaneously, adjustment should be made so that the final data have the same reference period.

(c) Universality

The population and housing census should provide data on the total number of

persons, households and housing within a precisely defined territory of a country. The counting (or benchmarking) of the population should include every person residing and/or present in the defined territory of a country at a defined singular point in time (commonly referred to as the Census day). The data provided by the census of the counting of the basic units should be validated with an independent coverage check.

(d) Small-area data

The census should produce data on the number and characteristics of the population and housing related to the smallest geographic areas of the country, and to small population groups, consistent with the overriding requirement to protect individual confidentiality.

(e) Defined periodicity

The census should be taken at regular intervals so that comparable information is made available in a fixed sequence. It is recommended that census data be produced at least every ten years.

#### **IV. Strategic objectives and criteria for the selection of census topics**

12. Given the costs required and the massive involvement of the population, the content and the methods used in a census should be carefully scrutinized to make sure that all the aspects of collection operations and the dissemination of results comply with the highest standards of relevance, quality, confidentiality, privacy and ethics. Data should be collected using the most viable and/or cost effective means. The content of a census should be decided after looking into: (i) the demand for data at national and local levels, (ii) the availability of data from other statistical sources and (iii) the constraints of a census for data collection where (for traditional censuses at least) only a limited number of questions can be asked on single topics and where sensitive or more complex topics that require extended modules and specialized training of interviewers can be covered only to a limited extent. In addition to these three factors, the countries following the census programme of the European Union will have to take into account possible requirements of that programme<sup>3</sup> in terms of census content, aimed at ensuring international comparability of census results specifically among countries of the European Union.

13. Each census topic should meet a number of key user requirement criteria:

- (a) The topic carries a strong and clearly defined user need;
- (b) Data on the topic are required for small population groups and/or at detailed geographical levels;
- (c) The topic is of major national importance and relevant at the local level;

<sup>3</sup> At the time of preparation of the Conference of European Statisticians Recommendations for the 2020 census round, information is not yet available on the 2021 census programme of the European Union and possible requirements in terms of census content. The topic content of the 2010 round of censuses carried out in the European Union member States had a very strong focus on enabling international comparisons regardless of the methodology adopted. Indeed the content of the censuses in these countries was as much influenced by the EU as by the requirements for information at the national and local area levels so that it would be possible for Eurostat to provide international comparisons. It is likely that, as the EU Framework Census Regulation will also be in place for the 2020 round, the range of topics to be reported to Eurostat will remain unchanged, though the scope and content of the prescribed outputs may be less detailed

(d) Data on the topic are expected to be used in multivariate analyses with other census topics;

(e) The data to be collected has been shown to be reliable and accurate; and

(f) The content does not differ drastically from previous censuses and where appropriate a new or modified topic can still provide comparison with previous censuses.

14. The user requirement for data should be balanced against a number of other factors when evaluating what topic can be collected from the census. A topic should NOT be included in a census if:

(a) It is sensitive or potentially intrusive, or requires lengthy explanations or instructions to collect;

(b) It imposes an excessive burden on the population, or seeks information not readily known;

(c) Its inclusion is likely to have an detrimental impact on coverage;

(d) It enquires about opinions or attitudes; or

(e) It is likely to present major coding problems or extensive processing or significantly add to the overall cost of the census.

15. In addition to these factors, the census should be considered as an exercise carried out purely for statistical purposes, and should not, therefore, be used to collect data that will deliberately promote political or sectarian groups, or sponsor particular causes.

16. In case a paper census questionnaire is used the limited space available on the questionnaire should be optimised; the design and size of a question will also be an important factor in a census questionnaire in deciding whether certain data can be collected.

17. The inclusion of new topics should always be tested to ensure successful collection and production of reliable results. In general, a population and housing census should be seen as part of an integrated programme of data collection and compilation aimed at providing a comprehensive source of statistical information for economic and social development planning, for administrative purposes, for assessing conditions in human settlements, for research and for commercial and other uses. The value of a population and housing census is increased if the results can be employed together with the results of other investigations.

18. A list of recommended topics can be found in Appendix I to the CES Recommendations. The list is divided between core and non-core topics and reflects the recommendations contained in chapters III-XIII of the Recommendations. Core topics are those considered to be of basic interest and value to CES members, and it is recommended that these countries cover these topics in their 2020 round of population and housing censuses. Non-core topics are those topics that countries could select based on their national priorities. Criteria for the selection of these topics are presented in paragraph 13. Some topics are referred to as derived topics. Derived topics are those for which information is obtained indirectly (i.e. from other topics), and therefore are not required to be collected separately. The derived topics are presented in general after the topics from which they are derived.

## **A. Census phases**

19. Censuses conducted by means of a field enumeration do not necessarily follow a uniform pattern among CES member countries, but they have certain major common elements. In general, census operations can be divided into seven phases which are not

entirely separate chronologically or mutually exclusive: involvement of stakeholders, preparatory work (including legislation, testing and outsourcing), enumeration, data processing, evaluation of data prior to its dissemination, dissemination of the results, evaluation of the coverage and analysis of the results. It is important that appropriate quality assurance strategies (see the chapter on Quality and evaluation) be applied in all these phases to make sure that all aspects of data quality (relevance, accuracy, timeliness, accessibility, interpretability, coherence) are taken into consideration, and that each choice made in all census phases is the best trade-off that “fits-the-purpose”.

## **V. The relation between traditional censuses and sample surveys**

20. While population censuses go back at least 6,000 years, as suggested by clay tablets found in ancient Babylon, the history of modern censuses can be traced to the mid-seventeenth century. Sampling, in turn, is a much more recent technique, dating back more than a century.

21. Traditionally, censuses started out as a simple enumeration of people. Over the years they grew in size and scope as requests started to be made for information on other areas of social and economic life in addition to basic demographic characteristics. Consequently, as new issues emerge, there are pressures to ask more census questions. Allowing for too many extra questions may result in exceedingly large census forms. This can cause concern for the quality of all of the information collected. Indeed, the advantages of simultaneous investigation of several topics may be offset to some extent by the additional burden on the respondent and on the enumerator resulting from the increased amount of information that must be collected at one time.

### **A. Use of long and short forms**

22. In order to reduce the burden on the respondent when information is collected on many topics in a traditional census environment, the data collection could involve both a short form (with selected questions) and a long form (with more questions for specific topics). The long form is completed for a sample of dwellings, households or people.

23. The use of long and short forms has made it possible to collect more information while keeping planning, training and field operations relatively simple, and costs in check. However, in view of the ever increasing demand for information, this strategy may lead to new compromises since the number of questions comprising the long form cannot itself keep growing for reasons already explained. Should “simultaneity” become an overriding principle, countries may wish to consider data collection involving both a short form and two or more longer forms (with more questions for one or more specific topics) and with each long form being completed for a separate sample of households or people. However, having more than one form introduces complexities in terms of designing the necessary ‘interlocking’ samples, keeping track of forms in the field, and weighting results to the total population. It also introduces restrictions for multivariate analysis.

### **B. The census as benchmark and frame**

24. The value of either a population or a housing census is increased if the results can be used in connection with the results of other data collections. These could take the form of use of the census data as a basis or benchmark for statistics in the same field, or to furnish the information needed for conducting other statistical investigations. It can, for example, provide a statistical frame for other sample surveys or an agricultural census. As an

example it can be mentioned that in each of Canada's last several Censuses follow-up surveys on disability or on aboriginal populations have been conducted. They are called post-census surveys that are regarded as using the census as the frame. The population census is also important in developing the population estimates needed to calculate vital rates from civil registration data. In addition, these censuses are a major source of data used in official compilations of social indicators, particularly on topics that usually only change slowly over time.

25. The purposes of a continuing coordinated programme of data collection and compilation can best be served, therefore, if the relationship between the population census, the housing census and other statistical investigations is considered when census planning is under way and if provision is made for facilitating the joint use of the census and its results in connection with such investigations.

26. An essential ingredient of sample design is the existence of a complete, accurate and up-to-date sampling frame. A sampling frame is defined essentially as comprising the materials from which a sample is selected. It may be a list of structures, addresses, households, or persons. The census can be used to construct any of these lists. Indeed, most countries use their census for such a frame. The census frame is almost always the departure point for the design of a household sample survey.

27. It is important to recognize that any census - even one that is only one or two years old - will be out of date and may not be suitable as a frame. In such cases, it is essential to update the census frame with current fieldwork or from administrative records before using it as a frame for a household sample survey.

28. Population and household counts for the enumeration areas, taken from the census, are a highly useful ingredient to establish measures of size for the selection of first- or second-stage sampling units, or to help in various stratification schemes. Whenever the census captures socio-economic information, this can be used to complement such stratification schemes.

### **C. Inter-census surveys**

29. Regardless of whether or not information on a wide number of topics was collected simultaneously, the rapidity of current changes in the size and other characteristics of populations and the demand for additional detailed data on social, economic and housing characteristics that are not appropriate for collection in a full-scale census has maintained the need for continuing programmes of inter-census household sample surveys.

30. The population and housing census can provide the frame for scientific sample design in connection with such surveys; at the same time, it provides benchmark data for evaluating the reasonableness of the overall survey results as well as a base against which changes in the characteristics investigated in both inquiries can be measured. To permit comparison of census and survey results, the definitions and classifications employed should be as nearly alike as possible, while remaining consistent with the aims of each investigation.

## **VI. The relation between population and housing census and agricultural census**

31. While the population and housing censuses have a close inter-relationship, their relationship with the agricultural census is less well defined. However, as the result of increasing integration within programmes of data collection, the relationship between the

population and housing census and the agricultural census is in some countries now closer than in the past, and these countries are increasingly looking at new ways to strengthen this relationship.

32. The relation between population and housing census and agricultural census is discussed in the chapter of the recommendations on agriculture, where two non-core topics are presented for consideration by countries that may want to collect in the population census some information on agricultural activity, that could be used for instance for a subsequent agricultural census. [NOTE: The rest of the text on this topic was moved to the chapter on agriculture]

## VII. Methodology approaches in the ECE region

33. There are three basic approaches to conducting a census, based on the method of data collection:

(a) The traditional method of full enumeration (possibly supported by registers as frame or control only) based on field operations at a given moment, either with an exhaustive collection of all characteristics or an exhaustive collection of basic characteristics with a collection of selected characteristics on a sample basis (long form/short form). This approach also includes alternative enumeration methods applied in two large ECE countries: A traditional (short form) enumeration with annual updates of characteristics on a sample basis (United States), and a rolling census where traditional enumeration is spread out over a few consecutive years with a continuous and cumulative survey (France);

(b) A combined census of registers with full field enumeration for selected variables or by sample field data for selected variables; and

(c) The method of using registers and other administrative sources (possibly supported by data from existing sample surveys for selected variables).

34. These approaches are described below. Necessary conditions, advantages and disadvantages, implications for the phases of census taking, and implications for content are addressed for each approach.

35. Registers and other administrative sources are increasingly becoming a viable alternative to the traditional census in so far as they contain the relevant topics, use similar definitions and classifications and cover the entire population. Sample surveys alone cannot provide equivalent data but they can be used in combination or to supplement census information on specific topics. Whatever alternative method is adopted it is important that the aim should be to adhere to the essential features of the census as closely as possible. When choosing and planning to work according to a new method statisticians have to keep these features in mind, as these give the basis for all censuses in all countries.

36. Whichever method of data collection/data provision is to be used should take into account a wide range of issues such as:

- (a) Users' needs;
- (b) Quality of the data;
- (c) Completeness of the count;
- (d) Data protection and security;
- (e) Comparability of the results between countries and over time;
- (f) Burden on the respondents;

- (g) Timeliness of outputs;
- (h) Costs;
- (i) Political and legislative implications; and
- (j) Public understanding and acceptance

37. The results of the UNECE Online Survey on National Practices in the 2010 Census Round, carried out in spring 2013, show that the traditional census approach is still the most common approach in the ECE region. However, since the previous census round, more countries than ever before moved away to a combined or register-based census. Countries with such censuses normally use a register as the backbone for their census. A register can be defined as a systematic collection of unit level data organized in such a way that updating is possible. Administrative registers are administrative information systems used for decision on individuals. [Note: maybe these definitions have to be moved to an earlier paragraph] Registers thus play a more prominent role in census taking than in previous census rounds. It is expected that this trend will continue in the future. Some countries that have previously taken a traditional approach may move towards a combined census, while others may move directly towards a register-based approach. Some countries that conducted a combined approach plan to move to a register-based approach next time. This can either be a fully register-based approach or an approach where some variables are 'recycled' from an existing survey.

38. Even among those countries planning to continue with a fundamentally traditional approach, several reported that they would introduce significant methodological changes that will utilize additional sources of administrative data to develop information to support a conventional enumeration. The administrative sources would then be used as frame or control only.

39. Whatever the methodology applied, most countries in the ECE region will continue to collect information on both each individual person and on housing through the same operation.

## **VIII. Traditional census**

### **A. Description**

40. The traditional census is the total process of collecting, processing, evaluating, disseminating and analysing demographic, economic and social data pertaining, at a specific time, to all persons and the housing stock in a country or in a well-delimited part of a country. It is taken in a given limited period immediately near to a given reference date (census day). Data are generally recorded on census questionnaires, being either in paper or, increasingly, electronic format, or via a secure online service provision. There are two major methods of enumeration: (a) where the information is recorded directly by an enumerator (or interviewer) or (b) where the household complete the questionnaire on their own (self-completion).

41. In the enumerator method information for each individual (in a population census) and for each set of living quarters (in a housing census) is collected and entered in the questionnaire by a census official designated to perform this operation in a specified area during a specified and (usually) short period of time to meet the requirements of universality and simultaneity. In the self-completion method, the major responsibility for entering the information is given to a responsible person in the unit being enumerated

(usually the head of the household or reference adult person), although the questionnaire is often distributed, collected and checked by a census official.

42. Questionnaires can be distributed by post or by enumerators. In a “most traditional” census, addresses are listed by enumerators who also conduct the enumeration or drop off questionnaires. Many traditional censuses now use pre-existing address lists in all or part of their enumeration. It might depend on whether the census address list is based on the official post address list or the statistical office has its own address database. It may also depend on the quality of the list(s); coverage may be weak or even entirely missing in some parts of the country. The postal services should be used to distribute the census forms only when a comprehensive up-to-date and nationally agreed list of addresses is available or can be prepared. In traditional censuses a big problem used to be how to cover all addresses in the country as there might exist many types of address lists in a country made for different purposes. These are reasons why in some cases there is field verification of address lists before the census. However, this is costly, and a cost/benefit analysis of carrying out such an operation is important. In some cases, instead of address checking, using enumerators for catching new addresses during the enumeration might be a better approach than using the post for delivering questionnaires. If an address list is not judged adequate for delivery of questionnaires by post, then enumerators must deliver them. In this situation when an address list is available a decision must be made regarding how exactly to use that list. Alternatives include: ignore it; a pre-enumeration address checking operation; or provide it to enumerators and ask them to update it as they drop off questionnaires. Another consideration might be to outsource the delivery to an official postal service. Questionnaires can be collected by enumerators, or can be sent back by post or can be delivered by the respondents to an agency (statistical office or local government) or on-line in case of internet data collection. Traditional methods can use these delivery and collection methods depending on the circumstances.

43. In some countries, postal distribution of the questionnaire, with or without postal return, is used in conjunction with the self-completion method. This mail-out and mail-back procedure can be used exclusively or combined with on-site checking by a census official.

## **B. Necessary conditions**

44. Both short and long forms may be used within the context of traditional censuses, or there may be an exhaustive collection of all characteristics data. If the former approach is used, the short form contains only those questions intended for universal coverage, while the long form is used to collect information only from a sample of households and population. This form usually contains detailed questions on particular topics in addition to covering complex enquires such as fertility or disability. Both are utilized during the same time frame of the census, with no content data collected outside of that time frame.

45. The traditional census approach to census-taking is the one utilized by most countries. It has a long-standing tradition of use, and is fully described in the United Nations’ Principles and Recommendations for Population and Housing Censuses [NOTE: Add reference here when the new Revision will be available].

## **C. Advantages and disadvantages**

46. The main advantages of the traditional approach are in providing a snap shot of the entire population at a specified period and the availability of data for relatively small areas.

47. Traditional censuses have been singled out as the most elaborate, complex and costly data collection activity that national statistical offices undertake. In addition to the

cost, this complex task requires full awareness and cooperation of the public to participate in it. Because of their complexity and expense, such censuses are usually mounted only once every five or ten years, so that even the most recent census data available are often several years out of date.

48. Each enumeration approach (interview or self-completion) also has its own advantages and limitations. The interview method is the most common practice for collecting data and may result in good quality because of the trained enumerators. This is the only approach that can be used in largely illiterate populations or in other population groups that may be unwilling to complete the census forms themselves, or find it difficult to do so. However, this usually requires a large number of field staff and public acceptance (meaning that people let the enumerators enter their home).

49. On the other hand, in countries where literacy is virtually universal and educational attainment relatively high, the self-completion method may often yield reliable results too, at substantially lower field costs, particularly if a mail-out/mail-back procedure can be used. This method is often preferred by respondents, finding that completing the questionnaire by themselves is more convenient than waiting for the enumerator and offers a greater degree of confidentiality. In addition it can help reduce general costs but the statistical office must design the questionnaire, instructions and related materials so as to be user friendly, encourage response and minimize respondent error and item nonresponse. To achieve these results, a thorough questionnaire testing program is very important.

50. It may sometimes be desirable to rely on one method for enumerating most of the population and to use another method in certain areas or for special groups of the population [NOTE: Add reference here to the relevant text in the new Revision of the Principles and Recommendations]. However, overly complex designs should be avoided.

#### **D. Implications for the various phases of census-taking**

51. The decision regarding the method of enumeration to be employed should be taken at a very early stage in planning on the basis of thorough testing of the various alternatives in terms of their costs, the quality of the data produced and their operational feasibility. Even where a method has been followed traditionally, it is well to periodically reassess its relative advantages in light of current census needs and changing techniques. An early decision is required because the method of enumeration used affects the budget, the organizational structure, the publicity plan, the training programme, the design of the questionnaire and, to some extent, the kind of data that can be collected. Especially when using more types of data collection, testing the public acceptance of the methods used is important as it affects the number of enumerators employed and thus the costs. It is very much a matter whether the self-completion method is used by a small (e.g. 5) or large (e.g. 70) percentage of the population.

52. Timing and length of the enumeration period is of great importance. The main consideration should be to select a period in which the census is likely to be most successful in terms of coverage and quality of the data collected. This may depend on a number of factors. Firstly, it is necessary to avoid those seasons in which it will be difficult to reach inhabited areas, or in which the work will be particularly arduous, because of severe or extreme weather conditions. Secondly, a time should be chosen when most people are staying at their place of usual residence; such a choice will simplify the census operations both in a de jure census (where people are enumerated at their place of usual residence) and in a de facto census (where people are counted where they happen to be present at the time of the census). It can make the results of a de facto enumeration, in particular, more meaningful. The season of peak agricultural activity should be avoided because of the difficulty of contacting persons who work late every day and who may even

stay on their land at night if the land is far from home. Periods of traditional holidays, festivals, pilgrimages and fasting are also unsuitable times for the census enumeration.

53. It is very important that the timing of the census should not overlap with major political events such as state or local election campaigns, since the population may confuse the two events and be less responsive to the enumerator at home. Unfortunately statistical offices have sometimes little or no control of this. It is also very important that the census should be taken within a stable political and socially secure environment in the country. In times of political or military instability the public are less likely to be compliant and the security of enumerators may not be guaranteed. The level of security should allow enumerators to reach all parts of the country safely.

54. Once a particular census has been taken successfully and the census date is found to have been on the whole satisfactory, the next census should be taken at a similar time of the year, unless there are strong reasons for changing this date. A regular census date enhances the comparability of the data and facilitates analysis and also provides administrative discipline, motivating all those involved in the census to make necessary preparations in a timely manner.

55. It is desirable to keep the enumeration period short in order to minimize double counting and omissions, which can occur in spite of a single reference date. On the other hand, the shorter the enumeration period, the greater the number of field staff that have to be employed. This increases the cost and may lower the quality of the data. How these different considerations should be reconciled depends on the size and nature of the country and on the resources at its disposal.

56. In recent censuses, most developing countries have allowed about up to ten days for the training of enumerators, while the enumeration period itself has generally varied from a few days to two or three weeks. Short periods are sometimes feasible in small countries while longer periods may often be necessary in large countries.

## **E. Implications for content**

57. The traditional approach to census-taking creates fewer content limitations than those that might be found with a register-based approach. However, overall content in this approach must result from a careful balance between the statistical requirements of users and the desire to minimise respondent burden.

## **IX. Traditional enumeration with yearly updates of characteristics (United States)**

### **A. Description**

58. This design is a variation on the traditional census design and focuses on counting the population and collecting only the basic mainly demographic data in the census year. A large household survey collects and tabulates detailed demographic, social, economic, and housing data every year (or every non-census year) throughout the decade, replacing the need for a census long form to collect this detailed data from a sample of the population.

59. The survey samples a percentage of addresses each year to approximate a long form sampling rate over a certain period of the census cycle, such as four or five years. To achieve increased reliability and quality of the estimates for small administrative/geographical areas, larger proportions of addresses will need to be sampled.

In the United States, where this approach has been implemented, annual sampling rates for the American Community Survey (ACS) at various geographic levels range from 0.51 per cent to 15 per cent. Over a five-year period, the sampling rates range from 2.55 per cent to 75 per cent.

60. The sample is cumulated over time to produce the lowest levels of geographic detail similar to the long form sample in the traditional census. Again in the United States, five years of data are required for areas with a population of less than 20,000. Three-year estimates are produced for areas with populations of 20,000 or greater. Single year estimates are produced for areas of 65,000 population or greater.

61. The survey data must be weighted to the population totals to produce reliable and useable estimates. Survey data are weighted to reflect the sample design, to adjust for the effects of nonresponse, and to correct for survey undercoverage or overcoverage. This final weighting adjustment helps to ensure that estimates of the characteristics are comparable to the standard, which is the periodic census. Once the final weights are applied, the statistics are generated, including population estimates, proportions, means, medians, and ratios.

## **B. Necessary conditions**

62. Among a number of necessary conditions, this approach requires the agreement of census stakeholders and government policy makers to introduce such a major variation in design. Users of traditional census data products must be willing to accept the transition from once-a-decade products to a new set of annually updated multi-year products. The user desire for more timely data leads to a cost of more complexity (e.g. in terms of one, three and five year estimates). This approach requires annual funding, rather than funding clustered in a one to three-years period each decade. The funding level is thus smoother over the years than it was previously.

63. To have this method a country also needs to have high quality population estimates at lower levels of geography that will be used to control the results of the survey. These population estimates also need to be kept up-to-date from year-to-year. Operationally, this approach requires an address frame for sample selection. It is critical that this frame not only be maintained but kept up-to-date from year-to-year, especially in rural areas.

64. Conducting a traditional enumeration with yearly updates of characteristics requires an on-going high level of professional staff throughout the decade to support the implementation of the survey. In addition, it requires staff to oversee a program of early and comprehensive planning, development and testing designed to continually seek efficiencies in the management and conduct of the short form only component of the census.

## **C. Advantages and disadvantages**

65. The primary impetus for this approach is twofold: to provide more frequent and relevant data on the population than is available from a census that is conducted only once a decade, and to reduce the operational risks associated with the census. Such a program requires a multi-year program of comprehensive planning, development, and testing. In the United States with its legal requirements for complete counts of the population at intervals, the complete count component of the census design is crucial.

66. In a traditional census design, even when detailed census data are released as soon as possible after the census year, data users are required to work with results that are, on the average, seven years old. The production of timely data to support decision-making at all levels of government is a major motivation for this approach. These timely and, therefore,

more relevant data can greatly enhance the value of the information to government officials, policymakers, and businesses that are currently obtained from a once-in-a-decade long form.

67. Removing the responsibility for the collection of detailed data from a sample of the population as part of the census will allow the short-form-only census to focus more directly on meeting the most basic census objectives.

#### **D. Implications for the various phases of census-taking**

68. This design transfers to the on-going survey the responsibility to provide estimates of detailed demographic, socio-economic and housing data throughout the decade. This transfer eliminates the data collection, data processing and tabulation responsibilities for these data from the decennial census enumeration. By removing the need for a long form during the census year (which requires collecting information on many more questions from a sample of households), census planners may be able to focus more on coverage improvement in the census year itself. Innovation, including the use of some technologies, may become possible when the census task is limited to short-form data collection. Eliminating the need for the census to capture, process, and tabulate detailed data will reduce the processing workload and allow the census to develop processing methods specific to the short-form requirements. Tabulation and release of census data will also be dramatically reduced. The focus of the census office thus switches to an on-going smaller (although still substantial) activity instead of a very large spike of activity every 10 years.

69. Many components of the census now must be coordinated across the operational period surrounding the census and the survey (throughout the decade). This includes consultation, outreach, promotion and publicity, and partnership programs designed to increase stakeholder and public cooperation and awareness. It also includes maintaining a master file of addresses that must be updated regularly, rather than established for a once-in-a-decade endeavour.

70. The fact that the survey is on-going throughout the decade provides an opportunity to develop a strong foundation to support data collection during the year of the census. Information obtained from the survey itself (for example, language spoken) can be of great use in planning for data collection in the census year. The survey-taking experience can be used to better allocate resources during the census.

#### **E. Implications for content**

71. Just like the census long form, the on-going survey can provide data on a wide variety of subjects including: families, children, and the elderly; income and poverty; educational attainment and school enrolment; work and unemployment; disability; immigration and language ability; housing; and many more. In the most obvious approach, the content for the survey is defined to be the content of the census long form. Requirements for adding or revising content must be clearly defined. A survey that relies on multiple years of sample data to support the production of estimates cannot easily accommodate content changes.

## **X. Rolling census (France)**

### **A. Description**

72. A rolling census represents a further alternative approach to the traditional model of census taking by means of a cumulative continuous survey, covering the whole country over a period of time rather than relating to a specific day (census day). There are two main parameters in a rolling census:

(a) The length of the period of time which is linked to the frequency of update required; and

(b) The sample rate which depends on the budget and the geographical levels required for dissemination (country, regions, towns, local areas, etc.)

73. For example, it is possible to build a sample framework in order to produce national results with a single annual survey; regional results by cumulating a few consecutive annual surveys; and small area results by cumulating a more substantial number of years' data. An annual survey may be conducted over the course of a year, or in a particular month or shorter time frame.

### **B. Necessary conditions**

74. Necessary conditions depend on the complexity of the sample framework. If the sampling units are addresses, a master address file is a necessary prerequisite. But if the sampling unit is larger, for example at the municipality level, it is only necessary to have enough information to spread the municipalities over the different years as each will be representative. However, it will be necessary to explain to the users of census data what will be the consequences and how they should use these data, because people are more used to snapshot data rather than periodic cumulative data.

### **C. Advantages and disadvantages**

75. The main advantage of the rolling census approach is the greater frequency of updating data: a traditional census provides decennial, or sometimes quinquennial, benchmarks, whereas the rolling census can provide annual updates. Furthermore it offers the possibility of improving the census process year after year, to introduce new topics as and when they become relevant, and to adopt new technologies as they emerge. The rolling census allows a continuous quality measurement and professionalization of fields actors.

76. A big disadvantage is that the essential feature of simultaneity is lost in that it no longer provides a snapshot of the whole population, at any one time, complicating comparisons between areas due to different enumeration periods, although a range of mathematical techniques (for example, averaging and/or projections and/or interpolation) may be employed so that the data are a statistical depiction of the average situation at a specific period of time. Also, as the rolling census covers the whole country over a long period of time, some respondents move and may be surveyed several times while some other people will not be surveyed at all, resulting in the potential loss of another of the essential features of the census, universality, unless careful methodological adjustments are made. For these reasons, the rolling Census needs an important methodological and data processing workforce.

## **D. Implications for the various phases of census-taking**

77. It is better to begin a rolling census just after a full traditional census, in order to reap the benefit of recent information in which to build the sample framework. As the operation is annual, the process must be very carefully prepared, since any delay can be problematic for the following stages.

## **E. Implications for content**

78. A rolling census is able to include all usual census topics and (as noted above) there is the possibility of changing the questions more regularly than in a decennial cycle. This enables the census to be more reactive to changes in social conditions and the needs of users, even if comparability over time must be preserved. However, for regional information data of consecutive years have to be merged and that implies the need for some stability of the questionnaire. According to the organisation of the census, it may be possible to add some thematic surveys.

# **XI. Combined census with registers and full field enumeration**

## **A. Description**

79. While many countries utilizing the traditional approach start with administrative lists, such as address lists, to support to the enumeration (but do not use administrative data to derive directly census results), the essence of this type of census approach is to make use of more complete population and other registers relevant to a census to directly produce census results (to reduce costs and to lessen the response burden), but complemented with a sample or exhaustive field data collection, with a twofold aim: to improve the accuracy of population counts, on one hand, and to obtain the census variables not available from the combination of registers, on the other hand. Countries using this approach must be cognizant of the fact that, as discussed in paragraph 6, if the data collection results are used to update population or other registers, this may violate one of the Fundamental Principles of Official Statistics stating that “individual data collected by statistical agencies...are to be used exclusively for statistical purposes.” Legislative requirements to support this approach are discussed in paragraph 82 below. [Note: changes to these paragraph numbers may be necessary]

80. There are two major differences from the most similar (combined) census type (combination of administrative registers with sample surveys) that is discussed in paragraphs 92-97:

(a) Variables not available from administrative registers are not obtained by sample surveys but through an exhaustive field operation, as in a classical census; and

(b) The population count based on the population register is not immediately accepted as the best possible, but checked and corrected against reality through the complete enumeration. The census thus acts as an exhaustive evaluation of the coverage of the population register, and allows for the reduction of both undercoverage and overcoverage.

## **B. Necessary conditions**

81. The main technical and legal conditions for this type of census to be suitable are:

(a) Availability of a population register: the register need not be completely reliable for demographic purposes, but it must be reliable enough to serve as an initial means for determining how many people will be counted and where; and

(b) Other administrative registers usable for census purposes: examples include tax files, social security files, public registers of unemployed, educational qualification records, and so on.

82. This type of census, with regard to its relationship with the population register, has two variants, depending on whether it is simply the case that the census is supported by the population register, or whether the benefits are mutual, such that the population register uses the census operation to update and improve its information. In the latter case, two additional conditions are required:

(a) The specific legislation governing the population register must explicitly provide such use of the census operation to update the population register while preserving the statistical confidentiality in the strictly census-related information; and

(b) Technical measures must be applied to ensure that the population register information to be checked, and which will be used for both (administrative and statistical) purposes, is treated in a proper and different way, throughout the whole operation. In questionnaire design this may be achieved by isolating the population variables on specific pages. In the processing phase files containing personal identifications should not contain statistical information.

### **C. Advantages and disadvantages**

83. The population counts in this approach may be more precise than in a wholly traditional census (because of the up-to-date information contained in the population register that helps to minimise coverage error and may also be more accurate than in an exclusively register-based census (because of the checking against reality that a complete enumeration provides, which could correct for accumulated errors in a population register).

84. Information not available in the combination of registers is obtained in an exhaustive, classical way, permitting maximum geographical and conceptual detail.

85. The longitudinal perspective that the use of registers provides is also present in this census approach, for example, by using the population register as the frame.

86. Disadvantages come from the intermediate-point condition of this approach. For example, such combined approaches are much more expensive than exclusively register-based censuses, because of the field enumeration component of the data collection operation. However, they will generally be cheaper than the wholly traditional census since the knowledge of the location where every person is registered permits the use of more efficient collection methods.

87. Response burden, other factors being equal, also falls somewhere between the minimum imposed in those censuses without any specific data collection operation and the maximum for censuses with a full enumeration and no previous information support.

### **D. Implications for the various phases of census-taking**

88. Pre-filling questionnaires with population register information is a complex technical task, especially when associated with the large census volume and with the constraints imposed by available technology. In addition problems of personal data confidentiality would rise as some prefilled data about other household members could

become available for the responding person or in case other persons living in the particular dwelling are visited. However, pre-filling could lower the response burden and fieldwork costs.

89. Previous census information and related administrative data facilitates a much more informed correction and imputation of erroneous and missing responses.

### **E. Implications for content**

90. The combination of registers and a full field enumeration permits maximum flexibility in the topic content, while reducing the response burden in comparison with a wholly traditional census collecting the same information.

91. Compared with combining registers and sample surveys that will be discussed next, the main advantage is the complete geographical and conceptual detail of all the variables, whether available in the registers or not.

## **XII. Combined census with registers and sample field data**

### **A. Description**

92. Some countries create census type statistics by using registers and other administrative sources, together with information from sample field data for selected variables. This option is (like the combined census with full field enumeration described above) open to those countries that do not have all census information available in registers. If this option is chosen, some census tables can be produced by simply counting from register information, while for other census tables information from sample field data have to be weighted to population totals. This is, however, only one way to merge register and survey data to obtain census-type results. Israel, for example, uses a population register as the basis for a 100 per cent enumeration, and a census sample survey to evaluate the accuracy of the register addresses and to collect the traditional long-form census data.

### **B. Necessary conditions**

93. Legislation provides a key foundation for the use of administrative data sources for statistical purposes. National legislation must allow the use of existing administrative data sources for statistical purposes rather than the re-collecting of data whenever it is possible. Such legislation should also give powers to National Statistical Offices to access administrative data at the unit level with identification data and to link them for statistical purposes. Furthermore, the appropriate legislation should provide a detailed definition of data protection. A country can only choose the option of a combined census if all variables not available in the administrative register sources will be dealt with in one or more census sample surveys. Moreover, it is a prerequisite to be able to link the information from the different sources at the record level.

### **C. Advantages and disadvantages**

94. The advantages of a combined census with sample field data are that it is much less costly than any census with an element of full field enumeration and imposes less of a burden on respondents. A traditional census can face many privacy objections to the intrusive collection of personal (and sometimes potentially sensitive) information. This

increases the extent of non-response. There are generally fewer objections, and less risk of non-response, to a combined census in which data are collected only from a sample of the population. It may be the case that data of a combined census will provide highly reliable results, because they are based on information on the whole register population.

95. A disadvantage is that it involves more work to produce the tables from the microdata as weighting problems may arise. Moreover, it may be more difficult to get public awareness as not all people will be involved in the census survey(s). Other disadvantages may be a lack of transparency in that no one external to the process may be able to reproduce the information.

#### **D. Implications for the various phases of census-taking**

96. It would be expected to take longer to start a combined census than a traditional census since it does not make sense to start producing census tables until all sources are available. Nevertheless, once off the ground a combined census with partial enumeration is normally quicker to produce results as it has the advantage that fewer incoming census (survey) forms need to be checked and corrected and less data processed. However, because only sample information is available for some variables it is sometimes impossible to meet the level of statistical and geographical detail required in some tables.

#### **E. Implications for content**

97. Some required variables have to be constructed from different sources and information in registers may be rather different from the result obtained if all the people in the country were interviewed. This implies some extra work as combined censuses are expected to mimic the results that traditional censuses would have given if they were still held. If this work is not done successfully, it may damage the comparability of the results among countries and over time. Registers have, on the other hand, the advantage that complete information on the register population is available. It is crucial that statistical bureaus make use of registers where they are relevant for the census.

### **XIII. Register-based census**

#### **A. Description**

98. The development of a register-based population census system is a long process, which might take many years or even decades. Many countries will choose to continue to use a traditional census in some way even when they start moving towards a register-based approach. The first data items taken from registers can be addresses, basic demographic data items, civil registration information, and income data. Usually the share of administrative data increases stepwise census by census. It is vital that countries have introduced a comprehensive and high quality population register and common identification numbers before they can attempt to combine data from different administrative sources. The continuous updating of registers together with communication between the register systems must be effective.

99. The register-based population census system is built around a set of basic registers that contain comprehensive data on the units that are to be described in the population and housing census. These registers may include the data maintained in a population register and a register of buildings and dwellings, as well as data from a business register. Such registers cover all people resident in the country (although the criteria for determining

'residency' in a population register may vary from country to country), the buildings and dwellings in the country and all the business companies (including all the institutions in the public sector) and their establishments. All statistical units can be linked to one another by means of the identification systems: persons can be linked to household-dwelling units and to the dwellings and buildings in which they live, while employed persons can be linked to their employers. Employers and buildings also need to be linked in order to determine workplace. Similarly, all units can be located on the maps by using local area codes or map co-ordinates.

100. Population census data are produced using the method of register estimation, in which several register sources are used simultaneously to define for each statistical unit the value of the relevant variable. The decision rules should be defined in such a way that the data they produce come as close as possible to the data collected by means of traditional census questionnaires. Data from earlier population censuses and register data from the same point of time are also consulted in constructing these rules. These include rules on prioritisation between different sources in the event of contradictory data.

## **B. Necessary conditions**

101. Legislation provides a key foundation for the use of administrative data sources for statistical purposes. National legislation must allow the use of existing administrative data sources for statistical purposes rather than the re-collecting of data whenever it is possible. Such legislation should also give powers to National Statistical Offices to access administrative data at the unit level with identification data and to link them for statistical purposes. Furthermore, the appropriate legislation should provide a detailed definition of data protection.

102. It is also important that the general public appreciates and understands the benefits of using register sources for statistical purposes and that there is broad public acceptance of the use of these administrative data for purposes of statistics production. Open discussion and debate, explaining the rationale and benefits of the use of personal information held in registers must always be considered a key principle. It is also important that the national register legislation is up-to-date and the activities of register authorities are open and transparent.

103. One major factor that facilitates the statistical use of administrative data records is the application of unified identification systems across different sources. The data linkage must occur at the individual level. In the absence of such unified systems it is extremely difficult and laborious, if not impossible, to link different registers, which is absolutely central to register-based statistics production.

104. Definitions of data items should be the same in the administrative sources and the census, or they should be transformable to meet the census definitions. It is essential to harmonize the concepts and definitions when linking registers, and forming the linkages will be difficult when no universal personal identifier exists. Quality assessments should continue to be conducted. If these conditions are not met, the country should continue to rely on the traditional population census as the primary source of benchmark population statistics.

## **C. Advantages and disadvantages**

105. Reduced costs and increased frequency of outputs are without question the biggest advantages of using administrative register sources over the more traditional census approach. With the introduction of a register-based system initial costs may be high, but

field costs can be reduced to zero and particular census statistics (such as employment, buildings and dwellings and housing conditions) can be compiled on an annual basis. A further key advantage of administrative sources is that the need for processing is confined to those data items that have changed. In the long run it is much cheaper as information is collected just once and processed only if and when it changes, such as for example changes of address. Nationality, religion and marital status, completed education and degrees change quite seldom. In most dwellings the floor area and number of rooms rarely change.

106. Register statistics are obtained from all geographical areas, since registers aim to cover the target population in its entirety, and because detailed geographical information can be obtained for all geographic units, municipalities, freely defined sub-areas and map grids of different sizes.

107. Register-based statistics are generally available every year. Growing information needs create new pressures to step up the production of regional statistics, but regional data from a decennial population census may not be up to date enough to satisfy those needs. Again, this is a major asset of using register sources, allowing for more frequent statistics production.

108. The use of administrative data sources, however, involves certain drawbacks that need to be taken into account. One such drawback is the fact that register-based descriptions have to rely exclusively on the information contents that can be formed on the basis of the registers available. This imposes some restrictions with respect to the variables that are available for analysis and may also undermine international comparability. However, exceptions exist. In the 2010 census round three register-based countries (Belgium, Iceland and the Netherlands) decided each to add one or two variables from the Labour Force Survey (LFS) that were missing in their registers. The LFS microdata were then weighted to the population totals. In this way these three countries can compare all core variables with other countries, but of course the level of detail in analyses with LFS variables is limited by the sampling fraction of the LFS.

109. The use of registers also imposes on the statistical agency a dependency on the authorities responsible for holding and maintaining the registers as well as on any changes in legislation and administrative practices. It is therefore crucially important to have close collaboration with the relevant authorities so that information on any such changes reaches the agency as soon as possible. Ideally, where changes to the content and/or structure of those administrative registers being used for the census are being considered, the National Statistical Office should be consulted in advance.

#### **D. Implications for the various phases of census-taking**

110. Register-based systems can create problems with reference periods and consistency. For reasons of statistical reliability it is important that change events are accurately recorded according to their true date. Information on dates of birth and death is usually accurate because it is recorded on the basis of certificates issued by the authorities, in most cases the reference time point is therefore right. Accurate information is also usually obtained on the dates of employment, unemployment and pension periods, whereas for periods of studying the dates are less accurate. In the event of a change of address the person who is moving may be late to provide notification, or neglect to do so altogether, particularly when moving abroad.

111. The linking of a person's data on such variables as place of work, occupation, and income from different register sources may sometimes give rise to consistency problems. It is, for instance, not always clear that the information on occupation and branch of industry describe the same period of employment.

112. There might be some items in the register system where data linking has caused particular difficulty. The data on employment pension, for example, may not use the same business code as the taxation and business registers, and therefore extra work is needed to link individuals to the company where they are employed. As another example, the linking of business enterprises to the building where they are based is not always straightforward since the company address data may not necessarily be fully accurate, or they may differ from the information in the buildings register.

## **E. Implications for content**

113. There are some data items that may have to be dropped from the register-based population census system because the relevant information is simply not available from any register: these may include household composition, ethnicity, religion, language, mode of transport to work, and mobile work. An option is to include such information from survey samples, but then the census becomes no longer fully register-based. [Note: this is discussed in more detail in other chapters when it comes to specific variables that differ depending on the kind of census conducted; one or two references could be helpful here]

114. Furthermore, without a traditional census questionnaire, there is no longer any collection tool for ad hoc needs. In many countries the population census system is an important tool of data collection that is used to meet emerging information needs to reflect changing social conditions (such as measuring economic migration). This flexibility is lost when data are no longer collected by means of traditional questionnaires.

## **XIV. Confidentiality and security**

### **A. Confidentiality principles**

115. The Census collects information on each person and household in the country. In its uses it is not concerned with facts about individuals as such. Its purpose is to provide statistics about the community, and groups within the community, as a whole. The public, therefore, has a right to expect, and needs to be assured that, personal information provided in confidence will be respected. Names, addresses and PINs should be separated from other data as soon as possible in the census process, and not released, so that the data output contains no personal identifiers. The confidentiality requirement encompasses the whole census operation, ranging from the security of the completed census questionnaires both in the field and during processing to the protection of the information contained in the outputs and made publicly available.

116. Assurances should be given to the public that all the information given will be treated in strict confidence by the census authorities and any person who is employed by, or provides a service to, the census authority for the purposes of carrying out the census. Many countries will have domestic legislation that protects such information in the form either of specific census legislation or of more general legislation relating to data protection and freedom of information.

117. The following additional principles should govern the treatment of the information given in the census returns:

- (a) Only persons under the management of the census authorities, or agents acting on their behalf, should have access to personal census information;
- (b) Completed questionnaires should be collected or returned in such a way that will not reveal information to other members of the public. Additionally, individual

household members should, if they wish, be able to give personal information on a separate questionnaire in a way that will not reveal it to others in their household or establishment, or to the enumerator;

(c) All members of the census organisation and outside agents providing services to the census authority in connection with the census should be given strict instructions, and be required to sign legal undertakings, about confidentiality. They should be liable to prosecution for any breaches of the law;

(d) The physical security of census documents containing personal information held by the census authorities, by field staff or by authorised agents should be strictly enforced and, if felt necessary, independently reviewed;

(e) The computer systems handling census data should have strict safeguards to prevent unauthorised access to the information;

(f) In releasing statistics from the census, all possible steps should be taken to prevent the inadvertent disclosure of information about identifiable individuals and households. Special precautions may apply particularly to statistical output for small areas.

## **B. Statistical Disclosure Control**

118. Statistical Disclosure Control seeks to protect statistical data in such a way that they can be released without giving away confidential information that can be linked to specific individuals or entities<sup>4</sup>. Statistical Disclosure Control methods should also be applied to ensure the confidentiality guarantee. Core qualities of Statistical Disclosure Control are minimum information loss and maximum data utility.

119. Measures to ensure Statistical Disclosure Control of tabular data may include some, or all, of the following procedures:

- i) Swapping the microdata before the tables are produced by selecting a sample of the records, finding a match in the data based on a set of predetermined variables and swapping all or some of the other variables between the matched records;
- ii) Restricting the number of output categories into which a variable may be classified, such as aggregated age groups rather than single years of age (this is called a global recode);
- iii) Where the number of people or households in an area falls below a minimum threshold, suppressing statistical output (a so-called local suppression) – except, perhaps, for basic headcounts – or amalgamating it with that for a sufficiently large enough neighbouring area;
- iv) Adding noise to the tables produced, e.g. by rounding cell values up or down to the nearest multiple of the predefined rounding base (conventional rounding).

120. In case of release of census microdata (such as microdata under contract or public use files) it is important that all information from databases relating to name, address and any unique characteristics that might permit the identification of individual respondents is

---

<sup>4</sup> See Hundepool, A., J. Domingo-Ferrer, L. Franconi, S. Giessing, E. Schulte Nordholt, K. Spicer and P.P. de Wolf, 2012. *Statistical Disclosure Control*. In: *Wiley Series in Survey Methodology*. Wiley, Chichester, United Kingdom, 2012.

removed. In addition applying global recodes and local suppressions to the microdata can be used to diminish the risk of disclosure. Also disturbing the microdata or making use of synthetic microdata may help to protect confidential information.

### **C. Public access to closed census records**

121. Many National Statistical Offices receive requests from time to time from genealogists, social historians and individual members of the public, to allow public access to, or reduce the period of closure for, census records for the purpose of researching family histories.

122. The period of closure of census records in many countries is prescribed specifically by statute but may vary from country to country. Other countries may rely on more general provisions within data protection and/or freedom of information legislation to keep confidential records closed until such a time that minimises the risk of disclosure of information about living individuals. Some countries, however, may choose to destroy the census forms once processing of the data has been completed.

123. While national governments may recognise both the sociological and commercial value of historical census records, they should also recognise that the ability of National Statistical Offices to collect information from the general public may be seriously compromised if assurances given about the confidentiality of the information collected were not honoured. Public confidence in the security and confidentiality of the information given in the census should be regarded, therefore, as paramount.

---