United Nations ECE/CES/2018/4

Economic and Social Council

Distr.: General
26 March 2018

Original: English

Economic Commission for Europe

Conference of European Statisticians
Sixty-sixth plenary session
Geneva, 18–20 June 2018
Item 4 (d) of the provisional agenda
Use of registers and administrative data for population and housing censuses

Guidelines on the use of registers and administrative data for population and housing censuses

Note by the Task Force

Summary

The document presents an extract of the *Guidelines on the use of registers and administrative data for population and housing censuses*. The Guidelines were prepared by a Task Force composed of representatives of the following countries and international organizations: the Netherlands (Chair), Austria, Canada, Estonia, Germany, Ireland, Israel, Italy, New Zealand, Norway, Poland, Portugal, Republic of Korea, Slovenia, United Kingdom, United States, Eurostat, United Nations Food and Agriculture Organization (FAO), United Nations Population Fund and UNECE.

The current extract of the Guidelines is prepared for translation purposes. It includes selected parts of the Guidelines: (i) introduction; (ii) scope of the new UNECE guidelines and definitions of register-based and combined censuses; (iii) essential features of a population and housing census; and (iv) considerations when transitioning from a traditional census to a register-based or combined census.

The full text of the Guidelines has been sent to all members of the Conference of European Statisticians for electronic consultation. It is available at: http://www.unece.org/index.php?id=47411. Subject to a positive outcome of the consultation, the CES plenary session will be invited to endorse the Guidelines.

GE.18-04680(E)
1804680

I. Introduction

A. Background

1. Between 2012 and 2015 the UNECE Steering Group on Population and Housing Censuses coordinated the preparation of the *Conference of European Statisticians (CES) Recommendations for the 2020 Censuses of Population and Housing*. The Steering Group managed the work of nine topic-related Task Forces established to prepare initial drafts of the various chapters of the Recommendations. The CES subsequently adopted the Recommendations in June 2015, and these are available both in electronic format on the UNECE website[1] and in printed form in English, French and Russian.

2. In October 2015, the CES Bureau conducted an in-depth review of the diversification of population census methodologies and sources, based on a paper by Finland and Turkey (ECE/CES/BUR/2015/OCT/3) and a note by UNECE (ECE/CES/BUR/2015/OCT/3Add.1). As an outcome of the review, the Bureau supported the preparation of new guidelines on the use of registers for population and housing censuses, and requested the Secretariat to prepare new terms of reference for the Steering Group on Population and Housing Censuses and for a Task Force on Register-Based and Combined Censuses (Report of the Bureau meeting: ECE/CES/BUR/2015/OCT/21).

3. A draft of the proposed new Guidelines produced by the Task Force was presented and discussed at the meeting of UNECE Experts on Population and Housing Censuses in Geneva in October 2017 and consequently revised in the light of comments made by countries at the meeting. This publication presents the Guidelines as subsequently agreed by the CES.

4. Before the guidelines are presented, the following section summarises the census methods adopted by countries in the UNECE region and their evolution over time.

B. Census methods in the UNECE region and their evolution over time

5. There are many different ways to conduct a population and housing census. For the sake of simplicity, this document summarises only the three main categories of census methods: the ‘traditional’ census, the ‘register-based’ census, and the ‘combined’ census. However, a more detailed discussion of the various census methodologies is given in the CES Recommendations.

6. The traditional census is here intended to mean a census based on the direct count of all individuals and the collection of information on their characteristics through the completion of census questionnaires, either in paper form or electronically. The information is collected in the field across the whole country in a relatively short period of time, normally no more than two weeks. Questionnaires can be completed either directly by the households (with delivery and collection of paper forms undertaken by enumerators, the postal service or other methods, or online in the case of electronic questionnaires), or by the enumerators during an interview of the household.

7. The traditional census has a number of disadvantages. First of all, it is a very complex and expensive operation, mainly due to the need to employ a large temporary workforce for the field data collection (enumerators, supervisors and managers), and to print, distribute, and process a very large number of forms. Moreover, in most countries there are increasing difficulties in enumerating certain population groups, particularly those

[1] http://www.unece.org/publications/2020recomm.html

characterized by high mobility and multiple residences, and an increasing reluctance of the respondents to be enumerated, for various reasons. Finally, the traditional census is normally conducted only every 10 years (because of its cost and complexity) and the results often only become available after a relatively long time after data collection, while many users would like to have timelier and more frequently updated information.

8. Some countries have addressed some of the disadvantages of the traditional census either by using sampling (where most households complete only a short form with basic information, while a sample completes a more detailed long form, thereby reducing the total amount of information collected and processed), or by facilitating an online self-response option, which may result in field cost savings and improved quality but requires very careful planning and implementation. Another approach is to spread the fieldwork over time and adopt sampling, as it is done in the 'rolling census' approach developed in France.

9. A totally different approach from the traditional census is the register-based census that was developed by the Nordic countries in the 1970s. Denmark was the world's first country to conduct a fully register-based population and housing census in 1981. Under this approach there is no direct collection of data from the population, and the traditional enumeration is replaced by the use of administrative data held in various registers (population register, building/address register, social security register, etc.) through a matching process, normally making use of personal identification numbers. This approach permits the production of census data at a much reduced cost and with relatively limited manpower, once a good quality system of statistical registers has been established.

10. Since the 1990s, a number of other countries in Europe have developed innovative methods to conduct the census, combining the use of administrative data with a limited collection of data from a field enumeration of the population for specific variables. Under this approach, called a combined census, the field data collection can cover the whole population or just a sample. Often this approach is adopted in the transition from a traditional to a register-based census.

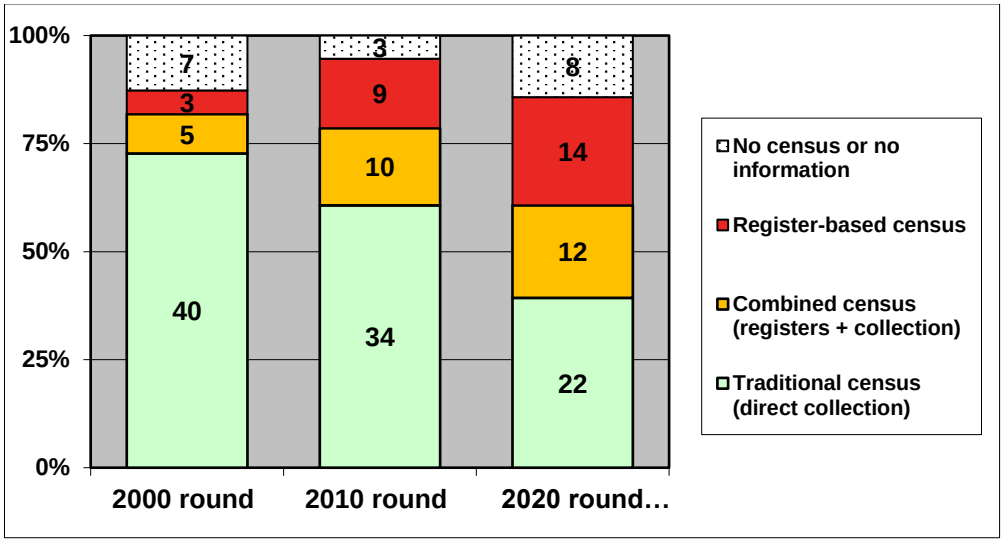
11. In the 2000 census round only few countries in the UNECE region[2] conducted a register-based or combined census (three and five countries respectively) and the traditional census was still by far the most popular approach in the region (40 countries)[3]. However, in the 2010 round, there was a significant increase in the number of countries conducting a register-based census (from three to nine) or a combined census (from five to ten), and a corresponding decrease of the traditional census (from 40 to 34 countries) (see Figure 1).

12. Based on information on tentative plans for the 2020 round, the trend of moving away from the traditional census continues: out of 48 UNECE countries for which information is available, 14 countries plan to conduct a register-based census (29 per cent), 12 countries are planning a combined census (25 per cent) and 22 countries are continuing with their traditional census (46 per cent). If only the 32 member countries of the EU (European Union) and EFTA (European Free Trade Association) are considered, then 13 countries plan a register-based census in the 2020 round (41 per cent), 9 countries a combined census (28 per cent), and just 10 countries will continue a traditional census (31 per cent).

[2] The UNECE region includes countries in Europe, North America, Central Asia, plus Turkey and Israel.

[3] Source: Valente, 2015, From the 2010 to the 2020 census round in the UNECE region – Plans by countries on census methodology and technology. Paper submitted to the Meeting of the UNECE-Eurostat Group of Experts on Population and Housing Censuses, Geneva, 30 September to 2 October 2015; http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2015/mtg1/UNECE_paper_Paolo_draft_0925_rev2.pdf

Figure 1
Number of UNECE countries by census method in the 2000-2020 Census rounds



II. Scope of the new UNECE guidelines and definitions of register-based and combined censuses

13. The scope of these new UNECE Guidelines is not on traditional censuses, but on register-based and combined censuses. Therefore, only definitions of register-based and combined censuses are given. More information about traditional censuses can be found in the Recommendations for the 2020 census round. In these new UNECE Guidelines different kinds of registers (on persons and buildings) are noted with a focus on those used in censuses.

14. For some of the definitions we can refer to those presented in the UNECE publication *Register-based statistics in the Nordic countries*[4]. On paragraph 63 of that publication a ***register*** is defined as a systematic collection of unit-level data organized in such a way that updating is possible. Updating is the processing of identifiable information with the purpose of establishing, bringing up-to-date, correcting, or extending, the register, that is, keeping track of any changes in the data describing the units and their attributes. ***Administrative data*** sources are data holdings that contain information collected primarily for administrative (not research or statistical) purposes. This type of data is collected by government departments and other organizations for the purposes of registration, transaction and record keeping, usually during the delivery of a service. They include administrative registers (with a unique identifier) and possibly other administrative data without a unique identifier. ***Statistical registers*** are registers created for statistical purposes. They are typically created by transforming data from registers and/or other administrative data sources.

15. In some countries the term ‘administrative data’ is used as a synonym for ‘register-based’ data. In other countries a distinction is made between the two, and ‘administrative data’ is taken also to include administrative sources other than registers.

[4] http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

16. A register-based 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 (see para. 123 of *the CES Recommendations for the 2020 census round*). Some register-based census countries miss some of the census variables in all of the available registers and choose to support their census with unit record data (microdata) from an already existing sample survey. All register-based census countries have in common the fact that no specifically designed census questionnaires are used to collect information about the population. Therefore, register-based censuses are in general much cheaper than combined censuses and especially so compared to traditional censuses.

17. In a combined census, statistics are created by using registers and other administrative sources, together with information from either sample field data or full field enumeration for selected variables (see paras. 52 and 116 of the CES Recommendations for the 2020 census round).

18. The remainder of the Guidelines are organized as follows. Chapter III describes the essential features of a census and how these may be met by register-based or combined censuses. Chapter IV describes a number of elements that need to be taken into account when planning a transition from a traditional census to a register-based or combined census. The following chapters are available only in the full English version of this document: Chapter V outlines a common framework that describes the process of conducting these non-traditional censuses. Chapters VI - IX provide more details on the processes and methods associated with each aspect of the framework and the role that quality assurance plays at each stage. Chapter X, together with Annexes A-I, present case studies from a number of countries that have transitioned or plan to transition from a traditional full-enumeration census to a register-based or combined census.

III. Essential features of a population and housing census

19. The essential features of a population and housing census were originally defined by the International Conference of Statisticians as early as 1853 in Brussels, and by incorporating these features, countries have been able to carry out censuses that have been internationally comparable in terms of methodology and quality over time. Nowadays, these five essential features have been redefined and highlighted by the CES[5] with the aim of ensuring the coherence of census data gathered in different countries with different levels of technical development and different cultures. Adopting all these features – regardless of the methodology of data collection – enables NSIs (National Statistical Institutes) to collect population data of internationally comparable quality that allows making decisions and forecasts on the population development.

20. The five essential features of a census are:

(a) Individual enumeration;

(b) Simultaneity;

(c) Universality (within a precisely defined territory of a country);

(d) Small area data;

(e) Defined periodicity.

[5] Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing (United Nations, 2015), see http://www.unece.org/publications/2020recomm.html, paragraphs 23-28.

How register-based and combined population and housing censuses can be designed to satisfy each of these features is discussed in the following sections.

A. Individual enumeration

21. The principle of individual enumeration is a fundamental feature for any census of population. Traditionally, this has been effected by providing questionnaires that ask questions of each individual within a household. In the case of register-based censuses a different approach is adopted where the data are taken from administrative registers. In such circumstances it is important that each census unit has a special, uniquely identified, record in the registers used. Then the registers become a useful source for the census. In the case of a combined census only some of the variables are derived from administrative data sources, and then this same approach is used for those variables.

22. If a single identifier for a particular unit does not exist across a range of registers, it is necessary to create a new statistical identifier (based on a group of identifying variables) to link the variables held in the respective registers, and to check carefully its quality (for errors and uniqueness).

23. Sometimes it is necessary to create the necessary census variable using information from several administrative registers and composing special algorithms for its calculation. This is possible if the units in all these registers are uniquely identified by the same identifier. In this case the variable created in such a way should be uniquely identified as well and saved in a statistical register. In the event that not all units are uniquely identified by the same identifier it may be possible to create a new statistical identifier as explained in the next paragraph.

24. The basic counting units of a population and housing census include not only persons, but also households, families, and dwellings (whether occupied or vacant). All of these units need identification, but there is no need to use five different identification variables. The minimal necessary identifiers are the one for persons (person ID) and the one for dwellings (dwelling ID). These IDs must be linked with each other (a dwelling ID is assigned to each person) and for each occupied dwelling the list of person IDs of people living in it must be given. The dwelling ID makes use of the address code, which may contain also spatial coordinates.

25. Information about households is usually collected on the basis of the housekeeping concept[6] by those countries conducting a traditional census. This definition can be achieved through asking questions on a survey or census, but is more challenging for countries conducting a register-based census. Many such countries instead use the household-dwelling concept, which considers all persons living in the same housing unit to be members of the same household. While adopting this definition has minimal impact on the total number of private households, it can have a larger impact for certain household types, such as one-person households. This bias in the number of private households and in the estimated structure of the household types depends on the traditions of the country and on living conditions. These challenges for register-based countries also extend to construction of families within households using relationship information.

26. In some countries (such as, Slovenia) a household register exists. The existence of such a register eases the organization of a register-based census, especially when household IDs are included in the register. Then there is accurate information available about which person ID belongs to which household ID. A household register might therefore improve the quality of a register-based or combined census significantly. However, the situation of

[6] See paras. 768-769 of http://www.unece.org/publications/2020recomm.html

Slovenia is an exceptional one. Ireland is researching the potential of using a decision tree algorithm to determine relations between people in the same dwelling so that the current housekeeping definition of the household can be continued. As part of the work to understand the impact of transitioning to its so-called Administrative Data Census, the Office for National Statistics (ONS) in the United Kingdom is currently exploring the potential impact on users of changing from a housekeeping concept to a household-dwelling concept.

27. Sometimes it is useful to use identification codes for other units such as enterprises and organizations. If these are linked with person IDs and dwelling IDs they form a helpful tool for deriving other statistics, such as on commuting between place of residence and place of work.

B. Simultaneity

28. The fixed census moment is the condition defining the simultaneity of the census data. Traditionally, to ensure this condition, the enumeration is carried out over a very short time frame, ideally during one day only. Though most modern day enumerations are conducted over a two-three week period, all the data collected should refer to a specified reference period. This essential feature should be respected also in the case of a register-based or combined census.

29. If the administrative registers in use are regularly updated, then it is necessary to fix the census period and to take the data from all registers with reference to this period. Sometimes the registers are updated regularly at some specific date such as the beginning of a year, and then it is possible to use this date as the census period, and the simultaneity of the census is guaranteed.

30. In the case of a combined census it is important that the census reference period mentioned in the questionnaires and the reference period of the information taken from the registers are the same or as close to each other as possible.

31. When several administrative registers are used in the census, it is important that all data taken from them have the same reference period. Usually, census variables derived via special algorithms take some time to calculate; hence those census variables are only ready for publishing sometime after the census period. For some specific variables in combined or register-based censuses different reference periods are defined for the particular administrative purpose of the register. Demographic data can normally be taken from population registers at the beginning of the year. However, labour force data might be more relevant somewhat earlier in the year - as fewer people tend to be in employment around Christmas and New Year’s Eve. For some administrative registers, it may not be possible to have Census Day as a reference day. For example, education registers often have relevant education data (e.g. referenced to a day early in the academic year) that may differ from a chosen Census Day. In such cases, the NSI may wish to take education data with a reference day as close as possible to Census Day as a compromise.

C. Universality (within a precisely defined territory of a country)

32. To ensure the universality of the traditional census the questionnaires used in the enumeration process are the same for all households and individual persons. If there are questionnaires in different languages, it is important to check if their content and the meaning of all questions is exactly the same.

33. If the administrative registers used in the census are common for the whole country and all population groups, the condition may be regarded as being met. However, if there

are different administrative data in different areas or for different population groups (such as an urban population register and a rural population register or if different administrative data are held in different cities), then it is necessary to analyse the possible discrepancies between the different administrative sources and find a way to define common census variables using these different administrative sources. In this case plausible results can be derived from these newly defined variables (via an appropriate algorithm) in a statistical register.

D. Small area data

34. Providing a rich wealth of information for small geographic areas and small population sub-groups (generically referred to here as ‘small area data’) is a key objective for a census of any kind, as there is generally no other single source of comparable data.

35. In the context of register-based or combined censuses, small area census data can be derived from administrative data providing the coverage is high, preferably covering the whole population. If there are some small areas that are poorly covered, resulting in the administrative data lacking some information (and thus showing poor universality), it will be necessary to seek to improve the administrative dataset before it can be used as a source for the census. Such poor coverage is likely to be a problem for the every-day usage of the administrative data, and so it should be improved anyway. Improving the statistical register can sometimes be done by adding information from another source, providing that linkage through common IDs is possible.

36. Sometimes there might be special administrative data for some small areas (particularly for some small groups of people). Then it will be necessary to combine these different administrative sources (see section III.2.). If the result is satisfactory, this combination is useable. In the case of a combined census it is also possible to supplement the lack of information in administrative sources with a survey that, if necessary, may use different data collection methodologies, such as doorstep or telephone interviews, or self-completion with paper or online questionnaires, to suit different areas. However, where a sample survey is used, problems can still arise with the coverage of small area data. In such circumstances, users’ needs regarding the required level of detail of the census outputs should be taken into account before any decision is made on the sample size of the census survey.

E. Defined periodicity

37. Nowadays, censuses are generally organized worldwide on a ten-year cycle. The United Nations recommend that countries conduct at least one census every ten years (between 2015 and 2024 for the 2020 census round). To meet European Union requirements, member countries were required (by an EU Regulation) to conduct a census in 2011, and will similarly have to conduct the next one in 2021. However, some countries (such as Australia, Canada, Ireland, New Zealand, and Slovenia) have shorter periods between censuses. The same underlying ten-yearly cycle should be kept in all censuses, regardless the methodology to provide international comparability. If a five-year period has been used, then, for EU member states, one of the two census years of the country should coincide with the census year fixed by Eurostat.

38. An advantage of a register-based census is, however, the opportunity to conduct the census more often than the usual ten-year cycle, as register data are permanently available and more regularly updated. It is advisable also to prepare census software in such a way that it is permanently ready for using for any reference date. Then the periodicity between censuses can be ten, five, or two years, or censuses could even be conducted annually.

Yearly updates of demographic data are a key objective in many European countries that now adopt different census methodologies.

39. It is also possible to produce some census updates with a shortened list of variables (but long enough to meet users' requirements). This could save resources. From this it follows also that in countries where a regular decennial census will be continued using a combined methodology, some updates with a shortened list of variables can more easily be done more often if the variables on the shortened list can be derived from administrative data sources.

F. Conclusions of the features defined by the Conference of European Statisticians

40. From the above it can be concluded that if a country has a good system of administrative registers that is consistent, easily useable and of high quality (that means, all units are uniquely identified using a common identifier), or if statistical registers can be built with equivalent quality from administrative data sources, then it is possible to organise a population and housing census that satisfies all required CES features.

41. If the list of administrative data sources cannot achieve the required quality for the whole range of census variables, it may still be possible and useful to derive census variables using a combined or register-based methodology with a shortened list of variables (including those covered in administrative registers) collected in between regular censuses.

IV. Considerations when transitioning from a traditional census to a register-based or combined census

42. The decision to move to a combined or register-based census is motivated because of the several advantages to be gained. However, the move needs to be carefully managed, and there are necessary conditions relating to data, technology, legal and stakeholder issues for a successful transition from a traditional to a combined or register-based census. This transition may therefore give rise to some challenges. These are set out in section IV.2 and can present significant obstacles for some countries.

A. Advantages

43. Conducting a combined or register-based census has a number of advantages and opportunities that are described below.

1. Lower per capita costs

44. Traditional censuses are very expensive. In many countries that conduct a traditional census it is common that the census costs are equivalent to about two annual budgets of the NSI. It is understandable, therefore, that governments put pressure on the institutes to cut these expenses, especially when other data sources are available.

45. If a combined census is conducted with full field enumeration for some of the variables, the cost savings made by shortening the census questionnaires are partly lost again in the effort required for combining the information from the field with administrative data sources. While the savings may be modest, the approach may still be preferable, particularly where the transformation process allows more data from administrative sources to be used in future censuses.

46. If a combined census is conducted without full field enumeration for selected variables much larger savings can be achieved. Practice shows that introducing a combined census could lead to reducing the costs by 22 per cent compared to a traditional census[7]. Moreover, if a register-based census is conducted no surveying is required specifically for the census and large savings can be expected. Several countries have proved that on average 98 per cent of the costs of a traditional census can be saved this way[8]. However, one should realise that such census cost savings can only be reached once the necessary access to the appropriate registers has been established.

47. Some countries have switched from a traditional census to a register-based census in one census cycle. If this is attempted, then all the costs of making the change must be met within the decennial period. More usually, however, such a move is done in several stages, often by first adopting an intermediate combined census approach before moving to a register-based census. The cost of the change can then be spread over two or three census cycles.

48. Moving to a combined approach and, especially, to a register-based approach contributes to a more cost-effective census. It is clear that meeting government-imposed budgetary constraints provides an incentive for such moves, even if registers are incomplete or of insufficient quality to be used as sources for the census. In such cases, however, it should be made clear from the beginning that a country should not attempt a move immediately and should continue conducting some form of traditional census. However, even when countries continue to do so, innovations making greater use of administrative data could help the NSI work more efficiently. It helps if the relevant public authorities make administration data sources available to the statistical institutes to produce proto-type register-based statistics. The government itself can help a great deal both by removing any legal barriers to data sharing and by subsidising the transition.

2. Quicker to conduct

49. In countries with an established register-based statistical system, the total production time to conduct a register-based census is much shorter than any other kind of census, due principally to the fact that no field enumeration has to be conducted. This, of course, does not hold for combined censuses in which some fieldwork is still necessary. Without fieldwork, the census may generally require less time in the planning stages (thus having the advantages of saving both time and money) while maintaining - or even improving - the time taken for the delivery of output. However, one has to realise that the first time a register-based census is conducted it may take more time than in later census rounds as census planning has to be set up anew.

3. Fewer problems with non-response and reduced response burden

50. With no need to conduct any field enumeration, fewer problems with non-response and a reduction in response burden to zero can be expected if only, or mostly, administrative data sources are used, and providing these are comprehensive and cover the whole population. With ever-increasing non-response rates being reported in international censuses and surveys this is going to be a more and more important aspect. Even if the move is to a combined census, the response burden on the population will also be lower, in particular if only sample field information is collected.

[7] Calculation based on PPP information in Table 7.2 of http://www.unece.org/fileadmin/DAM/stats/publications/2013/Measuring_population_and_housing_2010.pdf

[8] Calculation based on PPP information in Table 7.2 of http://www.unece.org/fileadmin/DAM/stats/publications/2013/Measuring_population_and_housing_2010.pdf

4. Possibility of a continuous census

51. The more administrative data a country uses in its census, the better the possibilities of an annual or even more frequent census. In theory, with good quality administrative data updated outputs could be produced on a daily basis. Such real-time censuses may be something for the future, but more regular than decennial data are now becoming increasingly expected by users. As information from other sources can provide such real-time information, censuses will be expected to keep pace. As a by-product, it is easier for the NSI to keep the knowledge and IT-infrastructure up-to-date if annual census updates are produced.

5. Better cooperation between units within the NSI

52. In some NSIs the different divisions or directorates are often structured to perform in isolated silos without too many contacts with each other. If an NSI moves to a register-based statistical system there is potential for this silo structure to be abandoned. By moving from a survey-based statistical organisation towards a register-based system, the traditional one-to-one relationship between sources and statistics is replaced by an m to n inter-relationship across all statistical branches. A better cooperation of units within the NSI thus becomes essential. Moreover, by better integrating statistics, the coherence of the statistical framework within the NSI is improved.

6. More time and resources for innovations

53. Innovations are crucial for the long-term perspectives of an NSI. As introducing statistical registers in the process of producing official statistics saves both time and money, it becomes easier to innovate. The resources saved could, instead, be used to stimulate administrative and technological innovations so that the processes of data production remain up-to-date.

7. More flexible and responsive to new information requirements

54. If all data are stored appropriately, not only can the regular statistics be produced more frequently, but there is potential for the creation of new statistics to meet changing user needs. The NSI can then become more flexible and responsive to new information needs and increase its value to society. Although this may not be an aim in itself, it can lead to a greater level of user satisfaction.

B. Necessary conditions for a successful transition to a register-based or combined census

55. If a country wants to move to a combined or register-based census a number of conditions are necessary before information from administrative registers (and other sources) can successfully be integrated to create the underlying statistical register. A number of these necessary conditions can present some challenges to the NSI, and these are discussed in the following paragraphs.

1. Legal base

56. Whatever type of census an NSI conducts, it should be within an established legal framework. To be able to conduct a combined or register-based census, in particular, there must be legal provisions that prescribe the access to, and protection of, the administrative data. Such a legal base is normally enshrined in a Statistics or Census Act.

57. The NSI must have legal authority to access the relevant administrative data, ideally free of charge, from whatever public authority sources, preferably including personal

identifiers. To avoid legal uncertainty or dispute, it should be stipulated that the right of access applies except in legal cases pertaining to the protection of public order or the security of the country. A further key issue that should be addressed in legislation is the legal provision for the NSI to have some influence or authority in the creation, revision or deletion of those administrative data that are to be used in the statistical registers.

58. In turn, the NSI must have a legal obligation to protect the confidentiality of the administrative data it obtains, and to adhere to the 'one-way traffic' only principle, except under specific circumstances mentioned in the legislation. Indeed, the relevant legislation could do more by generally prohibiting other data controllers having access to data held on the NSI's statistical registers.

59. In certain countries, legal requirements may constrain how a census can be conducted[9]. In some countries, the NSI first started exploring administrative data sources and, thereafter, found a legal base to make register-based statistics possible. In other countries, the legal base was first established and, thereafter, register-based statistics were produced and published. To gain experience in moving progressively to a register-based census it is often simpler for NSIs to start with producing register-based statistics covering just a selection of those variables collected in a traditional census, though it should be noted that the legal obstacles to overcome may be no fewer.

60. It is always the case that moving to a census methodology where administrative data sources play a role needs careful preparation including, in particular, pilot studies. NSIs should realise that once (part of) the fieldwork operation for the census is abandoned, reinitiating it becomes rather difficult. After the passage of time, the knowledge of how to conduct a traditional census is lost, and especially so in the case of a register-based census where no fieldwork is undertaken at all.

61. Legislation on access to administrative data may need to be supported by policies and directives that are internal to the NSI and that translate legislative requirements and central government policies and directives into requirements and responsibilities for the managers and employees of the NSI.

2. Public approval

62. While the law might give a legal licence for the NSI to a combined or register-based census, public approval is also necessary to ensure that such a census is acceptable. This might be more difficult to achieve than establishing the legal base. While in some countries people may get the impression that 'big brother is watching you' in the course of a traditional census, in some other countries using and linking administrative data collected for non-statistical purposes may be seen to be even more intrusive as the public has no control at all over the information about them that is to be disclosed.

63. In a traditional census, privacy concerns may lead to lower response rates or the deliberate giving of wrong information. It is getting more and more difficult to correct for such unit and item non-response. So, on the one hand the public might prefer a situation where fewer questions are asked if the equivalent information is already available. On the other hand, part of the population may prefer to answer census questionnaires directly instead of having their information taken from, and combined with, several administrative sources.

64. In a register-based or combined census people may feel uneasy about, or even object to, information from different administrative data sources is reused and linked in a census. It may not be clear to them that in the census the information is only used for statistical purposes. If no more census forms have to be filled in and only registers are used for the

[9] Particularly where representation in the national legislature depends on census results.

census, the public will generally be less aware that a census is being conducted. However, the absence of any public reaction should not be misinterpreted as public approval.

65. It is desirable, therefore, to prepare for possible specific questions on privacy, confidentiality and security issues in conducting a register-based or combined census. In combined censuses a discussion can be expected about which variables are to be included on the census questionnaires and which variables are to be derived from administrative data sources.

3. Stakeholder approval

66. Stakeholders - or, more specifically, data users - typically want each census to provide at least the same level of detail of information as in the previous census. However, this is not always possible when the census methodology changes.

67. It is important to inform and consult stakeholders beforehand. Users in particular can become critical if their expectations are not met. Even well-informed users can become very critical if they believe that they are going to lose some of the information that they had access to in the previous census. However, it is usually not possible to satisfy all users, and disappointment among some of them is often unavoidable when adopting a new census methodology.

68. It is important, therefore, to have a communication strategy for stakeholder engagement that should encompass some, or all, of the following goals:

- Create a transparent environment concerning the plans of the NSI;
- Assure users that their requirements will be taken into consideration;
- Inform stakeholders of the benefits of using administrative data and demonstrate that the information will continue to be kept secure;
- Strengthen partnerships with the stakeholders so that the NSI can benefit from outside expertise;
- Make stakeholders part of a successful transition to a new census approach.

69. Openness and the clear identification of new opportunities and benefits for the stakeholders will help to gain their approval. It is particularly important that there should be adequate consultation on any change in the provision of those statistics that have financial consequences for stakeholders (such as transfers of money to municipalities).

4. Cooperation between the NSI and other authorities

70. Good cooperation between the NSI and other (mainly government) authorities is vital in using administrative data sources in the census. The NSI needs to know when microdata (the administrative unit records) and the accompanying metadata can be made available before any register-based statistics can be produced. In a combined census, and even more so in a register-based census, the NSI is heavily dependent on administrative data holders to comply with their agreed or legal obligation to provide good quality data on time. If data holders fail to deliver, it is usually the NSI that is held responsible for the failure to publish census statistics on time.

71. It is vital to inform administrative data holders how important their data are for the NSI and how their data are to be used. Additional to a legal base (see IV.2.1.) and good contacts with other authorities, signing cooperation contracts or service level agreements could help in supporting the census process. In theory, administrative data sources from non-governmental authorities could also be used in the census, but this often creates

privacy and data quality concerns and involves commercial considerations; private sector data are more often than not only acquired at a substantial cost to the NSI.

5. Comprehensive and reliable statistical register system

72. A statistical register system that is comprehensive and reliable (in that it contains accurate and timely data) is essential for conducting a combined or register-based census. Administrative data sources, including administrative registers such as a population register, are not normally set up for statistical purposes such as conducting a census. A transformation process is therefore necessary in order to create a reliable statistical register system.

73. To assure the use of register-based statistics in official statistics it is important to have good working relations with administrative data holders. Conditional on the provisions of the legal base, and if the administrative bodies are conducive to it, in some countries there is also a potential to improve the relations between the data holders and the NSI by introducing new, or extend existing, register-based statistics, such as longitudinal studies to evaluate policy implementation. Of course, there has to be contact between the NSI and the relevant administrative data holders whenever an administrative data source is introduced as a new source. However, permanent contacts, facilitated for example via account managers, are vital to keep both the administrative data holders aware of the important role that their data play and the NSI informed about any changes to the microdata and metadata they receive. Only with regular contact between the NSI and the administrative data holders can the success of register-based statistics be maintained.

74. In many register-based census countries the system of administrative data sources is used by many different public authorities. The more users this system has, the better the quality one can expect. In using such a system for the census it is the quality of the resulting statistical register rather than the quality of the underlying administrative data sources that counts: are the data of good enough quality on which to base reliable census outputs?

6. Unified identification system

75. A unified identification system across different administrative data sources greatly facilitates register-based censuses. It is preferable to have unique ID-numbers at the unit record level that are common across all registers. For countries where unique ID numbers for persons do not exist, the ability to link data efficiently and accurately is a particular challenge.

7. Knowledge of administrative sources

76. When a country wants to move from a traditional census towards a combined or register-based census, building up a wide-ranging knowledge of the data held in administrative sources is important before actually making the move. Although building up knowledge can be effected in stages well before the census planning, the effort needed to make this process successful should not be underestimated. Many lessons about failures and successes can be learnt from countries conducting combined or register-based censuses, but the national context should always be taken into consideration. It is never advisable for an NSI to simply adopt the methodology of another country when setting up a combined or register-based census. However, by learning from the experiences of others the transformation period can be shortened drastically.

8. Transparency

77. If there are planned moves to a different census methodology, it is good practice to be transparent and share with stakeholders information on plans and tests as much as possible. As discussed in section IV.2.3 above, it is particularly important to inform users of any decision to move towards a register-based census as such an important change in methodology is likely to have an impact on the content and availability of output. Transparency and openness facilitate external review and feed-back on the new processes.

C. Difficulties that may arise

78. Despite the advantages noted in section IV.1 above, conducting a combined or register-based census has a number of disadvantages and risks that are described below.

1. Dependency on public authorities

79. In moving to a combined or register-based census, the NSI becomes heavily dependent on the public authorities holding the administrative records being used. NSIs have to realise that, for such authorities, the production of statistics is not a core activity to which they would normally give priority. For the NSI, any failure or shortcomings in the administrative registers will affect the quality of the derived official statistics, for which it must take responsibility.

2. Differences in concepts and definitions

80. Registers and other administrative data sources often adopt different concepts and definitions of population-related variables than those that generally apply in traditional censuses. NSIs should be aware that such differences may exist and decide whether these differences are acceptable when moving from a traditional to a combined or register-based census. What may, in one country, be considered an acceptable difference when assessing the balance between the continuity and coherence of the resulting statistics and the reduction in field costs, may be considered unacceptable to users elsewhere. NSIs should weigh up the balance before deciding whether they are willing to pay this price when moving towards a register-based census or a combined census without full field enumeration for selected variables. Sometimes, original definitions and concepts can be approximated rather accurately by derivations from different sources or by editing information from newly acquired census sources. However, this is not always the case and the NSI should then weigh up the balance between the acceptability of the differences and the costs of continuing full field enumeration for selected variables.

3. Timeliness of administrative registers

81. Public authorities responsible for maintaining administrative registers do not hold the data for statistical purposes, and, as a result, will have other priorities that could cause delays in the delivery of the relevant administrative data and metadata to the NSI. This can cause issues for the NSI regarding the timeliness of their register-based statistics, particularly where the timeliness of the delivery of data from different sources varies considerably.

4. Different reference periods

82. A particular problem that NSIs encounter when moving towards a combined or register-based census is that different sources of administrative data often have different reference dates. Sometimes a source gives the option of distinguishing clearly between reference dates and dates of events, but this good practice does not always apply. If these

problems cannot be resolved sufficiently, the risk is that not all sources will be harmonised to the same reference date. Then the question 'What is an acceptable difference in reference dates?' arises. However, the answer to this is dependent on the variable concerned. Some variables are rather stable over time and then a small difference in reference date is normally not a problem. Large differences in reference dates are always unwanted. Finally, it is relevant to realise that also in the case of census questionnaires not all information may in practice refer to the single reference date of the census. Especially in case the census information is given on a moment further away from the reference date recall effects may play a role and respondents do not always give the answers specific to the census reference date.

5. Privacy and security concerns

83. Using administrative data for purposes other than those for which the information was originally obtained inevitably leads to privacy and security concerns. These concerns often relate to the linkage of personal data from different sources. In some countries the legal framework has been specifically adapted to provide for this, suggesting that there is public approval (or at least acceptance) in the use of administrative data in official statistics. For other countries such a consensus has not yet been achieved.

6. Difficulty in identifying sub-populations

84. In a census it is important to ensure universality. However, the range and detail of outputs in a register-based census will be limited to those variables that can be derived from existing sources. These may not all relate to the entire population. Moreover, even for those countries that use sample surveys to collect data on information not available in administrative sources, it is sometimes difficult, or even impossible, to produce accurate outputs for small areas or specific sub-groups of the population because of the size of the sample population.

7. Keeping knowledge and IT infrastructure up-to-date

85. In countries that conduct census projects with large gaps between them, it may be difficult to retain staff within the NSI with the necessary experience and expertise to keep the knowledge and IT infrastructure up-to-date during the inter-censal period. However, when yearly census updates are introduced this difficulty is minimised.

8. Diminishing interest

86. In countries where censuses are carried out using questionnaires, not only the users but the general public itself will be interested in knowing the results. However, in register-based census countries, where people no longer complete census forms, there is often a decline in public interest in census results. Many people will not even be aware that a census has been taken and, as a consequence, national interest in the census is greatly reduced. Users will still have an interest in the statistical outputs, though evidence from register-based census countries suggests that their interest in the choice of original sources and the methodology used to produce the census data diminishes over time. This is also due to the fact that in those countries other outputs are often much quicker available than census results.

V. Chapters available in the full version of the Guidelines

87. The following chapters are available only in the full English version of this document: Chapter V outlines a common framework that describes the process of conducting these non-traditional censuses. Chapters VI - IX provide more details on the

processes and methods associated with each aspect of the framework and the role that quality assurance plays at each stage. Chapter X, together with Annexes A-I, present case studies from a number of countries that have transitioned or plan to transition from a traditional full-enumeration census to a register-based or combined census

Annex.

Glossary of terms and definitions

Accessibility: A measure of data quality relating to the conditions and modalities by which users can obtain, use and interpret the data.

Accuracy: A measure of data quality relating to the closeness of estimates to the unknown 'true' values.

Activity register: A register that holds information about residents' different activities that indicate a presence in the country or area. Such activities can include information on, for example, employment or other economic status, receipt of benefits or pensions, or student status.

Administrative data: Data holdings that contain information collected primarily for administrative (not research or statistical) purposes. This type of data is collected by government departments and other organizations for the purposes of registration, transaction and record keeping, usually during the delivery of a service.

Anonymization: The process of protecting the confidentiality of personal information by removing all unique identifiers from the unit records.

Coherence: A measure of data quality relating to the degree to which the census data can be combined in different ways and for various purposes with statistical information from other sources.

Comparability: A measure of data quality relating to the degree to which statistics are comparable between geographic areas and over time.

Combined census: A census in which some information on the numbers and characteristics of the population are derived from information taken from administrative data sources held for non-statistical purposes, but where other information that is not available from such sources is collected directly from individual persons and households by means of full or partial field enumeration or from other sample surveys.

De facto census: A census based on a count of persons at where they were present on the reference date.

De jure census: A census based on a count of persons at their place of usual residence on the reference date.

Deterministic method: A method without a random component that thus always leads to the same outcome

DSE: Duel System Estimation - a statistical method, based on a capture-recapture technique, applied to estimate the size of a population.

Hypercube: A high-dimensional statistical tabulation of, typically, four or more dimensions

LAU: Local administrative unit. The classification of local administrative areas used by Eurostat. Levels 1 and 2 equate to those areas that were previously classified, respectively at the NUTS 4 and 5 levels.

Metadata: Information about the content, structure, quality and other relevant characteristics of a register

Microdata: As used in these Guidelines, information in a register relating to a single entity or entities

Numerical address: Code linking to the address

Population register: A register of residents of the country

Probabilistic method: A method with a random component that thus not always leads to the same outcome.

Process quality: The quality of a statistical process as evaluated by the methods used, cost effectiveness and response burden.

Punctuality: A measure of data quality relating to the delay between the date of the release of the results and the target date (the date by which the data should have been delivered).

Register: A systematic collection of unit-level data organized in such a way that updating is possible. Updating is the processing of identifiable information with the purpose of establishing, bringing up-to-date, correcting, or extending, the register, that is, keeping track of any changes in the data describing the units and their attributes.

Register-based census: A census in which the data on the numbers and characteristics of the population are derived from information taken from administrative data sources held for non-statistical purposes. No information is collected directly from individual persons or households.

Relevance: A measure of data quality relating to the degree to which statistics meet current and potential needs of the users.

Statistical register: A register created for statistical purposes. They are typically created by transforming data from registers or other administrative data sources.

Synchronisation: The transmission of data recorded on handheld devices in the field to a central server.

Timeliness: A measure of data quality relating to the period between the availability of the information and the event or phenomenon it describes.

Traditional census: A census based on the direct count of all individuals and the collection of information on their characteristics through the completion of either a self-completion or interview-based questionnaires, either in a paper or electronic format.

Unique ID/key: A single alpha numeric identifier that relates a characteristic or variable to a particular entity (person, household or dwelling) across a range of different registers or administrative data sources.

XML: Extensible markup language. In computing a language that that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.