

Distr.: General
13 September 2019

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Twenty-first Meeting

Geneva, 18–20 September 2019

Item No. 5 of the provisional agenda

Future censuses beyond 2020

Future Censuses Beyond 2020

Note by the Steering Group on Population and Housing Censuses

Summary

This paper is a draft of a paper being prepared by the Steering Group on Population and Housing Censuses at the request of the Bureau of The Conference of European Statisticians. The draft is presented at this meeting in order to obtain feedback to inform subsequent drafts of the paper before it is finalized.

I. Introduction

1. Methods employed for Population and Housing Censuses have been constantly evolving over the past decades, either to adapt to changes in society and in information requirements, to leverage opportunities with new or different data sources such as registers and administrative files or based on opportunities offered by innovations in enabling technologies.
2. This is certainly the case for many UNECE countries. Information available on the plans by UNECE countries for the 2020 round of censuses clearly shows that a large proportion of countries in the region are planning or evaluating the introduction of important methodological and technological innovations, and the adoption of alternative approaches based on the increased use of registers and administrative data. For many countries, the ultimate objective is to move to a fully register-based census in the 2020 round or later. Others are expanding the application of traditional approaches to further introduce self-enumeration, particularly via the Internet, and to optimize various census-related processes with new technologies.
3. The introduction of such changes is not without challenges. UNECE and Eurostat experts discussed such challenges at the Expert Group Meeting on Censuses (Geneva, 4-6 October 2017), including a panel discussion on the future of censuses beyond 2020. Some work on this subject has also been taking place at the European Union (EU) level (doc. ECE/CES/GE.41/2017/21). In the discussion, consensus emerged that providing data on a decennial basis will no longer be acceptable for many or most countries in the future, and that more frequent updates will be necessary. It was also observed that the discussion on the future of censuses should not focus solely on incremental improvements based on the past, but may require a paradigm shift for many countries.
4. This paper has been prepared at the request of the Bureau of the Conference of European Statisticians by the Steering Group on Population and Housing Censuses, representing the UNECE Group of Experts on Population and Housing Censuses. The paper provides an analysis of many of the challenges and issues for the future of population and housing censuses after the 2020 round. It is hoped that this analysis will inform the drafting of recommendations for the 2030 round of censuses.
5. In developing the analysis, the following issues are considered:
 - a) The possible need for the use of different population measures;
 - b) The innovative methods being developed, multi-mode approaches using surveys, administrative data, and new data sources;
 - c) The increasing importance of geo-referenced information;
 - d) The move from a decennial census to annual data collections;
 - e) The use of 'big data', and the potential for competition with 'big data' producers;
 - f) The need to capture societal change in combination with a requirement to preserve comparability over time and across jurisdictions: will concepts such as 'household' and 'usual residence' remain relevant?

II. The situation in the UNECE region - summary of survey results

6. As part of the present work, in September 2018 the Steering Group conducted a survey on future censuses beyond the 2020 round among UNECE countries. The objective of the survey was to explore countries' thinking for the next census after the 2020 round, which will possibly take place in 2030 or 2031 in countries with a decennial censuses, or in 2025 or another year in countries where the census will be conducted on annual basis or with annual updates. For simplicity, in this document we will refer to "2030 round".

7. The survey was carried out by UNECE in September 2018, through a short online questionnaire. Responses were received from 47 UNECE countries, representing a very good response rate of 90 per cent. The information collected in the survey and presented in this section should be considered as unofficial and tentative, given the long time frame.

8. Among the countries that responded, 13 countries are conducting a traditional census (28 per cent) in the 2020 round, 17 a combined census (37 per cent) and 13 a fully register-based census (28 per cent). The percentage of UNECE countries conducting a traditional census is dropping significantly, considering that this was by far the most common methodology in the 2000 round (adopted by 83 per cent of the countries) and the 2010 round (64 per cent of the countries).¹

9. With regard to the 2030 round, 19 countries (41 per cent) indicated that they are contemplating – even tentatively – changes in the methodology compared to the 2020 round. Moreover, seven countries (15 per cent) replied “maybe” to the same question.

10. Of the 26 countries contemplating possible changes, 11 are conducting a traditional census in the 2020 round (five of which are planning to move to a combined census, and two to a register-based census); 13 countries are conducting a combined census (eight of which are planning to move to a register-based census, five planning to still do a combined census); and two countries are planning to conduct a register-based census both in the 2020 and 2030 rounds.

11. Of the 13 countries that are not contemplating methodological changes between the 2020 and the 2030 round (28 per cent of all responding countries), 10 countries are conducting a register-based census, and three a combined census.

12. Only seven countries (15 per cent) responded “do not know” with regard to the possibility of a change in methodology between the 2020 and 2030 rounds. This indicates that a large majority of the countries already have some plans – at least tentative – on the methods to be used for the 2030 census round.

13. Figure 1 presents the distribution of countries by census methods used in the 2020 round and planned to be used (tentatively) in the 2030 round. All 13 countries that are conducting a register-based census in the 2020 round plan to use the same method in the 2030 round, as expected, although two of these are contemplating methodological changes as mentioned above. Almost half of the countries that are conducting a combined census in the 2020 round (eight out of 17) are planning to move to a register-based census in the 2030 round. Finally, only two of the 13 countries conducting a traditional census in the 2020 round are planning to still use this method in the 2030 round. Among the others, five countries are planning to move to a combined census, two to a register-based census, and four countries do not yet know.

Figure 1

Method 2020 round	Census method planned 2030 round					TOTAL
	Traditional	Combined	Register-based	Rolling	Other / do not know	
Traditional	2	5	2		4	13
Combined		8	8		1	17
Register-based			13			13
Rolling					1	1
Other/do not know		1			1	2
TOTAL	2	14	23	-	7	46

¹ Valente, P., “From the 2010 to the 2020 census round in the UNECE region – Changes in census methodology and technology, plans by countries and opportunities for international cooperation”, paper presented at the UNECE Group of Experts on Population and Housing Censuses, Geneva, 30 September to 2 October 2015.

14. Countries were also asked whether they are contemplating other important changes – other than the main census method – or innovations for the 2030 round. Only four countries, all conducting a register-based census, indicated that they expect no major changes or innovations; nine countries replied “do not know”, and 33 countries (72 per cent) indicated one or more changes or innovations expected for the 2030 round.

15. The change expected by most countries is the move from a decennial census to a more frequent interval (21 countries, 46 per cent of the total), followed by increased geographical detail in the census output (18 countries), increased use of geo-referenced information for data collection (15 countries), larger integration of the census with other statistical programmes (13 countries) and use of “big data” or other sources (11 countries) (figure 2).

Figure 2

Change/innovation	Census method planned 2030 round				
	Traditional (2 countries)	Combined (14 countries)	Register-based (23 countries)	Other/do not know (7 countries)	TOTAL (46 countries)
Move from a decennial census to a more frequent interval		5	15	1	21
Increased geographical detail in the census output		6	11	1	18
Increased use of geo-referenced information for data collection	1	7	5	2	15
Larger integration of the census with other statistical programmes		4	9		13
Use of “big data” or other sources		6	5		11
None			4		4

16. Considering the changes and innovations separately for the different census methods, countries with register-based censuses seem more likely than those with a combined census to be considering moving to a more frequent interval. On the other hand, countries with a combined census seem more likely to make increased use of geo-referenced information for data collection. Both results could have been expected, since a combined census is less suitable than a register-based census to be conducted frequently, and can increase its efficiency by making use of geo-referenced information.

17. It is interesting to note the significant number of countries considering the use in the 2030 round of “big data”, as part of a combined census (six countries) or a register-based census (five countries).

18. Countries were asked whether, in view of the 2030 round, they foresee needs for changes in key concepts. Only 14 countries (30 per cent of the total) foresee needs for changes in key concepts. This includes eight countries that are planning a register-based census and six countries that are planning a combined census in the 2030 round (figure 3).

19. The concept that the most countries (11) consider in need of change is that of ‘usual residence’. For many of these countries (including Canada, Cyprus, Czechia, Israel, Russia, and Spain), changes may be needed when administrative data are used as part of a combined or register-based approach. In Canada, expansion of the concept of measured population is also needed, as a consequence of societal changes and user demand.

20. The concept of ‘household’ may need changes, according to six countries. Various countries (including Cyprus, Czechia, Hungary and Serbia) noted that the housekeeping concept is difficult to apply when register data are used, so the household-dwelling concept should be used.

21. Changes may also be needed for some key concepts related to the core content, according to four countries. In particular, Iceland and Spain noted that the current activity status concept using the ILO definitions of employment and unemployment is problematic when using register data.

Figure 3

Key concepts for which changes may be needed	Census method planned 2030 round				TOTAL (46 countries)
	Traditional (2 countries)	Combined (14 countries)	Register-based (23 countries)	Other/do not know (7 countries)	
Definition of a household		1	5		6
Usual residence		5	6		11
Core content		1	3		4
Other		2			2

22. Different reasons can lead countries to consider changes in their census. With reference to the 2030 round, the driver for changes reported by most countries (11) is technological changes, followed by societal changes (8), anticipated user needs (5), and expressed user needs (4).

23. Finally, countries were asked whether the Conference of European Statisticians Recommendations on Population and Housing Censuses for the 2020 round are aligned with their thinking on the 2030 round. The large majority of countries (35) replied “yes”. Only Canada replied “no”, particularly indicating that changes are needed for the population definitions (a comment that was shared by Germany), and noting that the essential features should be clarified in the context of combined approaches. France noted that censuses should try to count each individual in one place.

24. To summarize, the following main results emerged from the survey:

- a. The trend of moving away from traditional censuses towards combined and register-based censuses is expected to continue strongly in the 2030 round;
- b. In the 2030 round several countries are planning to move from a traditional to a combined or register-based census (at least seven), or from a combined to a register-based census (at least eight);
- c. About 70 per cent of the countries are contemplating important changes – other than the main census method – or innovations for the 2030 round, including the move from a decennial census to a more frequent interval (21 countries); increased geographical detail in the census output (18 countries); and increased use of geo-referenced information for data collection (15 countries);
- d. The use in the 2030 round of “big data” will be considered by at least 11 countries, as part of a combined or register-based census;
- e. Changes are considered necessary in key census concepts such as usual residence (11 countries), household (six countries), and current activity status. For most countries, changes are needed in connection with the use of data from registers;
- f. The main driver for changes is technology (reported by 11 countries), followed by societal changes (8), anticipated user needs (5), and expressed user needs (4);
- g. The large majority of countries indicated that the CES Recommendations for the 2020 round are aligned with their thinking on the 2030 round, although changes are needed to selected parts and key concepts.

III. Evolution of approaches to population censuses – a paradigm shift?

25. As can be seen in the results of the survey of UNECE countries, many are exploring or moving to new approaches to conduct their census and produce population counts in the 2020 or 2030 census round. As new approaches are explored and implemented, some challenges in adhering to existing *Recommendations on Censuses of Population and Housing* are brought to light, especially in the case of approaches based on registers or administrative sources. These challenges are further compounded by the natural evolution of social and societal factors, and of changing requirements from data users, adding pressure in ensuring the relevance of national census programmes.

26. Some countries benefit from a strong system of registers from which their census outputs have been derived for some time already. This is the case for Sweden, Finland and the Netherlands just to name a few. Others, such as Italy and Israel, are building such systems, and are in the normal evolution process of implementing their census from these.

27. Some countries do not have the benefits of universal registers of any kind. This does not, however, preclude the production of population counts from administrative sources, especially with a combined census approach. Statistical registers are being developed for this purpose in some countries (e.g. Canada and the UK). These are created by linking several administrative data sources to represent a universe, either of persons or of dwellings. The term ‘statistical registers’ is used to indicate that the registers are constructed from various administrative data sources and will be used only to support the NSO’s mandate to produce protected pseudonymous statistical outputs².

28. When a country decides to move from a traditional census to a register-based approach (or a combined approach), it is normally part of a strategic decision by the NSO to develop a register-based statistical system which will remain at the centre of national statistical production. Such changes represent a paradigm shift, and are often associated with a generalized rethinking of the overall statistical production process, including the frequency and timeliness of the various statistical activities. With regard to the census, in many countries the transition from a traditional to a register-based approach is considered a key step towards the possibility of producing census-type data at a higher frequency compared to the traditional census (normally taken every ten years), and of publishing the results in a shorter timeframe after the reference date. For most, this shift has as an explicit objective the production of at least a limited set of basic population and housing statistics on an annual basis. Moreover, moving to the use of registers or administrative data for the census often implies rethinking the population bases and key applied concepts such as those of place of residence and private household. In doing so, consideration has to be given to the characteristics of the information available in the registers, but also to the evolution of the society and of information needs.

IV. Population base – various challenges with new methods based on administrative sources

A. Coherence when produced from multiple sources

29. Information on individuals and dwellings in a census should normally refer to a unique and well-defined reference period. This requirement aligns with one of the essential features of a census, that of simultaneity. The use of various sources in the production of census counts can often result, though, in different reference dates or periods being used. For example, the reference period of various administrative sources is often not the same, and may not match the reference date of the portion of the census coming from direct enumeration in the case of combined approaches. This may make it difficult to choose the most appropriate source to

² That is, outputs from which direct identifiers have been eliminated or transformed but indirect identifiers remain intact.

produce the counts. One source could have a similar reference date to the enumeration date, but could also have important weaknesses. As more and more countries move to partial or full register or administrative data approaches, it may be desirable to review the essential features of a census in the next version of census *Recommendations*, to possibly better align with the imperatives of the new approaches.

B. Coherence in concepts for place of residence

30. The direct enumeration of the base population can be conducted using either a *de jure* or a *de facto* approach. The place of residence concept in this context is then defined based on the enumeration concept. By definition, the concepts of place of residence may be different for directly enumerated population versus those produced from registers or administrative sources. If such differences exist, they may be difficult to reconcile in the production of the base population in a census using a combined approach. This could even be more complex if multiple administrative files are used. In this context, it could be challenging to establish the place of residence based on current concepts for population groups which are more mobile, or for people with more than one residence. It may also be difficult to determine whether a ‘usual residence’ concept in administrative data sources corresponds to the one applied in direct enumeration in traditional census approaches.

31. The challenges with the use of administrative data sources in a census context can be further compounded by a lack of stability in coverage in some of these sources. Changes in concepts found in administrative data sources could take place over a short period of time, with limited capacity to evaluate their impact. Some could even go undetected. In this context, the use of administrative data in the census represents a possible vulnerability as changes may take place in coverage, reference dates, available variables, or other characteristics. Such changes could impact comparability over census cycles, a key element for census data users.

32. Future *Recommendations* should cover more extensively the various challenges associated with the use of multiple sources in the production of census output, particularly in the context of combined approaches.

C. Households with several homes

33. There is an increasing proportion of persons and/or households who have or who may be staying in more than one home during the course of a year. As indicated in the Recommendations on Censuses of Population and Housing, the census prescribes that the unit of enumeration be clearly defined. This is generally done with the registration of the permanent residence of a person, although this may be increasingly difficult to determine, or the permanent residence may change over the course of time.

34. To reflect these increasing changes in population behaviour, and to ensure relevance to data users, it would be advisable in the future to register, in addition to the permanent residence, a secondary residence if one stays there regularly for a sufficiently long time, e.g. more than three months a year. Criteria for this determination should be included in future *Recommendations*. In many national censuses, a dwelling that is no-one’s permanent residence is considered an unsettled or vacant dwelling (not occupied by usual residents). It could also be advisable to identify and differentiate dwellings that are used regularly but not permanently.

D. The challenges of measuring migration

35. One of the most complicated and prevalent population processes nowadays is migration, particularly international migration (migration between countries or crossing national borders). In many countries, migration information can be received from officially registered migration files or registers, which, however, tend to underestimate migration, particularly emigration. Many emigrants do not register their departure from the country for several reasons – not knowing this obligation where mandated, forgetting it or considering it inessential; often they also wish to keep the status of a permanent resident of their earlier home country in order to continue using its benefits. Therefore, international migration generally consists of two parts – registered and unregistered migration. If international migration is mostly unidirectional, the error caused by

unregistered migration cumulates over the years and distorts the assessment of the total number of emigrants and the number of population when relying on register data.

36. Such was the situation, for example, in Estonia before the 2011 census when, because of under-registered emigration, the number of population calculated from population register data was overestimated by almost 3 per cent. Assessment of migration and the count of population is further complicated by some special types of migration that do not coincide with the usual definition of international migration.

E. 'Pendulum' migration

37. 'Pendulum' migration is not a new phenomenon; in Europe, it evolved as early as the 19th century with the development of railway transport. International pendulum migration, however, is relatively new and only moderately spread. It is found mainly in border areas, particularly if the border separates countries with different living standards and employment possibilities. Population statistics has also treated pendulum migration relatively narrowly and exceptionally, but in some countries, especially in Eastern Europe, pendulum migration has reached quite high levels in recent decades.

38. Pendulum migration is generally defined as migration with a duration up to one week. For pendulum migrants, the place of residence is the residence of their family. For persons who, because of their work, have stayed away from their home for more than 12 months but spent most of their days off with their household, the residence of their household is considered to be their place of residence.

39. Pendulum migration can also be treated more broadly than in the above definition, considering a stay in another country for a longer period than one week. Defining residency and migration of people who move between different countries due to their work or studies according to the rules of population statistics is generally difficult, even for the people themselves.

40. This phenomenon will certainly continue to expand in the future, and will potentially expand the challenges of counting population in the right places. Uniform understanding of pendulum migration and its precise consideration presuppose concrete internationally-agreed answers to the following questions:

- (a) Migration of what duration is considered pendulum migration?
- (b) Depending on the characteristics of work, the migration period can also be of variable length.
- (c) If, however, the period between visits to the regular household grows longer than generally defined, the pendulum migrant may become considered, according to national rules, a permanent resident of the country where he or she works.
- (d) Pendulum migration has currently been defined only for people with families. However, is a young person who lives at their parents' home and working abroad but who regularly visits home not a pendulum migrant?
- (e) The definition of pendulum migration pertains primarily to employment. It would be natural to broaden pendulum migration to studies, particularly for domestic pendulum migration.
- (f) Very often, students who frequently visit their parents' homes, for example on weekends, have not registered as residents of the place of study, as they consider the student residence or rented apartment a temporary residence. Some national census rules on where to include people may even suggest this approach.
- (g) Pendulum migration is not among these indicators of official statistics that are presented annually.

F. Transnationality

41. Although, in principle, transnationality is not a new phenomenon either, attention has been paid to it only recently, particularly in the present century. So far the concept has not been given a precise definition. Transnationality is usually understood as living in two (or more) countries, while one of them is considered the country of home or origin, and connection with it is also retained while living in another country or countries.

42. Pendulum migration clearly is a particular case of transnationality. In the case of pendulum migration, the migrants' connection with the homeland is particularly strong due to the family living there. Transnationality has been defined through migration, although migration is not always an obligatory condition (particularly in the case of great distances); close contact with relatives in the homeland may be sufficient.

43. The greatest problem related to transnationality at present is the ambiguity of its definition. Still, the indicator is essential for many countries, and society's interest in it is relatively large.

V. Ensuring coverage and quality with new approaches

44. One of the key aspects of the quality of a census is the measurement of the coverage of the enumerated population. Various methods currently exist for doing so, the most prevalent possibly being the post enumeration survey approach. The transition to census approaches partially or totally based on registers and administrative sources may require a review of the methods used to measure coverage. This topic is partially covered in the UNECE *Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses*³.

45. Quality should possibly be approached from a 'fit for purpose' perspective. Despite the fact that countries generally tend to adhere to international principles and recommendations in the conduct of their population and housing census, fit for purpose could mean something different for every country based on particular national use of census outputs, whether it be to support funding transfers between levels of government, the application of the electoral process, or even none of these things. One aspect of census-taking sometimes omitted is the proper measure of population coverage. This activity may even become more problematic and difficult in the context of an increasing periodicity of the release of census results, for example on an annual basis.

46. The census will generally aim at enumerating the population of a defined territory as exactly as possible, and this calls for very small levels of geography. Measured concepts must be relevant to the various uses for the data produced, coherent over time as much as possible while also addressing new emerging phenomena. All aspects of a census related to timeliness, relevance, coherence, accessibility and interpretability are important, but special emphasis should possibly be put on precision.

47. How can the use of data from administrative sources in a census context cause issues with coverage? Conceptually, these sources should enable adequate coverage of the entire population. This is especially critical in the absence of actual registers, as is the case in Canada and the UK for example. Administrative sources need to be available in a timely fashion. They also need to pertain to the appropriate reference period. When using a number of different sources, the reference period will generally differ from one source to the other. Different attributes within a single source may even have different reference periods. For example, the reference period in an income tax file may be the previous calendar year, while the address of the person may be the current one.

48. In this context, the choice of reference date or period should be ideal in relation to the availability of the optimum version of the registers or administrative sources used to create the population count. For countries

³ Available from <https://www.unece.org/index.php?id=50794&L=0>

contemplating a combined approach for the creation of their population count (mix of direct enumeration and of administrative sources), the selection of the proper reference date may become even more problematic as the optimum period to produce each segment may be different.

49. Additional challenges may exist with the use of multiple registers and/or administrative sources. One relates to the integration of these sources via record linkage approaches. The precision or quality of the linkage keys may have a direct impact on linkage errors, and by the same token, on coverage. One must find the proper equilibrium between false links (false positives resulting in under coverage) and missed links (false negatives resulting in over coverage). The complexity of this operation may grow with the increase in the number of sources used, increasing the number of links required, and depending on the attributes of these sources in terms on linkage keys, impact coverage and quality of the information.

50. The challenges may be further compounded by the uncertainty of signals. The absence of information on a person should not automatically be interpreted as an exit from the population, any more than the presence of a signal should be interpreted as being part of the population. The use of multiple administrative sources to improve the quality of the measure about each individual will most likely result in over coverage, as out-of-scope cases tend to be more difficult to identify.

51. Coverage studies tend to be designed to produce aggregate results, rather than for the smallest levels of geography at which census results have traditionally been produced. As indicated in the section on population bases, there will also be added complexity with the evolution of social and migratory patterns. Future *Recommendations on population and housing censuses* should have an increased emphasis on the measurement of the quality of census results.

VI. Identified needs for modernization of geospatial statistics

52. There is a growing awareness that traditional surveys and censuses with fixed output areas do not meet the requirements of data users for territorial flexibility. In order to meet these growing demands, a new territorial dimension to the production of statistics, supported by a point-based framework, is needed. By linking administrative and/or census data with address coordinates, it is possible to obtain information on a wide range of functional areas (town/village, coastline, etc.), including small areas such as grids.

53. The demand for location-related statistical information has increased sharply in all sectors in most countries: government, commerce, non-profit, academic and civic. The growing demand for geospatial statistics has emerged because people, administrations, government bodies and other organizations want to gain insight into specific populations or into a subject that is of increasing interest in temporal and spatial detail. An important aspect of this insight is the location of the populations under consideration; in many cases, this insight is particularly required at the local level. Moreover, the rapid growth in mobility of people, products and services implies the need for better insight into complex dynamics within and between regions. Previous assumptions about where people live, work, play or have access to services are becoming less valid in today's world.

54. Information on flows of people and business transactions between locations, as well as more traditional regional demographic and business statistics, are used in a growing trend towards “place-based decision-making”.

55. A key element of making spatial statistics more timely, accurate and cost-effective would be for all countries to establish a spatial reference framework for statistics. Its central element will be the geocoded administrative address, included in a building and/or dwelling register(s). These registers should provide a reference framework for the geocoding of all future censuses, as well as for the linking of administrative data sources to geographical locations. An important aspect of the spatial reference system is that it needs to be equipped with unique identifiers that are stable over time and that can be used as unique keys for referencing all relevant information to them. The data making up this spatial reference system must also be easily accessible.

56. Census information at accurate and standardized geographical levels is essential to facilitate benchmarking and to achieve better quality of statistical geospatial data production. In the context of a census, it is important to have georeferenced unit data, preferably at the level of x,y coordinates. Understanding the origin, production process and other aspects of the quality of geospatial data is essential for the production process.

57. The landscape for integrating geospatial and statistical data is very complex. The challenge is to understand how best to achieve this integration in an efficient and coherent way. As part of the next *Recommendations on population and housing censuses*, or even before, it might be desirable to develop a coherent and systematic approach to combining statistical and geospatial data. An approach of this type could be based on the internationally-agreed Global Statistical Geospatial Framework (GSGF). A paper produced by Poland⁴ provides a lot of insight in how this could be achieved.

A. Example from Poland

58. In Poland, processing of statistical information with geocoding was conducted for the first time during preparatory work for the 2010/2011 census round. Due to the lack of equal quality of reference materials, different steps concerning spatial accuracy geocoding objects were adopted. For introductory pre-census work, the reference geodetic data from the State geodetic and cartographic resources for a whole area of the country, contained: Administrative division borders, Geographical Names, Topographic Data Base, Cadastral Data, ortophotomap, LPIS (Land Parcel Identification System). From statistical resource materials such as: situation sketches, statistical maps with statistical division boundaries (statistical regions, enumeration areas) and descriptive data from The National Official Register of the Territorial Division of the Country (System TERYT) regarding ID of administration levels, city, town, village, statistical area, street, building, dwelling were used:

- a. The first stage of geocoding work was to obtain address data from the administrative sources and registers containing spatial data. For that purpose, the address points were obtained from the register containing administrative division borders, in which these data are available.
- b. Due to the lack of full coverage the data in the administrative division border register, the second step of work had to be conducted. The updating database procedures were conducted by all 16 Regional Statistical Offices (System TERYT units). To carry out this task, statisticians were granted access to a map application presenting address databases with map background of ortophotomaps and contours of buildings. For that purpose, all obtained geodetic and statistical resources were used. Editing options made it possible to modify the location and attributes of address points (city/town, street, address number), as well as to manual address point acquisition of new points directly in the application. The activity was to identify or to assess the spatial location of addresses and add it in the correct place.
- c. In case when the ortophotomap was older or was lacking data on buildings, address points were localized into cadastral parcels, close to a street and entrance (gate).

59. In the pre-census round enumerators had to verify whether buildings or address points existed, supplement the list with missing address points, verify the correctness of address points and determine their locations (x, y coordinates). The census enumerators were provided with mobile terminals containing an application for the pre-census round. The application allowed address data to be verified and corrected, and any non-existing address points could be removed; new address points, not entered in the list, could also be added. The enumerator could also view his/her current GPS location and assigned address points. The points were displayed as a list of address points, including their spatial positioning on the map. The application offered one of the following modes of work with address points: 1) address points – the list of address points available

⁴ http://www.unecce.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2012/census_meeting/14_Poland_Eng_01.pdf

for verification, 2) map – an orthophotomap with spatially distributed address points. The mobile device had orthophotomaps for the entire area of a given *gmina* (local community). With the map mode, backed by the GPS receiver, enumerators could locate address points that required verification, and adjust the placement of address points by pointing their correct location on the map. Spatial accuracy was determined by the GPS devices and by the possibility of access enumerators to points or buildings (for example: guarded residential with no access). The application had the function of automatic control for identifying or modifying the location of an address point within the *gmina* where the enumerator was working.

60. This Polish example has shown that the creation of a geocoding infrastructure for statistics and its integration into the statistical production process does not demand a complete redesign of enterprise architectures and statistical production processes. Small and stepwise improvements are possible. However, integration of statistical and geospatial information is a cornerstone in the modernization of official statistics.

VII. Increased reliance on estimation models in the production of census outputs

61. As countries migrate their census approaches from traditional methods to approaches largely based on statistical registers and administrative sources, new challenges emerge in producing or maintain certain census outputs. In some cases, the information found in the new sources may not be as complete or as precise for certain variables as the information previously produced with the traditional model. There might also be slight difference between the sources in the concepts measured.

62. This is case, for example, for identification of households. With a traditional approach, the household would be derived based on questions and concepts applied at the time of enumeration. This concept often does not exist in administrative data sources and needs to be constructed somehow. Most administrative sources contain information about individuals, and the relationship between individuals residing at a given address often needs to be derived as it may not be readily available in these sources. The concept of household may also actually differ somehow from one country to the next. Differences relate primarily to rules about where to include children in shared custody, students away from the dwelling while at school, people temporarily living in an institution (such as a hospital), people temporarily away at work, etc.

63. The next *Recommendations* should take into account the transformation in census approaches, and include content about the use of estimation approaches in the creation of census variables, and related considerations.

B. Example from the Netherlands

64. The backbone of the Dutch census is the central population register (PR), which combines all the municipal population registers. PR data for 1 January 2011 (the year of the last census) were used as the basis for a set of hypercubes (high-dimensional tables). The hypercubes focus on frequency counts, not on quantitative information. Data not available or derivable from the PR were taken from other registers. All register variables are now available from Statistics Netherlands' system of social statistical datasets (SSD), and their quality has been improved by applying micro-integration techniques. Micro-integration entails checking the data and adjusting those that are incorrect. It is widely assumed that micro-integrated data provide more reliable results when using registers, as they are based on a maximum amount of information. They also provide better coverage of subpopulations: if data are missing in one source, another source can be used.

65. In the past, micro-integration played a very important role in the SSD and in particular in censuses. However, in the future the role of micro-integration will be more limited, because in the Netherlands the new so-called system of base registers is better integrated than the set of registers that Statistics Netherlands received before. The system of base registers will ultimately consist of 13 registers of population (residents and non-residents), addresses and buildings, enterprises, real estate (boundaries, ownership, value, etc.), topography (maps: land, water, roads), motor cars (model, colour, ownership, etc.), taxable income, labour

(wages, employers, social benefits, etc.) and underground infrastructure (sewerage, cables, etc.). The system of base registers is based on legislation. One cabinet minister is responsible for each base register. The Minister of Home Affairs is responsible for the system of base registers.

66. Users of base registers can rely on their validity. Statistics based on base registers demand only a limited amount of data editing. As base registers are linked to one another (through identification numbers), it follows that statistical data are also coherent. Contrary to sample survey data, no estimation methods are needed for small area statistics and non-response problems do not exist. However, some data are delayed, for example information about self-employed persons. One should not forget that registers are not developed for statistical purposes and do not always contain the statistical concepts required. In some situations different statistical concepts are acceptable. In other situations one tries to derive information about the required statistical concept. If such derivations are not satisfactory, one could finally decide to fall back towards sample surveys with the required statistical concept, but with more limited publication possibilities compared to register data.

VIII. Use of Big Data in a census context

67. In addition to registers and government-generated administrative sources, NSOs will be pressured to look at alternative data sources for the production of official statistics, including census outputs. Many NSOs are already exploring or exploiting such sources for many of their programmes, such as scanner and other transactional data in the production of economic statistics, for example.

68. The methods for acquiring these types of data are also evolving. In addition to obtaining data files directly from data producers and providers, the use of methods such as webscraping and remote sensing, to name just a couple, is expanding.

69. The use of information from these sources may present more challenges than is the case with registers and administrative files from government sources. The information may not be accompanied by complete metadata, making it more difficult to assess the quality or the completeness of the data source. Again in this case, the inherent concepts related to the files may also not match exactly what is available from some registers and what was directly collected traditionally from respondents.

70. Future *Recommendations* should more completely address the various aspects of the use of big data and of other unstructured data sources in a census context.

A. Example from Estonia

71. Statistics Estonia (SE) conducted a pilot project in 2017 testing the possibility to use mobile positioning data (MPD) for their register-based census approach to solve the problem of differences between registered and actual places of residence, which cause biases in population statistics. Mobile positioning is the detection of the location of the mobile device. Many technologies allow mobile positioning, which can be network or handset based.

72. Statistics Estonia found volunteers to participate in the pilot project, and a set of potential addresses was created for each participant based on registers. Then the anchor points of the places of residence were estimated based on the mobile positioning data. The anchor points and other auxiliary information were then used to build a model for selecting the most probable place of residence for each person from the set of addresses. In addition, the data of actual place of residence of volunteers was collected to compare positioning data with actual address.

73. Analysis of the accuracy of identified home anchor points and home-workplace anchor points in comparison to actual home address data provided by the persons was done within the analysis of the accuracy indicator results.

74. A home anchor point or a home-workplace anchor point could be calculated using MPD for 93 per cent of the respondents. For some individuals, anchor points could not be calculated using the MPD method and, consequently, MPD could not be used for their residence validation due to their limited telephone use or technical issues associated with processing and interpretation of data, including calculation of anchor points.

75. In the case of 82 per cent of the respondents, the residence address was located within the range area of the home anchor point or the home-workplace anchor point. The disadvantage of observing settlement units in relation to the centroid of an anchor point range area stems from the fact that, normally, the range area of a mobile mast includes several settlement units (especially in sparsely populated areas) and this does not necessarily mean that the anchor point has been incorrectly calculated.

76. The sample-based comparative analysis provided an opportunity to analyze the sources of errors in anchor point calculations and the dependencies of the results on various factors. This knowledge can support further development and supplementation of the anchor point methodology, which can be expected to lead to significant improvement in data validation.

77. To conclude, census team assessed that the new data sources and outcomes were acceptable but did not implement this approach in census statistical process at the time. The main reason was the absence of a legal basis for the use of this information. The use of mobile data in the census requires an amendment to various Acts, which is a complex activity. However, the development of this methodology is promising for the future. The methodology will be improved as there is a need for statistics to meet user needs and to consistently map population changes and assess trends in a timely manner. In order to achieve this, there is a need to use new data sources.

IX. Balancing User Requirements

78. With the proliferation of data sources including big data, the advancements in the capacity and methods to exploit large data sets, and the increased value that is put on data assets in a world in constant evolution, there is an increased demand for timeliness in statistics. Discussions at the European level emphasize this need for timeliness for the future census rounds, increasing the move towards alternative statistical approaches based largely on the use of administrative registers and other data sources. The use of these data sources is often less expensive and can produce regular information in shorter timeframes than methods based on direct enumeration.

79. According to the principles and recommendations for population and housing censuses, data quality can be assessed through 10 dimensions: accuracy, relevance, reliability, timeliness, punctuality, accessibility, clarity, coherence, comparability and metadata.

80. In trying to meet the changing and increasing demands from data users, timeliness and accuracy or comparability over time often cannot be achieved simultaneously. When working with administrative data sources, it is possible to produce and disseminate data very quickly, but this could be at the cost of fewer quality control processes and/or the use of modelling to make up for some of the shortcomings of some of the data sources. Various quality assurance controls could be used, but generally at the cost of timeliness as the data would then be released a longer time after the reference period, reducing relevance for the users.

81. For many national statistical organizations, the opportunities with new data sources may create conflicts with increasing user expectations. For many NSOs, the emphasis on producing official statistics of measurable high accuracy has often been realized at the price of timeliness. The changes to new approaches may require a paradigm shift in that area. New criteria for quality could shift primarily to the satisfaction of user needs. The challenge would rest in how to assess user satisfaction in order to adjust the trade-off between accuracy and timeliness!

82. Census outputs are generally not exclusively produced for short-term use. They are structural data for a country that need to be reliable at a national level but also at a very local level. They are used to guide medium

and long-term policies and infrastructures choices, and support social and economic programme evaluation. Part of their strength is derived from the ability to measure changes over time.

83. Another challenge with increased timeliness may be linked to varied user needs, depending on their profile. Policy users and many researchers, for example, may need to have very precise results for small areas or small groups, but it takes time to assure accuracy at these levels. Users at a national level may prefer more timely data even if the quality is not as strong at a local level.

84. Another aspect linked to changes in census methods is the capacity of data users to understand these changes and their possible impact on census outputs. Some users may express concerns regarding the methods used to produce outputs, for example modelling, and the possible reduction or loss in comparability with previous censuses and other sources.

85. Future *Recommendations* on censuses of population and housing may need to provide direction in considering these potential trade-offs.

- (a) What is the best balance between timeliness and accuracy or completeness when taking into account user needs? What criteria should be considered?
- (b) Is it acceptable to change the definition of a concept (population base, for instance) to enable easier use of administrative data or to better suit new sources of data to achieve improved timeliness? Should definitions take into consideration this need for timeliness, or should definitions be defined independently of this need?
- (c) On some fronts, the measurement of certain characteristics may move from 100 per cent of the population to a sample, leaving the impression of a reduction in quality. This would be the case, for example, where the population count is produced from registers and characteristics traditionally collected directly by the census are now obtained via a sample, either through direct enumeration or by administrative sources.

X. Conclusion

86. The pace of change in the ways that censuses of population and housing are conducted internationally, and more specifically in the UNECE region, has increased in the last couple of decades. As more and more countries explore and implement new methods, it can be anticipated that this pace will be maintained for the next round.

87. Census *Recommendations* have traditionally played an important role in guiding the development and implementation of census approaches. As these important transformations take place in national censuses, it is increasingly important that they be guided by a set of common principles and guidelines, to ensure this important statistical exercise maintains its relevance for many more cycles to come.
