

New Frontiers for Censuses Beyond 2020



UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

NEW FRONTIERS FOR CENSUSES BEYOND 2020

Prepared by the Conference of European Statisticians
Steering Group on Population and Housing Censuses



United Nations
Geneva, 2020

Acknowledgements

The present document has been prepared by the Conference of European Statisticians Steering Group on Population and Housing Censuses, with contributions from the following members:

Marc Hamel (Statistics Canada) – Chair
Valérie Roux (France)
Eric Schulte Nordholt (Netherlands)
Janusz Dygaszewicz (Poland)
Danilo Dolenc and Barica Razpotnik (Slovenia)
Cal Ghee, Justine McNally, Becky Tinsley and Jen Woolford (United Kingdom)
Arona Pistiner (United States)
Fabian Bach, David Thorogood and Adam Wronski (Eurostat)
Irina Zbarskaya (CIS-Stat)
Eduard Jongstra (UNFPA)
Srdjan Mrkic (United Nations Statistics Division)
Paolo Valente, Andres Vikat and Fiona Willis-Núñez (UNECE)

UNECE wishes to acknowledge their contribution with much appreciation.

Contents

1	Introduction	1
1.1	Summary of issues to be considered in revising the CES Recommendations on Population and Housing Censuses for the 2030 Round.....	2
2	The situation in the UNECE region – summary of survey results	3
3	Evolution of approaches to population censuses – a paradigm shift?.....	7
4	Population base – challenges with new methods based on administrative sources.....	9
4.1	Coherence when produced from multiple sources	9
4.2	Coherence in concepts for place of residence	9
4.3	Households and/or persons with several homes.....	10
4.4	The challenges of measuring migration	10
4.4.1	‘Pendulum’ migration.....	11
4.4.2	Transnationality.....	12
5	Ensuring quality with new approaches: a focus on coverage.....	13
6	The need to modernize geospatial statistics	15
6.1	Increased integration of geospatial data into the statistical system: an example from Poland.....	16
7	Increased reliance on estimation models in the production of census outputs....	17
7.1	Register-based census system and micro-integration: an example from the Netherlands.....	17
8	Use of big data for censuses.....	19
8.1	Using mobile positioning to determine place of residence: an example from Estonia	19
8.2	Balancing user requirements	20
9	Privacy, public acceptability and ethics.....	23
9.1	Privacy.....	23
9.2	Public acceptability	24
9.3	Ethics	24
10	Conclusion.....	27

1 Introduction

1. Methods employed for population and housing censuses have been evolving constantly over the past decades: 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; and to capitalize on opportunities offered by innovations in enabling technologies.

2. Such evolution is especially evident in many UNECE countries. Information available on the plans by UNECE countries for the 2020 round of censuses shows clearly 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 the following one. 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 along with the adoption of new and adjusted statistical methodologies is not without challenges. UNECE and Eurostat experts discussed such challenges at the 2017 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 taken place at the European Union (EU) level (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, including a refocusing of attention towards the quality of census outputs for many countries.

4. This paper has been prepared at the request of the Bureau of the Conference of European Statisticians (CES) 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. This analysis will inform the drafting of the CES recommendations for the 2030 round of censuses¹.

5. In developing the analysis, the following issues were considered:

- a) The possible need for the use of different population measures;
- b) The innovative methods being developed, multi-source 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 more frequent, including annual, data collections and outputs;

¹ The *CES Recommendations for the 2020 Censuses of Population and Housing* can be found at: <http://www.unece.org/index.php?id=41398>.

- 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?
6. In developing the analysis, a number of issues emerged that should be considered when the recommendations on population and housing censuses are revised. They are summarized below.

1.1 Summary of issues to be considered in revising the CES Recommendations on Population and Housing Censuses for the 2030 Round

- a) Review the essential features of a census, such as simultaneity, to better align with the imperatives of new approaches which may make use of multiple sources with different reference dates;
- b) Provide direction in navigating the trade-offs in the various dimensions of quality and the evolution of these as user needs change;
- c) Cover more extensively the challenges associated with the use of multiple sources in the production of census output, particularly with combined approaches;
- d) Develop criteria for determining secondary residence;
- e) Develop criteria to identify and differentiate dwellings that are used regularly but not permanently;
- f) Improve definitions and understanding of pendulum migration, including methods to measure coverage of the phenomenon when using registers or administrative sources;
- g) Increase the emphasis placed on the measurement of the quality of census results; in particular the importance and challenges of measuring coverage (reflecting the results of the work currently being done by the UNECE Task Force on assessing the quality of administrative sources for use in censuses);
- h) 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);
- i) Provide guidance on the use of estimation approaches in the creation of census variables and related considerations;
- j) Provide guidance on the various aspects of the use of new unstructured data sources for censuses.

2 The situation in the UNECE region – summary of survey results

7. 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 take place in 2029, 2030 or 2031 in countries with a decennial census, 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 all of these are referred to as the "2030 round".

8. The survey was carried out by the UNECE secretariat in September 2018, through a short online questionnaire. Valid responses were received from 46 UNECE countries, representing a very good response rate of 88 per cent. The information collected in the survey and presented in this section should be considered as unofficial and tentative, given the long timeframe to the 2030 round.

9. Among the countries that responded, 13 countries (28 per cent) are conducting a traditional census in the 2020 round, 17 (37 per cent) a combined census – a mix of direct enumeration and administrative sources and 13 (28 per cent) a fully register-based census. 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).

10. For the 2030 round, 19 countries (41 per cent) indicated that they are contemplating – even tentatively – changes in the methodology compared to the 2020 round, and a further seven countries (15 per cent) replied "maybe" to the same question.

11. 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 full 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.

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

13. Only seven countries (15 per cent) responded "do not know" when asked about 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.

14. Table 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 full 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 full register-based census, and four countries do not yet know.

15. 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 already 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.

Table 1
Census method planned for the 2030 round, by method used in 2020 round

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

16. 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); greater integration of the census with other statistical programmes (13 countries); and use of ‘big data’ or other sources (11 countries) (Table 2).

Table 2
Changes/innovations expected for the 2030 round

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

17. 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 are to be expected, since a combined census is less suitable for frequent conduction than a register-based census, while the efficiency of a combined census can more readily be increased by making use of geo-referenced information.

18. It is interesting to note that several countries are considering the use of 'big data' in the 2030 round as part of a combined census (six countries) or a full register-based census (five countries).

19. 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 such needs. This includes eight countries that are planning a full register-based census and six countries that are planning a combined census in the 2030 round (Table 3).

20. 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 full register-based approach. In Canada, expansion of the concept of measured population is also needed, as a consequence of societal changes and user demand.

21. 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.

22. 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 International Labour Organization (ILO) definitions of employment and unemployment is problematic when using register data.

Table 3
Key concepts for which changes may be needed for 2030 round

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

23. Countries have various reasons for considering changes in their census. For the 2030 round, the driver for changes reported by most countries (11) is technological change, followed by societal change (8), anticipated user needs (5), and expressed user needs (4).

24. 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 responding 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 for the context of combined approaches. France noted that an essential feature of a census is that it should try to count each individual in one place only.

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

- a) The trend of moving away from traditional censuses towards combined and full 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 full register-based census (at least seven), or from a combined to a full 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 of ‘big data’ in the 2030 round 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.

3 Evolution of approaches to population censuses – a paradigm shift?

26. As can be seen from the survey results, many countries 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 employed, 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 societal changes and by resulting changes in requirements from data users, adding pressure in ensuring the relevance of national census programmes.

27. 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, to name just a few. Others, such as Italy and Israel, are building such systems, and are in the process of deriving their census data from these.

28. 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 or combined administrative data sets are being developed for this purpose in some countries, e.g. Canada and the United Kingdom. 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 mandate of a national statistical office (NSO) to produce protected anonymous statistical outputs.

29. 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 or adopt 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. This paradigm shift will also be reflected in the organization of the census production and the skills and capabilities required of staff. In many countries the transition from a traditional to a register-based approach to the census 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 must be given to the characteristics of the information available in the registers, but also to the evolution of the society and of users' information needs. It is also important, in register-based censuses, to define clearly for each census variable the main administrative source from which the information is derived.

4 Population base – challenges with new methods based on administrative sources

4.1 Coherence when produced from multiple sources

30. 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 produce the counts. One source could have a similar reference date to the enumeration date but could also have important weaknesses. Some of these weaknesses may be overcome by a model-based approach such as 'now-casting'. 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 better align with the imperatives of the new approaches.

4.2 Coherence in concepts for place of residence

31. Direct enumeration of the base population can be conducted using either a *de jure* or a *de facto* approach. The place of residence concept is then defined based on the enumeration concept. This concept refers to both the global level (country of residence) and the local level (place of residence within a country). 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. It could be challenging to establish the place of residence based on current concepts for population groups which are 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.

32. The challenges with the use of administrative data sources in a census 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, limiting the possibility of evaluating their impact. Some could even go undetected. The use of administrative data in the census therefore introduces 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 requirement for census data users.

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

4.3 Households and/or persons with several homes

34. 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 time.

35. 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 period of time, the length of which could be determined by local or national needs. 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.

4.4 The challenges of measuring migration

36. One of the most complex and increasingly prevalent population processes is migration, particularly international migration (migration between countries or crossing national borders). In many countries, information on migration is received from officially registered migration files or registers which tend, in most countries, to underestimate migration, particularly emigration. Many emigrants do not register their departure from the country for several reasons: not knowing that it is obligatory to do so, in cases where such an obligation exists; forgetting it or considering it inessential; or wishing to keep permanent residency status in the country of departure in order to continue using its benefits. International migration therefore 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 total population number when relying on register data.

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

4.4.1 'Pendulum' migration

38. 'Pendulum' migration, a particular kind of circular migration in which a person migrates repeatedly back and forth between two places, 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 opportunities. Population statistics have also treated pendulum migration relatively narrowly and as a rarity, but in some countries, especially in Eastern Europe, pendulum migration has reached quite high levels in recent decades. Pendulum migration has a spillover effect on measuring migration flows. The challenge will be the synchronization of (annual) census stock with population flows.

39. While not a new phenomenon in society, pendulum migration is a relatively new statistical concept for which there is as yet no internationally-agreed definition, either in the census recommendations or in international recommendations for migration statistics². In general, however, it is defined as migration with a duration of up to one week. For pendulum migrants, the place of residence is usually considered to be the residence of their family (according to the concept of usual residence). 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.

40. Pendulum migration can also be treated more broadly than in the above definition, by considering people who travel back and forth to another country, staying in each location for a period of longer than one week. Using the rules of population statistics to define residency and migration of people who move between different countries due to their work or studies is generally difficult, even for the people themselves.

41. This phenomenon will undoubtedly continue to expand in the future, compounding the challenges of counting population in the right places. Uniform and precise understanding of pendulum migration will depend upon concrete internationally-agreed positions on the following issues:

- What duration of stay is required for migration to be considered pendulum migration?
- Depending on the characteristics of work, the migration period can also be of variable length. How should stays of variable length be dealt with?
- If 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.
- The definition of pendulum migration pertains primarily to employment. It would be natural to broaden the understanding of pendulum migration to migration for other

² The 2016 UNECE publication *Defining and Measuring Circular Migration* considers pendulum migration as a particular kind of circular migration http://www.unece.org/fileadmin/DAM/stats/publications/2016/ECECESSTAT20165_E.pdf (pp. 11–12 & 19).

reasons, such as studies, particularly for domestic pendulum migration, or to retired persons who spend part of the year in another region or country.

- Students who frequently visit their parents' homes, for example on weekends, have in many cases not registered as residents of the place of study because they consider the student residence or rented apartment to be a temporary residence. Some national census rules on where to include people may even suggest this approach.

4.4.2 Transnationality

42. Although in principle transnationality is not a new phenomenon either, attention has been paid to it only recently. 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. Transnationality may also refer to potential migration, especially in cases of dual citizenship.

43. Pendulum migration clearly can be 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.

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

5 Ensuring quality with new approaches: a focus on coverage

45. One of the key aspects of the quality of a census is the completeness of coverage of the enumerated population. Under- and over-coverage caused by linking different sources must be considered independently of the quality of record linkage. Various methods currently exist for evaluating this aspect of census quality, the most prevalent being the post enumeration survey. 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 2018 UNECE *Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses*³.

46. Quality of content is also in question if several sources are used to augment a single characteristic for all individuals (for example, highest level of education attained). Quality could be defined as 'fitness for purpose'. Although 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 uses 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 things⁴. One aspect of census-taking that is sometimes overlooked is ascertaining how population coverage should be measured. This activity may even become more problematic and difficult with increasing periodicity of the release of census results, such as on an annual basis.

47. A census generally aims to enumerate the population of a defined territory as exactly as possible, and this calls for very small levels of geography. Concepts measured must be relevant to the various intended uses of the data produced, and as coherent over time as possible while also addressing emerging phenomena. All aspects of a census related to timeliness, relevance, coherence, accessibility and interpretability are important, but special emphasis may be put on accuracy.

48. How can the use of data from administrative sources in a census cause issues with coverage? Conceptually, such sources should enable adequate coverage of the entire population. This is especially critical in the absence of actual registers (especially for population), as is the case, for example, in Canada and the United Kingdom. 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. In such cases, the choice of the reference date or period for the census should take in consideration the actual or expected date of availability of the most complete version of the registers or

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

⁴ The question of defining quality for censuses based on new approaches and data sources is currently being tackled by the CES Task Force on Assessing the Quality of Administrative Sources for Use in Censuses, which will report its findings to the CES Bureau in 2021.

administrative sources used to create the population count. For countries contemplating a combined approach for the creation of their population count, 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 arise 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. A balance between false links (false positives resulting in undercoverage) and missed links (false negatives resulting in overcoverage) must then be found. The complexity of this operation may grow as a function of the number of sources used, increasing the number of links required, and depending on the attributes of these sources in terms of 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 a definite indication that they are part of the population. The use of multiple administrative sources to improve the quality of the measure about each individual can result in over-coverage, as out-of-scope cases tend to be more difficult to identify⁵. However, the analysis of administrative signs of life can help to correct for over-coverage.

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 assessment of the quality of census estimates, paying attention not only to coverage errors within countries, but also to coverage errors between countries.

⁵ Several UNECE publications provide guidance on techniques and considerations for integrating multiple administrative data sources: *Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses* (2018, *ibid.*); *Guidance on Data Integration for Measuring Migration* (2019) available from <https://www.unece.org/index.php?id=51143>; *Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices* (2012) available from <http://www.unece.org/index.php?id=28898>; and the *Guide to Data Integration for Official Statistics prepared by the High-Level Group for the Modernization of Official Statistics* (2018) available from <https://statswiki.unece.org/x/OQGzBw>

6 The need to modernize geospatial statistics

52. There is a growing recognition 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 census data with address points related to x,y coordinates, it is possible to obtain information not only within administrative or statistical areas, but also for a wide range of functional areas (towns or villages, urban areas, coastline, etc.), including small areas such as a grid (1 km² or hexagonal) or freely-defined areas of interest to users. Anchoring each source to the geospatial infrastructure provides record linkage that is not possible otherwise.

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, with enhanced temporal and spatial detail. An important aspect of this insight is the location of the populations under consideration; in many cases, such insight is particularly required at the local level. Moreover, the rapid growth in mobility of people, products and services implies the need for better understanding of complex dynamics within and between regions. Previous assumptions about where people live, work, play or access 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 being used together in what is now a growing trend towards "place-based decision-making".

55. A key means of making spatial statistics more timely, accurate and cost-effective would be for all countries to establish a spatial reference framework for statistics. A central element of such a framework is the geocoded administrative address, included in building and/or dwelling registers. These registers should provide a reference framework for the geocoding of all future censuses, as well as for linking 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. For a census, it is important to have georeferenced unit data, preferably at the level of x,y coordinates. This information allows NSOs to aggregate census data at any spatial level.

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)⁶. The text below summarizes a paper by Statistics Poland which provides insight in how this could be achieved.

6.1 Increased integration of geospatial data into the statistical system: an example from Poland

58. In Poland, the processing of statistical information related to geospatial information was introduced for the first time in the framework of the 2011 census. At the beginning, reference geodetic data from state geodetic and cartographic resources for the whole country were used. These data included geometry of: Administrative boundaries, Geographical names, Topographic Database, Cadastral data, Building register, orthophotomaps (satellite remote sensing) and the network of streets and roads. The statistical sources used such materials as: situational sketches, statistical maps with statistical division boundaries (statistical regions, enumeration areas) and descriptive data from the National Official Register of the National Territorial Division (TERYT system) concerning unique IDs of all levels of administration, city, town, village, statistical area, street, building, flat.

59. All the above data were collected in the statistical spatial database which supported the process of data collection (geo-coding) and were used for making census data available in relation to the spatial dimension.

60. Statisticians gained access to the application mapping address databases with the background of orthophotomaps and building contours maps. Editing options permitted updating the location and attributes of address points (city/city, street, address number), as well as manual acquisition of new address points directly in the application. It supported the identification or evaluation of the spatial location of addresses with the accuracy of x,y coordinates and added them in the right place.

61. This Polish example demonstrates 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.

⁶ Available from http://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/The_GSGF.pdf

7 Increased reliance on estimation models in the production of census outputs

62. As countries migrate their census approaches from traditional methods to approaches based largely on statistical registers and administrative sources, new challenges emerge in producing or maintaining 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 differences between the sources in the concepts measured (e.g. administratively formed households may differ from the household concept used in a traditional census). Small area estimation could be used where the information from sample sources is not sufficiently detailed.

63. This is the case, for example, for identification of households. With a traditional approach, the household would be derived on the basis of questions and concepts applied at the time of enumeration. The concept of household 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 differ from one country to the next, or even between different population groups within a country. 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.

64. In the examples above, various statistical methods may need to be applied to derive or estimate the required census variables. The use of estimation approaches may also be required in the production of multivariate outputs. Different methods would be required for topics with varying degrees of missing data, as will often be the case when using multiple data sources in the production of census results.

65. 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.

7.1 Register-based census system and micro-integration: an example from the Netherlands

66. 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 felt 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.

67. 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 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.

68. 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.

8 Use of big data for censuses

69. In addition to registers and government-generated administrative sources, NSOs are coming under increasing pressure 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. Using big data sources requires the development of processing methods. This development could be done in cooperation with academia or the private sector.

70. 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 web scraping and remote sensing is expanding.

71. The use of information from these sources may present new challenges compared with using 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. The inherent concepts related to the files may also not match exactly what is available from some registers and what was directly collected from respondents with traditional direct enumeration. Finally, in order to use such data for statistical purposes, public acceptability is needed, and a legal basis has to be developed if it is not already available.

72. Future *Recommendations* should more completely address the various aspects of the use of new unstructured data sources for censuses.

8.1 Using mobile positioning to determine place of residence: an example from Estonia

73. Statistics Estonia conducted a pilot project in 2017 testing the possibility of using mobile positioning data (MPD) for their register-based census 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 a mobile device. Many technologies allow mobile positioning, which can be network- or handset-based.

74. 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 on the basis of 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.

75. 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.

76. 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.

77. For 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.

78. The sample-based comparative analysis provided an opportunity to analyse 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.

79. The Estonian census team assessed that the new data sources and outcomes were acceptable but did not implement this approach in census statistical processes 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 legal Acts, which is a complex activity.

8.2 Balancing user requirements

80. According to *Recommendations for the 2020 Censuses of Population and Housing*, there are six widely accepted dimensions of statistical quality: relevance, accuracy, timeliness, accessibility, interpretability and coherence⁷.

81. With the proliferation of data sources including 'big data', advances in capacities and methods to exploit large data sets, and the increased value placed on data assets in a world in constant evolution, there is also 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.

82. In trying to meet the changing and increasing demands from data users, there is often a trade-off between timeliness, accuracy and comparability over time. The three 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 expense of timeliness as the data would then be released a longer time after the reference period.

⁷ (Chapter IV, Quality management, pp. 69–70).

83. For many NSOs, the opportunities with new data sources may create conflicts with changing user expectations. The emphasis on producing official statistics of measurably 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.

84. One challenge with increased timeliness may be linked to varied user needs, depending on the profile of the users. Policy users and many researchers, for example, may need to have very precise results for small areas or small groups, which usually require more time to assure accuracy. Users at a national level may prefer more timely data even if accuracy is not as high at the local level.

85. 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.

86. Future *Recommendations on censuses of population and housing* may need to provide direction in considering these potential trade-offs. This will entail formulating positions on the following issues:

- a) What is the optimum 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) The measurement of certain characteristics may move from 100 per cent of the population to a sample, potentially giving users 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. Perception management to communicate quality assurance methods would then become crucial.

9 Privacy, public acceptability and ethics

87. For most countries looking at transforming their population census design, successful transition to register-based censuses will depend on having access to a wider range of data sources. In many cases, NSOs already have access to what they consider to be key data sources for producing population counts. These are usually sources expected to have high coverage of the population, for example tax and social security records. However, for many countries, the need to continue producing census estimates for wider population characteristics – often at multivariate/small area level – requires the development of new microdata that have been constructed from multiple and diverse sources. Informing the public about the secondary use of data becomes relevant as this may not be covered by the informed consent obtained when the data were collected. Reputational risks always need to be considered when conducting censuses.

88. For a number of countries engaged in future census research, the focus has tended towards integrating data sources that are routinely collected by government departments. However, interest for some countries is extending to other forms of commercial data when there is unique potential for specific topics, for example mobile phone data for commuter flows. In such cases the balance between privacy and the usability of the data is of the utmost importance.

89. There is also a need to revisit the role of social surveys to support future censuses, with many countries still anticipating the need to collect information on topics that are not available from alternative sources, or to assess the coverage and quality of administrative data.

9.1 Privacy

90. A census model based on integrating multiple data sources brings new challenges to NSOs from a privacy perspective. Since the NSO is no longer responsible for end-to-end collection and processing of data, additional steps need to be put in place to ensure safeguarding of personal information. Business models may need to be redesigned across a range of functions, including data importing, data management, data access, record linkage, and disclosure control.

91. In some cases, onward sharing of data between suppliers and the NSO may not be able to proceed until certain conditions, such as data processing, data encryption, data architecture and data access, have been met by the NSO.

92. While it is important to recognize that the use of multiple sources elevates the risks to individuals' privacy, any radical changes to processes can have adverse impacts on the quality of statistical outputs. For example, encryption of data can result in errors in the record linkage process, and bias in the resulting outputs. Other attempts to preserve privacy by separating dataset attributes or reducing geographic granularity can undermine the potential to produce statistics to the required levels of detail and precision. The challenge for NSOs is finding the balance in preserving privacy while delivering products that still have the ability to meet users' needs.

9.2 Public acceptability

93. The majority of NSOs pursuing the goal of transitioning to a register-based census, or already conducting one, can rely on legislation enabling access to datasets to support the production of official statistics. However, members of the public are generally unaware of how their data are used when providing information about themselves during interactions with government services. As with traditional censuses, participation in future register-based censuses could be considered compulsory, as there would be limited scope for 'opt-out' provisions that entitle individuals to exclude their records from analysis. It is generally accepted that NSOs would need to resist any push for a requirement for informed consent from individuals to use their data in future administrative data-based censuses. Even a small percentage of individuals unwilling to provide their information for these purposes could result in very biased population statistics.

94. For those members of the public seeking assurances about the reasons for data collection and methods of processing of data, information is readily available during a traditional census collection. For register-based censuses, the uses of data are far removed from the original purposes of collection. NSOs need to consider the extent to which they will inform the public about the use of data sources compiled for official statistics. Statistics New Zealand, which is relatively advanced in the development of its Integrated Data Infrastructure (IDI) has made notable steps in promoting its vision to the public about the re-use of existing data for statistics and research⁸.

95. Depending on the level of sensitivity in the public debate around data sharing, NSOs' need to consider how actively they want to promote the uses of different sources of data in future censuses. Articulating the benefits is important to key stakeholders. However, it is also important to avoid any negative publicity about the proposed uses of data. An approach that is often favoured to try and understand potential concerns from the public is to undertake small-scale public acceptability research, using either focus groups or quantitative surveys. Studies such as these generally conclude that the public is supportive of data sharing once the benefits are understood, but that it is becoming increasingly complicated for participants to understand the processes involved when compiling statistics from multiple sources. The seemingly unlimited scope of potential research studies based on integrating data also makes it difficult for participants to consider all of the ethical considerations that are relevant in particular circumstances. For this reason, some NSOs are seeking to build support for new outputs by engaging closely with advocacy groups representing those under study.

9.3 Ethics

96. Alternative sources of data present new opportunities for NSOs. Some of these sources, however, also present more detailed ethical challenges. Data are generally made available to NSOs for statistical purposes only, with a clear statement that any resulting analysis will not be used to perform direct functions that are considered operational or administrative in

⁸ See <https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/> and the materials linked therein.

practice. In countries where register-based censuses have already been established, there are examples of NSOs using administrative data to improve the quality of survey collections. For example, telephone or field interviewers undertaking surveys are provided with the names of individuals listed on population registers and asked to verify with respondents which individuals still reside at the address.

97. The use of administrative data to support survey collections is sometimes referred to as 'dependent interviewing'. It has a particular benefit for measuring overcoverage in administrative data, which is a major challenge for countries that do not have a system of population registers in place. However, there are significant barriers to conducting dependent interviewing: information disclosure about previous residents; breaching data sharing agreements with suppliers; and reputational risk about practices being carried out by the NSO.

98. To consider propositions for new data uses, structures need to be put in place for ethical oversight and transparency. As an example, the UK Statistics Authority has established a Data Ethics Advisory Committee to advise the National Statistician on ethical access, use and sharing of public data for research and statistical purposes for the public good. Giving detailed consideration to each proposed innovative and novel use of data, from across ONS and the Government Statistical Service more widely, it provides independent advice to the National Statistician on the ethical appropriateness of these proposals. This work has included developing ethical principles and establishing a variety of transparent processes to assess proposed research against these principles:

- a) The data subject's identity (whether person or organization) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately;
- b) The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognized standards of integrity and quality;
- c) Data used and methods employed are consistent with legal requirements;
- d) The views of the public are considered in light of the data used and the perceived benefits of the research;
- e) The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public.

10 Conclusion

99. In the last two decades there has been an increasing pace of change in the ways that censuses of population and housing are conducted internationally, and more specifically in the UNECE region. As more and more countries explore and employ new methods, it can be anticipated that this pace will be maintained for the next round of censuses.

100. *Census Recommendations* have played an important role in guiding the development and implementation of census concepts and methodology. As these important transformations take place in national censuses, it is more important than ever that they be guided by a set of common principles and guidelines, to ensure this fundamental statistical exercise maintains its relevance and effectiveness for many more cycles to come. Under the overarching umbrella of the United Nations Fundamental Principles of Official Statistics⁹, which guide all official statistical endeavours, the *Census Recommendations* will continue to serve this guiding role. Their utility for future censuses beyond 2020 will be enhanced by revising them in light of the suggestions made in this analysis.

101. The topics covered in this paper suggest that the following issues should be revisited and potentially revised for the next edition of the *CES Recommendations*:

- a) How to ensure coherence in concepts and data when using multiple data sources;
- b) Essential features of a census;
- c) Definitions of population, usual (and possibly secondary) residence, and household;
- d) Measures of census coverage and quality of census results;
- e) Improving the use of geospatial data in census operations and products;
- f) The increased reliance on estimation models in the production of census outputs;
- g) The use of 'big data' in censuses;
- h) Better balancing user needs with expansion of data sources and statistical methods;
- i) Privacy, public acceptability and ethics.

⁹ <http://www.unece.org/stats/fps.html>