



# Economic and Social Council

Distr.: General  
26 June 2019

Original: English

---

## 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 4 of the provisional agenda

#### Research on the use of administrative data for censuses

## Innovative approaches planned for the 2021 Canadian census and beyond

Note by Statistics Canada\*

### *Summary*

The Canadian census is conducted every five years, the next one being planned for May 2021. Canada still uses a traditional approach, based primarily on a wave methodology generating high Internet response. Despite recent successes with this approach, Statistics Canada is researching the introduction of a combined census model for the future, relying first on information from statistical registers when possible, supplemented by traditional collection approaches where the availability of administrative sources (for the long form) or quality of the registers is not sufficient. This paper describes the research being done on the development of statistical registers to support the introduction of a combined approach, possibly for the 2026 cycle. It also briefly describes the collection methodologies planned for 2021, and more particularly how the work on the statistical registers can be leveraged, along with information from other sources, to support field follow-up operations in 2021.

---

\* Prepared by Sylvain Delisle, Marc Hamel and Sophie Lefebvre.



## I. Introduction

1. Statistics Canada is responsible under the Statistics Act for conducting the census of population every five years. The next census will be in 2021. The Canadian census programme includes a short questionnaire with approximately ten questions on basic demographics, and a longer questionnaire collecting more detailed socioeconomic information on a sample of 25 per cent of households. The programme relies on a combined approach, with data collected from households about individuals and about housing units at a specific point in time, and some information being obtained from administrative records. For example: income variables are obtained using income tax records, and information on category of immigrant is derived for the immigration longitudinal database maintained by Immigration, Refugee and Citizenship Canada.

2. Despite the success of the 2016 Census, with record high internet (68.4 per cent) and self-response (88.8 per cent) rates, it is expected that the census in Canada will face some challenges in the future. First, the current methodology is not likely to be sustainable over time, as is generally the case with any approach, and as society continues to evolve. The success of the current approach is highly dependent on a wave methodology, which relies on reaching households by mail.<sup>1</sup> Indeed, the use of an electronic questionnaire as the primary mode of response for the census requires the use of the postal system to provide a secure access code to every household. This mode of communication may gradually be abandoned by Canadians as digital communications become more and more the norm. Alternatives tested so far to deliver this information to respondents, such as e-mail and digital mail boxes, have not been successful. Second, Canadians have expressed concerns about the response burden imposed on them by different levels of government.<sup>2</sup> There is a certain expectation that before any direct collection takes place, the census programme should aim at leveraging data that are already collected by public, and even private organizations. Lastly, current international trends and research toward census-taking approaches based partly or wholly on administrative data cannot be ignored.

3. To respond to these challenges, Statistics Canada has embarked on a new research agenda called the “Census Program Transformation Project” (CPTP). The CPTP aims to make the census of population more user- and respondent-centric in order to be more responsive to emerging information needs, less burdensome for Canadians and also less costly. Past initiatives such as the 2016 Census Strategy Project<sup>3</sup> concluded that adopting an administrative census solely based on registers (as implemented in the Nordic countries, for example) would not be possible in Canada in the short or medium term, as the necessary conditions to implement such an approach are not in place.

4. This paper presents a brief description of the research conducted in relation to the expansion of a combined approach for future census cycles in Canada. It also presents the opportunities to leverage some of this work to support collection operations for the upcoming 2021 Census.

## II. Developing Statistical Registers

5. Statistics Canada has, since 2010, been examining how administrative data can be further leveraged as part of the Canadian census of population. Some information produced for the census has been obtained from administrative sources over the last three cycles (e.g., income and category of immigrants in 2016). Before a multi-mode approach was introduced, the “mode effect” of the different modes or sources of providing responses to the census was studied, and the quality of information was found not to have been significantly altered.<sup>4</sup> In fact, responses from administrative sources and from the online questionnaires are generally deemed to be better quality.

---

<sup>1</sup> Hamel, 2012

<sup>2</sup> Statistics Canada, 2019

<sup>3</sup> Howatson-Leo, Trépanier, 2011

<sup>4</sup> Grondin, Sun, 2008

6. The administrative census as implemented, for example, in the Nordic countries (Denmark in 1981, Finland in 1990, and Norway and Sweden in 2011) or the Netherlands, is often cited as the ultimate way of leveraging administrative data or registers to provide population and dwelling counts. This model, however, is not appropriate in the Canadian context since Canada does not have the fundamental elements in place: a unified system of administrative registers and a personal identification number for every resident.

7. The initial research on the use of administrative sources in the census context focused on the construction of a statistical population (Canadian Statistical Demographic Database (CSDD)) by compiling information from various administrative sources, primarily tax and immigration data, and information from birth and death registers. The initial promising results from the research on the CSDD then led Statistics Canada to undertake the development of a Statistical Population Register (SPR) and transform the Address Register (AR) into a Statistical Building Register (SBgR). These registers are statistical registers, meaning that they are being constructed from various administrative data sources and will be used only to support Statistics Canada's mandate to produce protected de-identified statistical outputs. The SPR will provide a continuously updated list of the people living in Canada at any given point in time, along with basic demographic characteristics such as age, sex at birth and place of residence.

8. The SBgR will be an evolution of Statistics Canada's Address Register and will provide an accurate list of all buildings in Canada, including collective dwellings, occupied and non-occupied private dwellings, and non-residential buildings.

9. The SPR and the SBgR coupled with the existing Business Register (BR) and a potential Activity Register (AR) will provide an infrastructure of core interconnected statistical registers that will be available to all Statistics Canada's programmes. For instance, the SPR could be used by the population estimates programme for their intercensal estimates or as a sampling frame for the social statistics programmes.

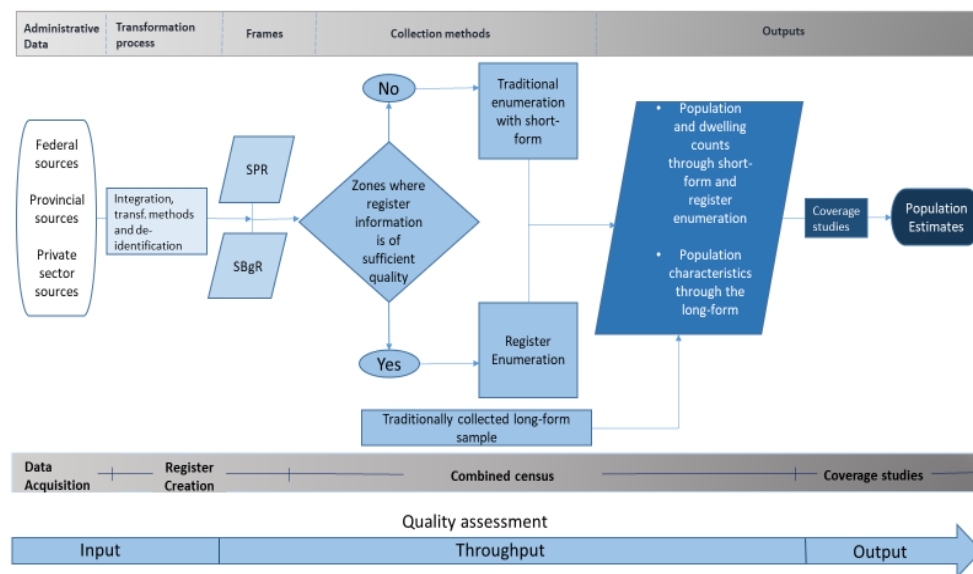
### **III. Expanding the Combined Approach for Future Censuses**

10. Countries such as the United Kingdom or New Zealand are investigating moving to a census approach based partly or totally on administrative data even without a robust unified system of governmental registers and personal identification numbers. Since 2016, Statistics Canada has been exploring the use of the information stored in an eventual SPR as a possible new mode of response to the census of population.

11. The main conclusion of the research conducted to date is that the raw administrative data currently in Statistics Canada's possession show great potential for providing good coverage of the Canadian population and could eventually be used to produce population counts. More work is required, as the research also demonstrates that sub-provincial coverage must be enhanced through improvements of current administrative files or acquisition of new ones. Despite this, information in the statistical registers could most likely be deemed of sufficient quality in many areas to be used in the census programme in the foreseeable future. This could be realized via an expansion of the current combined census approach.

12. An expanded combined census in Canada would integrate statistical register information with traditional collection. In a nutshell, this implies that the census frame of Canadian dwellings would be divided into two parts, using various characteristics (register coverage and accuracy, dwelling type, geography, operational considerations, costs, etc.). In one part, census outputs would be produced using administrative data of sufficient quality from the SPR and SBgR instead of direct enumeration. Thus, the SPR would be used to provide the population count with basic demographic information such as date of birth, sex at birth and geographical location. Traditional enumeration with the short-form questionnaire would be used for the second part, where the information in the registers is not deemed to be of sufficient quality or precision. To complement this approach, a representative sample of dwellings would be still be selected to provide detailed characteristics through the long-form questionnaire, as conducted in the current model. Figure I shows the various steps from the administrative data sources to the population estimates.

Figure I  
From administrative data sources to the population estimates through a combined census



13. Statistics Canada is presently conducting several simulations of this approach, comparing early iterations of the SPR to the 2016 census database and to results from the Demographic Estimates Programme, both to validate the quality of the statistical registers, and to determine the optimum use of these registers in a combined approach. Similar simulations will be conducted in parallel to the 2021 Census. The decision about the potential implementation of an expanded combined approach in 2026 will be derived from this research.

#### IV. Using the Statistical Registers for the 2021 Census

14. For the 2021 Census of Population Programme, Statistics Canada will largely reuse and repeat the approaches successfully implemented for the 2016 programme. However, opportunities for enhancements of the existing approach, which is focusing on the maximization of self-response, are now more limited. The largest potential for new efficiencies of the existing census approach is to take advantage of readily available information to reduce the need for follow-up on non-responding dwellings. The use of information generated by the collection operations as well as data from the new statistical registers, and new administrative data sources available at Statistics Canada could help decrease costs and burden while maintaining the high quality of the census. The strategy proposed for 2021 has three main streams:

- (a) Reduce over-coverage of the census frame: reducing the number of cancelled dwellings to 3.5 per cent from 4.5 per cent in 2016 via validation with signals from administrative sources;
- (b) Predict dwelling occupancy status: identify cancelled, unoccupied and occupied dwellings for non-responding cases;
- (c) Stop the follow-up of some non-responding cases and use data from the SPR as a proxy response.

15. The reduction of over-coverage will be largely realized through matches to various administrative files. The census has been using the Address Register (AR) as the frame for several census cycles. The majority of updates to the AR are done using administrative sources. The main file used for this purpose is the Points of Call file maintained by Canada Post. Less than 20 per cent of addresses are validated through an in-office process, or via field listing.

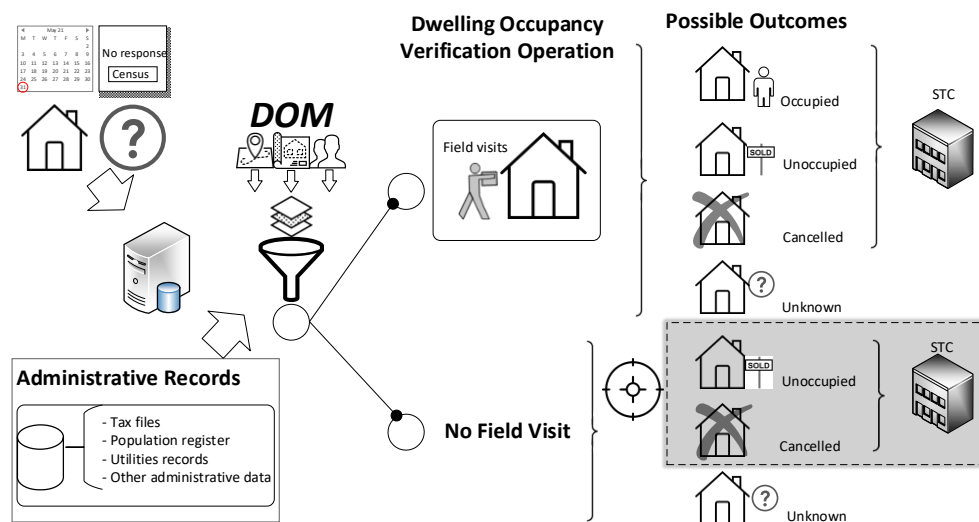
16. Historically, a new post-census iteration of the AR would include records cancelled for the first time during the previous census. These dwellings would then be part of the frame for the next census, on the anticipation that they may be valid dwellings at that time. This approach was changed with the first update of the AR post-2016 census. This update included only the valid private dwellings from the 2016 Census. Cancelled dwellings in 2016 will be rebirthed only if they show up on administrative files used to update the AR. A method is also being developed to identify duplicate records on the file pointing to the same dwelling. Other internal methods have been developed to fine-tune matching and reconciliation processes to minimize address versioning.

17. One of the current collection approaches that is part of the census process will be improved for 2021 by using new administrative data sources and by expanding its objectives. The Dwelling Occupancy Verification (DOV) operation, aiming at identifying vacant and cancelled dwellings as close to census day as possible, will rely on a more precise sampling approach based on new administrative data sources.

18. The Dwelling Occupancy Model (DOM) is being developed to predict the occupancy status of each dwelling. Based on information from various administrative sources such as tax files, utilities data and the statistical registers, each dwelling will be assigned a predicted probability of being occupied, unoccupied and cancelled prior to field operations. Figure II illustrates this process. Dwellings with high probabilities of being unoccupied or cancelled (i.e. lower probability of being occupied) will be in scope for the DOV sample selection. The DOV sample is also selected on the basis of operational criteria, such as clustering of dwellings to control collection costs. During the DOV operation the occupancy status is verified by field staff.

Figure II

**Process derived from Dwelling Occupancy Mode**



19. Depending on the performance of the DOM, new treatments could be applied prior to defining the DOV sample, removing field visits completely or reducing field visits at the time of Non-Response Follow-Up (NRFU) operation:

- Dwellings with a high probability of being unoccupied and low probabilities of being cancelled or occupied would be classified as unoccupied.
- Dwellings with a high probability of being cancelled and low probabilities of being unoccupied or occupied would be classified as cancelled.

20. The third segment of the strategy will allow follow-up to be stopped for some non-responding dwellings during the NRFU operation, using SPR data as a proxy response. The current strategy in Canada attempts to obtain a minimum response level of 98 per cent for each of approximately 45,000 collection units defined for the NRFU operations. In this context, the SPR could be used for a predetermined maximum number of dwellings per collection unit in areas where high effort has already been expended to secure a response

from non-responding dwellings without success, or even to increase response in collection unit where 98 per cent response or better has already been attained.

21. Among the criteria to be used to identify these dwellings which would be provided by the SPR are: the high probability of the dwelling being occupied from the DOM; the dwelling not being part of the long-form questionnaire sample; the relative quality of the SPR records; and the set limit of such replacement by collection unit. Limits have to be imposed on the use of this type of response for entire households to avoid impacting the variance of the census coverage studies results, which are themselves based on administrative sources.

## V. Conclusion

22. Statistics Canada has been progressing rapidly in its research of administrative alternatives to an enumerated census. On the other hand, any implementation of such approaches on a large scale will need to be considered very carefully. Population counts are the basis of the electoral system in Canada. Also, billions of dollars in annual transfers from the federal government to provinces and territories, and by extension to municipalities, are based on census results. Quality is going to have to be a key characteristic of any decision to move forward with a much expanded combined approach.

23. The implementation of a combined approach will also involve certain trade-offs. For example, some of the census outputs currently produced from the short form (100 per cent of the population), such as marital status and family formation, would have to be moved to the long form (25 per cent sample) because this information is generally not available in administrative files. This possibly does not represent a major issue, as the value of these variables is much greater when combined with other variables on the long form. Extensive consultations will be necessary with data users and stakeholders, however, before the combined approach can be expanded in the future.

## References

- Hamel, Internet data collection in the Canadian Census of Population, UNECE Expert Group Meeting on Censuses Using Registers, Geneva, 2012
- Statistics Canada (2019). Modernization: a key to Statistics Canada's effort to reduce response burden, [https://www.statcan.gc.ca/eng/about/smr09/smr09\\_100](https://www.statcan.gc.ca/eng/about/smr09/smr09_100)
- L. Howatson-Leo, and J. Trépanier, (2011). Final Report on the 2016 Census Options: Proposed Content Determination Framework and Methodology Options, Statistics Canada. [http://www12.statcan.gc.ca/census-recensement/fc-rf/reports-rapports/r2\\_index-eng.cfm](http://www12.statcan.gc.ca/census-recensement/fc-rf/reports-rapports/r2_index-eng.cfm)
- C. Grondin, L. Sun, 2006 Census Internet Mode Effect Study, Joint Statistical Meeting of the American Statistical Association, August 2008.
-