**UNITED NATIONS** **E**

# Economic and Social Council

Distr.
General

ECE/CES/2006/26
30 March 2006

Original: ENGLISH

**ECONOMIC COMMISSION FOR EUROPE** **STATISTICAL COMMISSION**

CONFERENCE OF EUROPEAN STATISTICIANS

Fifty-fourth plenary session
Paris, 13-15 June 2006
Item 6 of the provisional agenda

SEMINAR ON POPULATION AND HOUSING CENSUSES
SESSION III

## The important role of research, evaluation and continuous improvement of census taking in Canada[1]

Submitted by Statistics Canada

1. Canada's 25th Census since confederation in 1867 will be conducted in May 2006, introducing substantial changes in the collection and processing methodology. The new approach addresses a number of important issues that have built up over time, relying increasingly on computing technology and centralized processes. Canadians have become increasingly conscious of privacy issues and continue to expect total confidentiality protection of census responses. The issue of local enumerators collecting sensitive information had to be addressed. A critical mass of Canadians also expects to be offered an on-line response option, albeit a very secure one. Users demand detailed and timely public good data, available free of charge over the internet, while central funding agencies expect more cost-effective ways of conducting the census. Both the content and the manner in which a census is conducted have to continuously evolve and meet the demands of Canadians if it is to remain relevant and effective. This paper describes the important role that research, evaluation and continuous improvement play in the Canadian Census, focusing on the efforts that led to the innovative approaches implemented for the 2006 Census.

[1] This paper has been prepared at the invitation of the secretariat.

## I. GENERAL DRIVERS FOR CHANGE

2. A national census is often thought of as a portrait or snapshot of the country, reflecting the demographic, social and economic make-up of its people at a given point in time. But the target is a moving one as society evolves at an ever-increasing pace. A successful census operation takes notice of and acts upon the drivers of change. These drivers are categorized below by three sources of change, though it is recognized that these distinctions are not mutually exclusive nor necessarily comprehensive. Moreover, some countries may identify others drivers of change reflecting changing events or considerations in their region.

3. First, a census must remain relevant by adapting the content of its questionnaire to the evolving demographic and socio-economic characteristics of the country and its regions. The content can be influenced by emerging social policy issues such as the demands of an aging population; the role of continuous learning and the need to upgrade industry and occupation classifications. The challenge is to find the right balance between introducing change to census content and maintaining historical continuity to enable analysis of trends over time.

4. A second driver for change concerns emerging technologies associated with collecting, processing and disseminating statistical information. For example, many countries have responded with the introduction of automation technologies such as scanning and intelligent character recognition. This category of drivers also covers the need to respond to the increased sophistication of data users who expect more detailed geographic and socio-economic information in formats compatible to the latest data and cartographic software.

5. Finally, a census must take notice of respondent attitudes towards the census. In many countries there is an increasing hesitancy on the part of people to respond to surveys including the census. Declining response rates negatively impact both the significant costs of non-response follow-up as well as the data quality of the end results. It is important to understand the respondents' sources of concern so that, where possible, changes can be made to address them.

## II. BACKGROUND ON THE CANADIAN CENSUS

6. On May 16, 2006, Statistics Canada conducted its latest quinquennial census. Uniquely identified questionnaires were delivered to over 13 million households. For the first time, about two thirds of households received their questionnaires by mail while about one third of questionnaires were dropped off by enumerators. In both cases, an adult in the household was asked to complete the questionnaire for themselves and for members of their household and return it on-line or by mail. A small minority of about 2% of households were enumerated using the canvasser method, where an enumerator visited the household and completed the questionnaire by interview. This method was used in remote and northern areas of the country and on most Indian reserves. It was also used for transient residents in the downtown areas of large urban centers.

7. Most households (80%) received a short census questionnaire which contained eight questions on basic topics such as relationship to the reference person, age, sex, marital status, and mother tongue. One in five households (20%) received a long questionnaire which contained the eight questions from the short form plus 53 additional questions on a variety of topics such

as education, ethnicity, mobility, income and employment.

## III. MAIL OUT/MAIL BACK AND INTERNET RESPONSE CENTRAL TO CHANGES TO CANADIAN 2006 CENSUS METHODOLOGY

8. Prior to the 2006 Census collection process in Canada was a highly decentralized, manual collection operation involving a large and geographically dispersed workforce. Essentially unchanged since it was introduced in the 1971 Census, although robust, it became increasingly evident to Statistics Canada that a change from these traditional approaches was required.

### A. Drivers for change

9. One of the original drivers forcing consideration for change was the issue of confidentiality (in particular concerns with local enumerators) and security of personal information. The issue of confidentiality is of increasing concern to Canadians, whether it be with regard to public or private institutions. While Statistics Canada offers the assurance of both confidentiality and security guaranteed under federal legislation, complaints from respondents continue to be a source of concern. Under the traditional collection methodology, completed questionnaires are returned to the local enumerator for completeness checks and follow-up when necessary. These enumerators are generally locally engaged staff who, hired for their knowledge of the local geography, often are known to the respondents that they are following-up with.

10. A second driver was the federal revenue agency known as Canada Customs and Revenue Agency (CCRA). Since 1981 Statistics Canada had relied on its staff and facilities to key enter data from the census questionnaires. CCRA is moving away from its own key data entry operations as it aggressively promotes electronic filing of tax returns. They now receive half of the tax returns electronically, either directly from the tax filer or through companies that offer on line tax preparation services. This percentage increases each year. The resulting decrease in CCRA's keying and capture capability required Statistics Canada to examine alternative data capture options.

11. A third driver for change was the initiative to have all transactions that Canadians make with the federal government available on line. Having already made significant inroads in the dissemination of census information to Canadians via the Internet the next step was to provide Canadians with an option to fill out the census questionnaire on line in a secure and efficient application that is integrated with census processing and field operations.

### B. New collection and processing methodologies

12. The methodology for the 2006 Census that addresses these drivers for change has the following major elements:

(a) enhancement of a comprehensive address register containing specific civic addresses for a majority of the dwellings in Canada where questionnaires can be mailed-out instead of hand-delivered by enumerators. The same file subsequently forms the core of the "Master Control System", a centralized dwelling frame, which facilitates effective operations control at the dwelling level;

(b) the return, by mail, of all questionnaires to one central data processing center, with

automated editing functions and the follow-up for missing information in a prioritized manner from three call centers using Computer Assisted Telephone Interviewing (CATI). This approach eliminates the need for local enumerators to log receipt, edit and conduct follow-up for missing information, except for total non-response or refusal follow-up, significantly reducing the local enumerator privacy issues;

(c) the use of Intelligent Character Recognition software to capture questionnaire data, thereby virtually eliminating manual key entry. This addresses the need for timely receipt, capture, and processing of census returns. Automated registration of questionnaires upon arrival at the central data processing centre facilitates communicating non-response workloads at the dwelling level to field operations on a daily basis;

(d) the implementation of an integrated secure on-line census application that provides an efficient alternative to the paper questionnaire, potentially providing future savings in paper handling, postage and processing. With built-in edits, the same provides the opportunity for better quality data with reduced follow-up for incomplete information.

C. Long term planning, research and testing key

13. The changes made to the collection and processing methodology for the 2006 Census follow a clear path of planning, research, testing, evaluation and continuous improvement that spans at least a decade. As part of the 1996 Census, a “centralized edit test” was conducted in the region surrounding Ottawa, implicating some 325,000 dwellings. Approximately 200,000 dwellings were mailed questionnaires based on the address register list, their questionnaires returned to one location, with manual receipt, editing and centralized follow-up for missing information. The remaining 125,000 dwellings were hand-delivered a questionnaire but the receipt and follow-up procedures were the same as for the mail-out area. Although not automated, the process provided important lessons that proved invaluable in developing the 2006 process.

14. The address register was developed prior to the 1991 Census where it was first introduced as a post drop-off check to ensure that known dwellings through administrative records were not missed by the enumerator. Subsequent censuses have continued to use the address register in the same way until 2006, when it was verified by field staff in a large listing exercise across the country. The verified and expanded Address Register now serves as the core for the census dwelling frame which is used for pre-printing addresses on questionnaires, controlling and monitoring non-response follow-up and overall operations control.

15. In 1996, an automated system to facilitate enumeration at the dwelling level in urban core areas was developed, allowing for assignments to be generated daily and field operations to be controlled at the questionnaire level. The system was expanded to the major urban centers in 2001 and served as a significant learning tool for the Non-Response control systems being used by 36 local census offices, fed daily by the receipts in the Data Processing Center, for the 2006 Census.

16. The 2001 Census introduced the first on-line census application in two test areas of the country, providing much needed experience as well as an alternative to the paper questionnaire for those insisting on an Internet based questionnaire. Although the security and the application were robust, they involved a significant download to each respondent’s computer. It became

evident that a more user-friendly, but still very secure, interactive on-line application needed to be offered in 2006.

17. As part of the 2001 Census, a sample of 300,000 questionnaires was rigorously evaluated in an "Edit and Follow-up Study (EFUS)" to analyze the impacts of editing and follow-up on outgoing data quality. The study concluded that totally eliminating editing for incomplete questionnaires and using strictly imputed data to fill the gaps would bias these records to a point where it would negatively impact the quality of census results. The study guided the development of automated edit rules in place for 2006, along with a prioritization methodology for following-up on incomplete records using CATI.

18. The 2001 Census questionnaires were scanned and imaged using external contractor furnished hardware and software, with the subsequent questionnaire images substituting paper questionnaires for various coding and processing steps that follow data capture. The experience provided invaluable experience and lessons learned in setting up and operating a large processing center, including scanning and image handling.

19. In 2004, following extensive questionnaire format testing, Statistics Canada successfully conducted a sizable census test involving three major areas comprising 300,000 dwellings and 20,000 farms. The integrated systems and centralized processes were exercised thoroughly in advance of scaling-up for 2006.

20. In addition to the various research and testing processes, sharing international experiences and lessons learned with our international census colleges has also played a crucial part in the development of the overall 2006 Census methodology.

## IV. EXPLORING A PAPERLESS CENSUS

21. The introduction of an integrated secure on-line census application has contributed to a more secure and efficient process for collecting and processing respondents' information. Data capture is no longer required for questionnaires filled out this way and built-in edits provide an opportunity for better quality data.

### A. Driver for change

22. While the development of the on-line option is expensive, some costs are offset by savings in the reduced postage and paper handling. Moreover, as some edits can be built directly into the questionnaire application, less follow-up is required as questionnaires are generally more complete and accurate. Successfully promoting and achieving a high on-line response rate will increase these savings. It is expected that the use of the Internet in Canada will grow with future censuses and the cost reductions may be considerable.

23. The potential for even greater savings however, is constrained by the current methodology of mailing out questionnaires to all Canadians. Presently Canadians are provided with a paper questionnaire with an option to use the access code included with their package to respond on-line. Thus, considerable resources must still be expended to print, handle and mail out the census forms.

B. Initial steps in the planning, research and testing of a paperless census

24. Statistics Canada is exploring the concept of a paperless census, where Canadians would be provided electronically with an access code to respond on-line, with an option to request a paper questionnaire.

25. Though this approach is attractive for the many advantages it would offer, many issues immediately come to mind. For example, not all Canadians have access to the Internet. Of those who do, many might not wish to be contacted this way. Rather, the current multi-modal census collections model would have to be expanded to allow identification of households with a significant expectation of providing an on-line response.

26. As with the mail out/mail back methodology, changes of this scope with major implications for many areas of the census (respondents, operations and content) can only be carried out through a comprehensive research program spanning many censuses.

27. The 2006 Census has begun this process with a pilot study to test a method aimed at promoting Internet response by asking households to complete their questionnaires on-line, without prior receipt of paper questionnaire packages. Taking advantage of data identified a priori, from the 2001 Census and the 2004 Census Test, areas that include a significant number of dwellings likely to answer the census by Internet were identified.

28. The selected households receive a letter, instead of a paper questionnaire, with an internet access code and instructions encouraging them to fill in their census questionnaire on-line. Respondents not able or wanting to respond by internet could call a special telephone number, provided with the letter, for information or to request that a paper questionnaire be mailed to them. Households that did not respond are contacted as part of the regular non-response follow-up activities.

29. Statistics Canada will shortly conduct a follow-up survey with a sample of households from this study to understand why some respondents did not use the on-line option. The effectiveness of the letter and telephone support will also be examined.

30. This study should provide preliminary insights into the effectiveness of methods used to encourage the on-line option without having to print and mail out the census questionnaire. Over the coming years, further research and testing will build on this initial study.

## V. COMBINING CENSUS AND ADMINISTRATIVE FILES TO IMPROVE MEASURES OF INCOME

31. Significant changes to the content of the census questionnaire are also the result of extensive research, evaluation and testing. The change to the 2006 Census module on income is a case in point.

32. For the first time, the 2006 Census long form offered to Canadians the option of giving their permission to link to their income tax records for the previous year instead of having to answer the income questions.

### A. Drivers for change

33. The main drivers for inclusion of the tax permission option on the Census questionnaire are reduced respondent burden, increased data quality and privacy advantages.

34. The tax permission option in fact offers a significant decrease in respondent burden as answering the census questions accurately requires the respondent to consult their personal files. Respondents who give permission to use the data in their income tax files for the previous year do not need to answer the thirteen detailed questions on sources of income and income tax paid.

35. Studies have shown that the income data retrieved from the tax files more accurately reflects the respondent's income, compared with census self-reported figures. When respondents report their income figures on the census there is a reasonable likelihood that they have reported rounded figures or an amount from memory, as opposed to the exact amount.

36. Although the 2006 Census no longer uses enumerators to review census questionnaires for completeness, privacy of census information and of income data in particular may continue be of concern to respondents. Some respondents may not wish other members of their household to see their income. Where local enumerators are still used to deliver the forms there may also be a perception that someone they know personally will see this information.

### B. 2006 Census income module result of much research and testing

37. The methodology for a tax permission option involves a number of steps. These include offering and obtaining permission to use information from the respondent's income tax file, establishing a linkage between the census and income tax file using probabilistic matching techniques, integrating the linked income tax information with the general census processing data base and accounting for situations where the linkage was not successful.

38. Census variables that are planned to be used for the linkage include name, address, sex, date of birth, marital status, disability, citizenship, current immigration status, postal code one year ago and recent labour force activity and class of worker. Personal identification numbers from the administrative file are not used for the matching. Once the matching process has been completed the data file used for the linkage containing information on name and address is destroyed.

39. The feasibility and outcome for each of these steps has been the subject of much research and testing over the last twenty years.

40. Statistics Canada first looked to other areas of the Agency where work in this area had been conducted, including:

(a) the Survey of Labour and Income Dynamics (SLID) has been statistically matching their consenting records to the income tax file since 1994. This is a smaller scale linkage but with many of the same matching variables;

(b) the Census Reverse Record Check used to measure net census undercoverage has linked Census data to income tax records in the past to obtain an updated address and other information to assist in tracing persons selected for the sample;

(c) the Health Modelling and Analysis Group of Statistics Canada have also conducted record linkage activities using a derivative of the 1990-1991 income tax files and the 1991 Census in order to link to the Canadian Mortality Database.

41. The first testing for the tax permission option was conducted with the 1996 Census and the 1995 income tax file, using individual characteristics without the name of respondent. The studies were repeated with the 2001 Census and the 2004 Census Test. They examined the accuracy of the linkage which is the degree of confidence that the records linked between the Census file and the income tax files represent the same individual or an individual with similar income.

42. While use of name as a matching variable may be more sensitive for its privacy implications, including name in the 2004 study has shown it has an important impact on the quantity and quality of links obtained. The matching rates in the 2004 Census Test increased from 76% to 91% with the addition of the name components while lowering the error rate to a third of its original value.

43. The 2004 Census Test also included for the first time in a census environment the question offering respondents the option of allowing Statistics Canada to use their income tax information in lieu of their responding to the set of income questions on the census form. In the test, about 80% of respondents granted permission to use their income tax information. This compares to the permission rate obtained by the Survey of Labour an Income Dynamics. While this Survey and the 2004 Census Test were voluntary surveys, it is expected that the 2006 Census will yield similar results.

## VI. CONCLUDING REMARKS

44. Using a number of pilot tests and phasing changes over the 2001 and 2006 Census cycles, Statistics Canada has moved away from a decentralized, manually intensive collection and data entry operation to a more centralized and automated approach. As this is a continuous process, research continues through 2006 and beyond to examine ways to improve how census information is collected, processed and disseminated.

45. In this process of continual improvement it is important to take notice of drivers for change. They may come from many sources including the socio-economic environment of the country, the respondents themselves, emerging technologies and internal operational pressures.

46. Successful implementation of change is also the result of intense teamwork by dedicated and professional census staff. What makes the census a success is how the different workers come together. Teamwork is the most critical ingredient for success, whether when brainstorming about new possibilities, when testing new census methodologies, or in times of stress. Management must foster a climate of teamwork and risk tolerance if it wants the innovative ideas that feed research and testing to surface and sustain a culture of constant improvement.

* * * * *