

## Article

Symposium 2008:  
Data Collection: Challenges, Achievements and New Directions

### **The Internet: A New Collection Method for the Census**

by Anne-Marie Côté, Danielle Laroche

2009



## **The Internet: A New Collection Method for the Census**

Anne-Marie Côté, Danielle Laroche<sup>1</sup>

### **Abstract**

Major changes were made to the data collection process for the 2006 Census. One of those changes was the Internet response option, which was offered to all private households in Canada. Nearly one in five households chose to complete and return the questionnaire on-line. In addition, a new method of promoting Internet response was tested via the Internet Response Promotion (IRP) Study. The new approach proved very effective at increasing the on-line response rate. Planning for the 2011 Census, which is under way, calls for the use of a wave collection strategy, and wave 1 would be the IRP method. This paper provides an overview of Internet data collection in the 2006 Census – evaluations, results, lessons learned – and the methodology that will be used in the next census in 2011.

Key Words: Census, Internet, Collection method.

### **1. Introduction**

Using the Internet for data collection is not a new phenomenon for Statistics Canada. Indeed, some business surveys are already using this collection method. Over the last few years, Statistics Canada has been trying to incorporate it into its social surveys. With its 2001 feasibility test, the Census was the first social survey to exploit the new methodology. Then, in the last census on May 16, 2006, an Internet response option was offered for the first time to all private households in Canada. The introduction of the new collection method presented many methodological challenges in a number of areas, including questionnaire design, data security and confidentiality, data quality, mode effects, and data processing and analysis.

### **2. Overview of the Census**

The Census of Canada is conducted every five years. All Canadians are required to take part. Citizens are enumerated at their principal residence, regardless of whether they are there on Census Day. In the 2006 Census, most of the population was counted through self-enumeration. According to the Census, on May 16, 2006, there were 12.4 million occupied private dwellings and 31.6 million people in Canada.

Most of the data were collected using two types of questionnaires. A short questionnaire containing eight questions was distributed to 80% of Canadian dwellings. A long questionnaire containing about 60 questions, including the eight in the short questionnaire, was distributed to 20% of all dwellings. Each questionnaire was available in both official languages.

In the 2006 Census, two thirds of Canadian dwellings received a paper questionnaire by mail; for the remaining third, a questionnaire was dropped off by an enumerator. Each questionnaire included the URL of the census Web site and a unique access code.

Respondents thus had the option of completing the questionnaire either on the Internet or on paper; in the latter case they had to mail it back after completing it. They could also call the Census Help Line (CHL) for assistance in completing the questionnaire. Non-respondents were contacted in non-response follow-up (NRFU) operations.

---

<sup>1</sup>Anne-Marie Côté, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A 0T6; Danielle Laroche, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A 0T6

### 3. Characteristics of the Internet option

The Web application was available 24 hours a day, seven days a week. For security, Statistics Canada used session encryption with automated login, which provides two-way encryption at a significantly more secure level than what is normally used on Web sites. Every household could complete its questionnaire – long or short – on-line with absolute security. In addition, users who received the long questionnaire could save it so that they could complete it in more than one session and from different locations while preserving and protecting their personal information.

To facilitate the integration of data received on the Internet and maximize their quality, the electronic questionnaire was made similar to the paper questionnaire, at least with respect to the wording of the questions, the instructions and the presentation of response choices. Wherever possible, the application complied with the federal government's standards and guidelines for Web site presentation. In addition, Web standards such as radio buttons, check boxes and drop-down lists were used. To facilitate navigation in the questionnaire and reduce the response burden, automated skips were used so that each respondent was asked only the questions that applied to him or her.

The electronic questionnaire used an architecture with a number of interactive pages. With that architecture, the questionnaire was presented screen by screen. Each screen displayed a question or group of questions on the same subject. Whenever a respondent clicked the "Continue" button at the bottom of a screen, the application would check the data. If any of the data were erroneous, the system would display one of three types of validation messages: if the respondent left out a question or only answered part of a question, the application would display a non-response or partial response message; if the response was not in the proper format, the application would display an invalid response message. This function applied mainly to numerical fields and postal codes. In addition, amount verification messages could appear for questions about income and dwelling expenses. Any of the messages could be ignored by clicking the "Continue" button a second time.

## 4. Results

### 4.1 Response

The distribution of questionnaires received for occupied private dwellings by response mode is presented in the table below. The figures are from the 2006 Census Master Control System (MCS).

**Table 4.1-1**  
**Questionnaires received by response mode**

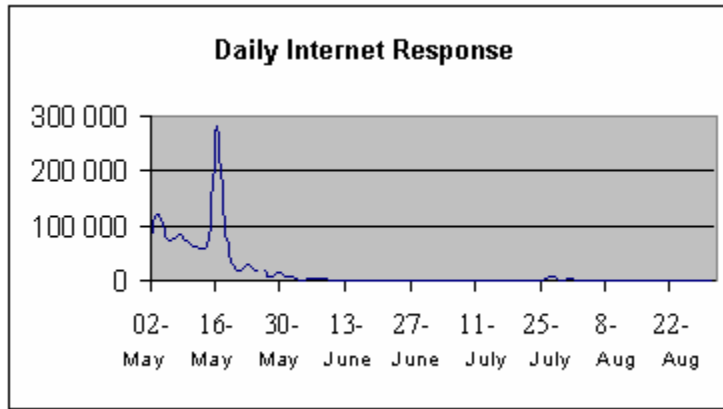
Response mode	Response %
Internet	18.3
Paper	63.1
Census Help Line	2.3
Non-response follow-up	16.3

One of the biggest successes of the 2006 Census is certainly its on-line response rate of 18.3%. In fact, Canada has the highest Internet response rate in the world for a census. The proportion of paper questionnaires received by mail was 63.1%. Overall, 81.4% of the responses were obtained by self-enumeration. This rate is similar to that obtained for the 2001 Census. The proportion of responses received through the Census Help Line (CHL) was 2.3%, while the proportion for non-response follow-up (NRFU) was 16.3%.

## 4.2 Reception of electronic questionnaires

The figure below shows the number of electronic questionnaires received per day throughout the 2006 Census collection period.

**Figure 4.2-1**  
**Daily volume of electronic questionnaires received**



In early May, we started receiving large numbers of questionnaires. An average of just over 87,000 Internet responses arrived daily in the first two weeks following mailing of the paper questionnaire. Volume peaked on Census Day, when we received more than 280,000 electronic questionnaires. From then until the end of the collection period, we received fewer and fewer questionnaires. The on-line response time lag is very short. Ten days after the Census, we had already received 86.3% of the electronic questionnaires.

## 4.3 Completeness of the questionnaires

The completeness of the questionnaires is a good indicator of the initial quality of the data received. Any questionnaire that does not meet a predetermined level of completeness fails the edit stage. The failure rate for the two types of questionnaires by response mode is presented in the table below.

**Table 4.3-1**  
**Edit failure rate by type of questionnaire and response mode**

Response mode	Short (2A) %	Long (2B) %
Internet	2.5	5.7
Paper	5.6	39.1

The results show that questionnaires completed on-line are more complete than paper questionnaires. For short questionnaires, only 2.5% of electronic questionnaires failed edit, compared with 5.6% of paper questionnaires. The difference was much greater for long questionnaires: 5.7% of electronic questionnaires failed edit, compared with 39.1% of paper questionnaires.

The differences are partly attributable to the respondents' profile and the features of the Web application (radio buttons, automated skips, help feature and validation messages). On the one hand, most Internet respondents are highly educated. In addition, validation messages are very effective at obtaining responses to questions overlooked by the respondent or pointing out inadvertent errors. However, in the qualitative tests and follow-up surveys, we found that respondents thought they had no choice but to provide a response for each question if they wanted to continue completing the electronic questionnaire. Similarly, a majority of those who received validation messages did not know that they could ignore them. As a result, they provided a response no matter what. The responses we

obtained may thus be different from the ones we would have received if we had used some other collection method. These differences are referred to as mode effects.

#### **4.4 Problems identified**

Since the option was first tested in the 2001 Census, Statistics Canada has been trying to find methods to assess the quality of its data and ways to compare them with data received through other response modes. During preparations for the 2006 Census, qualitative studies were carried out in-house and in respondents' homes to gain a clearer understanding of the Internet results. User-friendliness tests, expert reviews of the electronic questionnaires, and follow-up surveys of households were also conducted. Through these studies and surveys, we were able to gain a better understanding of respondents' behaviour toward this new response mode, correct and improve the electronic questionnaires and develop potential methods of identifying mode effects. Following the 2006 Census, a number of analyses were carried out to measure the impact of the Internet response mode compared with the traditional paper response mode.

Despite all efforts made to obtain high-quality data, evaluations following the 2006 Census identified some flaws that could easily have been corrected. One such flaw involved the drop-down menus in certain questions. When respondents selected a response using their mouse, they then had to click outside the field. If they did not, their response could change without their knowledge if they used the mouse wheel to scroll to the next question.

Difficulties also arose in some numerical fields, including hours worked. Respondents who did not work in the week preceding the Census had to check a box labelled "None". The box was located below the numerical field. However, some people did not see it and entered 0 in the numerical field instead of checking the box. As a result, they received a validation message, since valid responses to the question had to be between 1 and 168. Because they felt they had to answer the question, they entered 1 for hours worked.

In addition, problems occurred in some fields that required a postal code, particularly those associated with mobility. Having probably forgotten the postal code for their old address, some respondents entered the code for their current address instead of leaving the field blank, as they would have done if they had completed a paper questionnaire. As noted earlier, many respondents thought they absolutely had to provide a response for each question before they could go on to the next one.

These problems have been easily corrected through changes to the Web application for the next census. First, the use of drop-down menus for response choices has been confined to selected questions; second, the numerical field for hours worked now accepts 0; and third, the non-response messages associated with postal codes have been disabled.

### **5. Lessons learned**

There were a number of reasons for the changes in the Census's traditional collection method. Among the most important were concerns about local enumerators, the security of completed questionnaires and the hiring and retention of a large number of employees for a short period; the rapid growth of Internet access in Canadian households; and the federal government's strategic plan to make all its services available on-line by 2005.

Despite the few problems with the application, which were easily fixed, the Internet option proved very beneficial in a number of respects, including the following:

- rapid reception of questionnaires;
- reduced response burden (with automated skips, respondents saw only the questions that applied to them; they also did not have to mail their questionnaire back);
- efficient data processing (no data capture was required since the data were already in electronic form);
- increased completeness of questionnaires (validation messages prompt respondents to answer questions they have overlooked);
- greater access to additional information through the help feature; and

- as a result of the last two points, less follow-up required.

In view of all these advantages, it would make sense to try to obtain as many responses as possible via the Internet.

According to various sources, the Internet connectivity rate was approaching 70% in 2006. However, the Internet response rate in that year's Census was only 18.3%. The gap between the two figures suggests that there is potential for more on-line responses.

The evaluation surveys conducted at the time of the 2004 Census Test showed that part of the gap was probably due to the poor visibility of the Internet option on the paper questionnaire. Many Internet users did not know that the option existed. The surveys also revealed that many respondents chose to complete the paper questionnaire rather than the on-line questionnaire simply because it was readily available to them. Consequently, in 2006 we started looking for ways to promote the on-line response option. Changes were made in the front cover of the census questionnaire to make the Internet option more visible. A communications program was developed to further promote the option. In addition, the Internet Response Promotion (IRP) Study was launched.

## **6. Internet Response Promotion study**

### **6.1 Description**

The purpose of the study was to promote Internet response by encouraging households to complete their questionnaire on-line. To that end, we sent them a letter instead of a paper questionnaire. The letter provided the URL of the Web site, a unique access code for each household and a toll-free telephone number. Respondents could call the number to get help or to ask for a paper questionnaire if they preferred that option.

To minimize the burden on data collection operations, especially total non-response follow-up, we tried to target households whose members used the Internet. Since there was no list of those households, we developed a model to determine in advance which households were most likely to respond via the Internet. We based the model on the characteristics of respondents who used the Internet response option in the 2004 Census Test. The specificity of Internet respondents was mainly evident in the following characteristics: age, education, labour force activity, student status, the presence of children in the household, and income. From the households that satisfied this profile, we selected a preliminary sample of 40,000 households and split them into two groups of 20,000 households: the IRP sample, which received a letter; and the control group, which received a paper questionnaire.

The IRP Study has two dimensions that have to be kept distinct: the model and the method. The model refers to the tools used to determine in advance the profile of households that are likely to respond on-line, and the method refers to the mailing of a letter instead of a paper questionnaire.

### **6.2 Results of the Internet Response Promotion study**

The effectiveness of Internet response promotion (IRP) is evident in the table below, which shows the proportions of responses received for the IRP sample, the control sample and the 2006 Census in general.

**Table 6.2-1**  
**Questionnaires received by response mode for the IRP sample, the control group and the Census**

Response mode	Census %	Control group %	IRP sample %
Internet	18.3	23.8	61.8
Paper	63.1	61.1	7.6
CHL	2.3	2.1	7.7
NRFU	16.3	13.0	22.9

The proportion of Internet responses was 18.3% in the 2006 Census, 23.8% for the control group and 61.8% for the IRP sample.

The effect of the model, i.e. identifying households likely to respond via the Internet, can be measured by comparing the proportion of Internet responses in the Census with the proportion in the control group, since the households in the control group were all likely to respond on-line. The difference between the two was 5.5 percentage points. The effect of the method, i.e. sending a letter, can be measured by comparing the proportion of Internet responses for the IRP sample with the proportion for the control group, since only the former received a promotion letter instead of a questionnaire. The difference between the two was 38 percentage points. The method thus had a much greater impact on promoting the Internet response option.

While the method was very effective at increasing the number of Internet responses, it carries certain risks in a census. It increases the number of calls to the Census Help Line and the number of non-response follow-up cases, which represents an additional burden for collection operations. However, the method's negative impact was minor in 2006, since the IRP sample was small (20,000 households across Canada). In any case, we will determine later how we plan to deal with these issues in the next census.

### **6.3 Follow-up surveys**

We administered evaluation surveys to the households selected for the IRP Study for the purpose of assessing the on-line response option and improving the on-line response methodology for the next census.

The surveys showed that the method (mailing out a letter instead of a paper questionnaire) drew very little negative reaction from the households, especially, as one might expect, from households that chose to respond via the Internet.

Nearly 21% of on-line respondents in the IRP sample reported having had at least one technical problem with the application. The most common problems were accessing the Web site, accessing the on-line questionnaire and adjusting their computer's settings to the site's requirements. However, none of these difficulties prevented that group from submitting their questionnaire on-line.

Nearly one in five on-line respondents in the IRP sample (18%) mentioned receiving a validation message (non-response, invalid response or amount verification). Most of those respondents (88%) did not know that they could have ignored the messages.

Nearly 40% of on-line respondents in the IRP sample called the CHL to get help or to ask for a paper questionnaire. Respondents who asked for a paper questionnaire did so for various reasons: technical difficulties in using the Internet application, a preference for the paper option, and practical issues (the letter was lost, thrown out, ignored or misunderstood).

## 7. Planning for the 2011 Census

Planning for the 2011 Census is under way. To reduce costs and the response burden, and to speed up data processing and dissemination, we will attempt to obtain as many responses as possible via the Internet. We plan to do this by using a modified version of the IRP methodology on a broader scale. We will mail a letter to 60% of Canadian households instead of sending them a questionnaire. The letter will provide the URL of the Web site, a secure access code and the toll-free number to call for a questionnaire. We also plan to introduce an automated questionnaire request system, which should be more efficient than the service offered in 2006 and should prevent any additional burden on data collection operations, especially the Census Help Line. The remaining 40% of households will receive a census questionnaire in one of the traditional ways: by mail, by drop-off or through a visit from an enumerator. With this methodology, we expect an Internet response rate of 40%.

Collection for the 2011 Census will be based on a wave methodology. Canadian households that receive the Internet promotion letter will do so in wave 1, which will take place seven days before Census Day. In wave 2, two days after Census Day, those households will receive a reminder letter if they have not completed their questionnaire. In wave 3, which will occur 10 days after Census Day, they will receive a questionnaire in the mail if they still have not responded. This step should also considerably reduce the burden on data collection operations, especially non-response follow-up. Finally, 20 days after Census Day, non-response case managers will contact households that have failed to respond.

In addition, we will continue to target households whose members use the Internet, with the aim of minimizing the burden on non-response follow-up and avoiding placing a burden on households that do not use the Internet. To that end, we are currently developing a logistic regression model based on the results of the 2006 Census.

To avoid inconveniencing the public wherever possible, the geography unit will be the collection unit, about 500 dwellings. All the dwellings in the same collection unit will receive the same thing, either a letter or a paper questionnaire.

## 8. Conclusion

Providing Canadians with the option of completing their census questionnaire via the Internet is very promising. Initial experience with Internet data collection has shown that the new response mode has many advantages. However, collecting data with a paper questionnaire will always be necessary to enumerate households that do not have Internet access. The on-line response option entails the costly introduction of a second data collection infrastructure. Nonetheless, savings can be achieved if the proportion of Internet responses is high enough to reduce the costs associated with processing paper questionnaires. The difference between the proportion of Internet responses in the 2006 Census and the Internet connectivity rate in 2006 suggest that a high Internet response rate is possible, especially since the IRP methodology proved very effective. The wave collection methodology planned for the 2011 Census will use both infrastructures. In the first wave, the IRP method will be used to enumerate some of the population, and paper questionnaires will be used to enumerate the rest. This approach will exploit the full potential of both models while minimizing their drawbacks.

## References

- Laroche, D. (2007). 2006 Census Internet Data Collection. Technical report presented at Statistics Canada's Advisory Committee on Statistical Methods, April 30-May 1, 2007.
- Laroche, D. and Grondin, C. (2008). Impact of Online Edits and Internet Features in the 2006 Canadian Census, *United Nations Statistical Commission and Economic Commission for Europe, Conference of European Statisticians*, Vienna, Austria.

Grondin, C. and Sun, L. (2008). 2006 Census Internet Mode Effect Study, *Proceedings of the Survey Methods Research Section*, American Statistical Association

Côté, A., Kevins, C. and Laroche D. (2007). Enquêtes d'évaluation de l'option Internet au Recensement de 2006, unpublished report, Ottawa, Canada : Statistics Canada.