

Distr.

GENERAL

Working Paper No.18

10 March 2008

ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION

STATISTICAL OFFICE OF THE

EUROPEAN COMMUNITIES (EUROSTAT)

**ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)**

STATISTICS DIRECTORATE

Meeting on the Management of Statistical Information Systems (MSIS 2007)

(Luxembourg, 7-9 May 2008)

Topic (iv): User Perspective

CANADA'S 2006 CENSUS ONLINE OPTION

Invited Paper

Prepared by Graeme Gilmour and Karen Doherty, Statistics Canada

I. INTRODUCTION

1. In May 2006 Canada's conducted its 25th census since confederation in 1967. This census introduced the largest set of changes to the collection and processing methodology in the past 30 years. The new approach addressed a number of important issues that emerged in the previous census cycles, relying heavily on computing technology and centralized processes to achieve the desired outcomes. A critical mass of Canadians expects to be offered services online but Canadians are increasingly conscious of privacy issues and continue to expect total confidentiality protection of census responses.

2. The 2006 Census provided a secure online Internet based application so that everyone who lives in Canada could complete their Census questionnaire online. This short document highlights some of what we learned following the decision to offer an online option for 2006 and the sizeable dress rehearsal conducted in May, 2004.

II. BACKGROUND

3. The 2006 Census provided the capacity for respondents in all private households and agricultural operations in Canada to complete either the long (53 questions) or short (8 questions) questionnaire by Internet. It was anticipated that approximate 16% of respondents would use the online application, which was available 24 hours a day, 7 days a week. A comprehensive public communications strategy was planned to encourage respondents to complete their census form online. Census questionnaires were mailed to about 73% of the dwellings in Canada, with the remainder hand-delivered by an enumerator. A unique Internet access code was pre-printed on the front of each paper questionnaire. Respondents completed and transmitted their questionnaire directly to Statistics Canada's census Data Processing Centre, where the response was registered and its data integrated into the regular flow of census returns. Responses received on paper and through electronic channels were logged against a master control list of dwellings in order to track the status of all dwellings. This was essential in order to allow notification to field staff to either conduct or suspend non-response follow-up.

4. The Internet application was designed so that after a respondent uses it to submit their questionnaire, no software trace (footprint) was left on his/her computer. Respondents who completed the long census questionnaires were able to save their partially completed form and establish a password. By using their password in conjunction with their Internet access code, respondents could complete their form over multiple sessions. If a respondent chose to save their form and did not return within a pre-specified period of time, the partially completed form was submitted on their behalf automatically, and was passed on to subsequent processing steps. In addition, Internet respondents who saved their form could access it from different locations (roaming). For example, they could start the form at home, save their partially completed form and complete it from work. The Internet application was available in both of Canada's official languages (English and French), and respondents could toggle back and forth between languages as they completed the form. Once the online questionnaire was submitted, the subsequent processes were independent of the response mode, be it through paper or computer assisted telephone interviewing, thus making it a truly integrated approach.

5. In our experience, there are a few factors that should be considered essential, before a census Internet option is offered to a country's population. In addition, there are a few key elements, which if present, would greatly enhance the likelihood of a relatively successful census online offering.

III. PRECONDITIONS

A. Collection Methodology

6. Adequate literacy levels and the ability of a large portion of the population to self-enumerate are absolutely essential for an Internet census option. The ability to provide each household (or individual) with a unique Internet access code linked to a geographical location is key to having an effective tracking and control system.

7. For the 2006 Census, Statistics Canada created a list (the Master Control System or MCS) of all dwellings with a unique identifier linked to a physical address (where a questionnaire was mailed-out), and linked to a geographical code (where a questionnaire was manually dropped off). Each questionnaire was pre-printed with this unique identifier as well as a unique Internet access code. Whatever response mode was used by a respondent, the questionnaire was registered using this unique identifier and the MCS was updated. Non-response follow-up lists were created from the MCS and transmitted to Field staff on a regular basis, allowing contact with respondents only for those dwellings that had not responded.

B. Connectivity and Public Expectations

8. Clearly any country considering the viability of developing an application for online filing of census forms must consider the level of connectivity amongst the population, their access to high-speed Internet, and the level of trust they exhibit through indicators such as the extent and ease with which the population conducts "business" using the Internet (for example, online banking, filing tax forms).

9. In Canada, a majority of households are connected via the Internet. Based on the 2003 Household Internet Use Survey, an estimated 7.9 million (64%) of the 12.3 million households in Canada had at least one member who used the Internet regularly in 2003, either from home, work, school, public library, or other location. An estimated 6.7 million households (54%) had a regular user from home, with 4.4 million (65% of the home users) with high-speed access to the Internet. Furthermore, about 57% of households using the Internet at home had someone who accessed online banking services, well above the 44% in 2001.

10. This growth indicated that Canadians were becoming more confident in the security aspects of the Internet, and the enormous penetration of the Internet in Canadian households, along with the sophistication of its use in differing types of transactions, suggested that the public would have a strong expectation for a secure and efficient online option for the 2006 Census.

C. Mature Security Infrastructure

11. While Canadians in general are taking to the web, trust in the level of security offered by the vendor or institution plays a big factor in their decision to either conduct or not conduct an online transaction. Canadians insist on a high level of confidentiality from their governments and therefore the provision of a secure infrastructure was absolutely essential in ensuring that the trust factor was high enough for the respondent to be willing to provide confidential census data online. This led to the decision to offer a level of security that was

higher than most other online transactions to which Canadians have become accustomed. The fundamental difference in a statistical survey or census is that the respondent is being solicited for a response, rather than initiating or benefiting directly from the transaction, such as in the case of online banking or requesting unemployment insurance.

12. As part of its Government Online (GOL) strategy to have all services online by 2005, the Government of Canada pursued the creation of a “Secure Channel” through which citizens could transfer confidential information using the Internet. This initiative is offered to Federal departments as they implement their individual online strategies. In partnership with the Secure Channel team Statistics Canada invested in the development of bidirectional encryption between the STC servers and the user’s computer. This functionality was required in order to conduct online edits as well as to offer respondents the choice to save and resume a session in the future. This, coupled with the use of anonymous, limited-use PKI certificates, brought an exceptionally high level of security for the information.

D. Experience with Online Surveys

13. Many of Statistics Canada’s business surveys have offered respondents the option of providing data electronically. While these initiatives are largely geared to reducing respondent burden for companies who supply data on an on-going basis, and there are significant differences between the census and business surveys in terms of content and the respondent community, we have successfully leveraged from these initiatives to plan and implement the census infrastructure and to address a number of risks.

IV. STRATEGIC CONSIDERATIONS

14. Having worked through the essential and highly desirable pre-conditions for offering an online census option, a number of strategic considerations weigh heavily on the manner in which to proceed with the design, planning and implementation of such an undertaking. These include cost factors, potential savings and implementation strategies.

15. The main cost drivers for the Internet application are questionnaire development and testing, developing a secure infrastructure, the application development, testing (functional, integrated and volume) and support (technical and respondent) efforts, telecommunications infrastructure, hardware/software and public communications expenditures. In our 2006 Census context, about half the expenditures were associated with the customization of the Secure Channel infrastructure and could possibly be considered as a one-time investment.

16. The maximum number of concurrent users plays a critical role in determining the capacity and thus cost, and in our case, we sized the infrastructure to accommodate 15,000 concurrent user sessions. Our total target response rate over the Internet was 20% of households that respond prior to non-response follow-up (approximately 2 million questionnaires). Not meeting this target would have had a direct impact on the capacity of the paper-handling systems and operational costs.

17. One of the key considerations was the level of expertise and where this was best capitalized upon in order to minimize development, integration and implementation risks. For us, the conclusion was that we had the in-house expertise to design, user-test and specify the online questionnaire format, flow, skip-patterns, and rules for online edits. These factors influence the respondent experience, the overall look and feel of the application, comparability to the paper questionnaire (possible mode effects), as well as the resulting data quality and as such deemed core to the expertise of our agency. Conversely, it was concluded that the private sector was better able to meet the systems integration challenge associated with programming and linking the online application with downstream processing systems and ensuring that it was optimized to handle the anticipated volumes.

18. It is important to understand how Internet users use online applications and the possible mode effect that can be introduced by attempting to adhere too closely to the existing paper form. Respondents who choose to respond via the Internet tend to be experienced Internet users, and have expectations about how the form should work based on their experience with other web-based forms. We carefully considered the potential implications before introducing anything in the Internet form that was different from the paper form, and for 2006, achieved a balance with general adherence to the paper form, while incorporating many useful web standards.

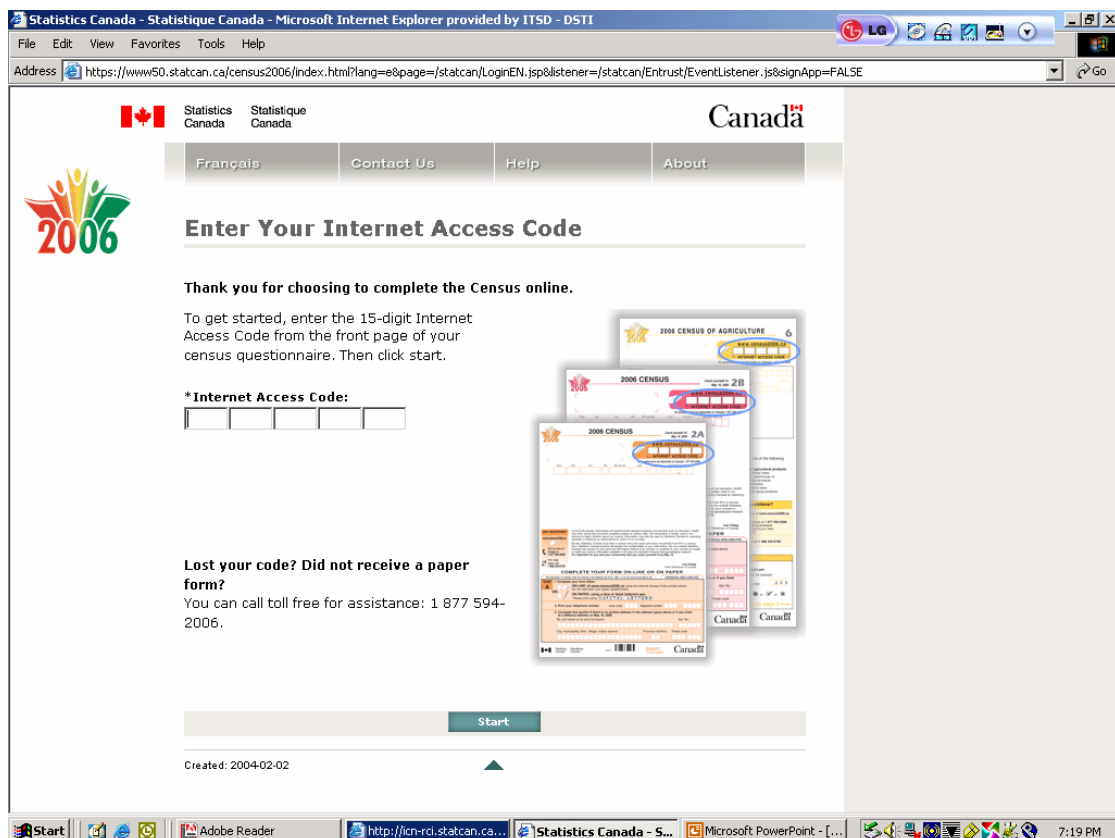
19. Finally, despite the stringent security measures in place to protect confidential data, careful consideration had to be given to how respondents would react to the private sector being involved with real or perceived concerns about data confidentiality and privacy. Any widespread negative reaction could have had an adverse impact on overall response rates and create respondent relations issues for the statistical agency.

V. RESULTS

A. Improved User Experience

20. Ensuring a high-level of satisfaction with the online experience is essential in retaining the respondent for the duration of the session, resulting in a completed questionnaire. The questionnaire was accessed through a web page offered in both official languages. Accessibility options for respondents with visual disabilities were also supported. The application required the presence on the respondent's workstation of a Java applet to allow the browser to function with the PKI solution used to ensure end-to-end encryption of the data. No other software was loaded on the respondent's desktop.

21. A large number of browsers were tested to ensure that the application would run on the majority of browsers likely to be in use across Canada. Prior to the Census the best information available indicated that approximately 6% of workstation did not already have the Java applet loaded. Respondents were provided with information to assist them in downloading the applet if their browser was not already Java-enabled. To access the application the respondent simply had to enter the Internet Access Code printed on their census form into the space provided on the Census questionnaire web page and proceed to complete the questionnaire.



22. User satisfaction is influenced by a number of factors, such as the ease with which the questionnaire is accessed, which is influenced by the minimum system requirements of the security solution and the complexity of the login process. Other factors include the absence of lengthy downloads or foot-prints, working with the default configuration of most users' machines, being compatible with common browsers, intelligent skip patterns, soft edits that permit the user to continue even if a particular response was not valid or provided, and the respondent's perception of the relative speed and ease of use of the application. After the 2004 Census test,

Statistics Canada conducted a follow-up survey with a small sub-set of those who responded over the Internet to judge their level of satisfaction. Some of the relevant results are as follows:

- the majority of the respondents (89%) indicated that they completed their Census test questionnaire at home;
- 79% of the respondents indicated that they had a “high speed” connection;
- when asked why they chose to complete their Census test questionnaire online, 52% said because it was easier, 30% said because it was faster, 18% said because of personal preference, 16% said because they did not have to mail it;
- 95% of the respondents rated their overall experience with the Census Internet application as favourable;
- 88% of respondents felt that the time it took to complete the Census Test questionnaire was acceptable, of which more were short-form respondents. Respondents who indicated that they did not have ‘high speed’ were more likely to say that it took too long;
- 98% of respondents indicated they would complete their Census questionnaire online in 2006;
- 57% of respondents were not at all concerned that the privacy and the confidentiality of their Census test questionnaire data were more at risk on the Internet, while 34% were concerned and 8% were very concerned;
- when asked how secure it is to transmit personal information over the Internet, 73% stated that it was secure, 13% stated it was very secure and 10% stated it was not secure at all.

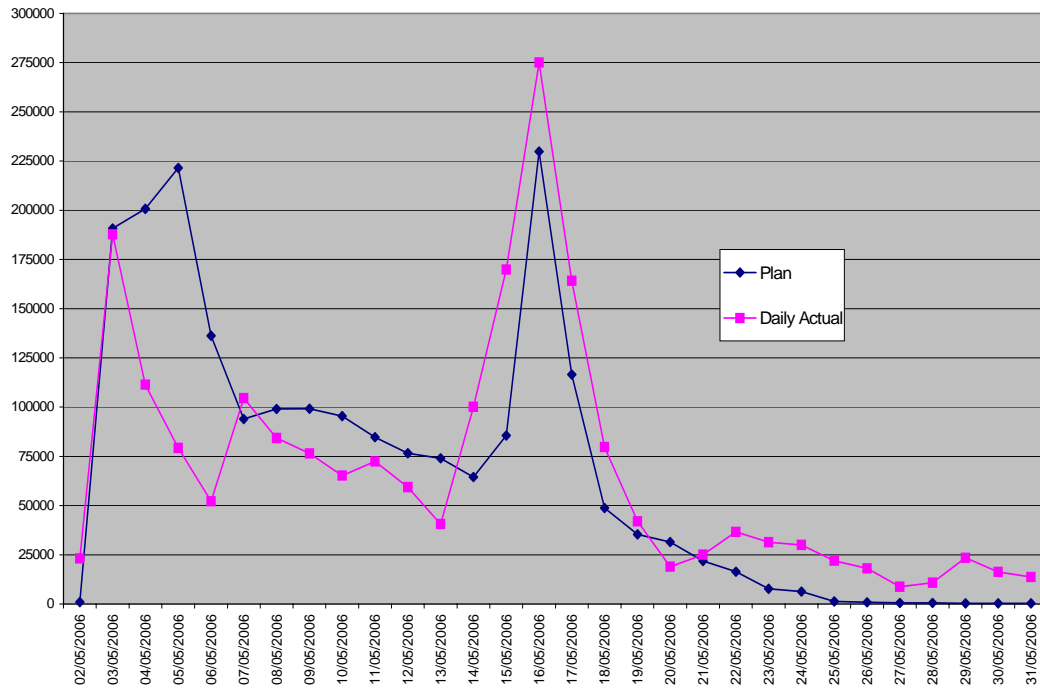
23. STC conducted a respondent satisfaction survey following the 2006 Census using a very small sample from the general population and a sub-sample of the respondents who were in the sample selected for the test of the “Push Strategy” (see the section *IV Future Opportunities*, paragraph 35). As of the writing of this paper the results had been tabulated but not yet released for distribution. The analysis however generally support the findings from the test conducted following the 2004 Census Test.

B. Internet Response Rate

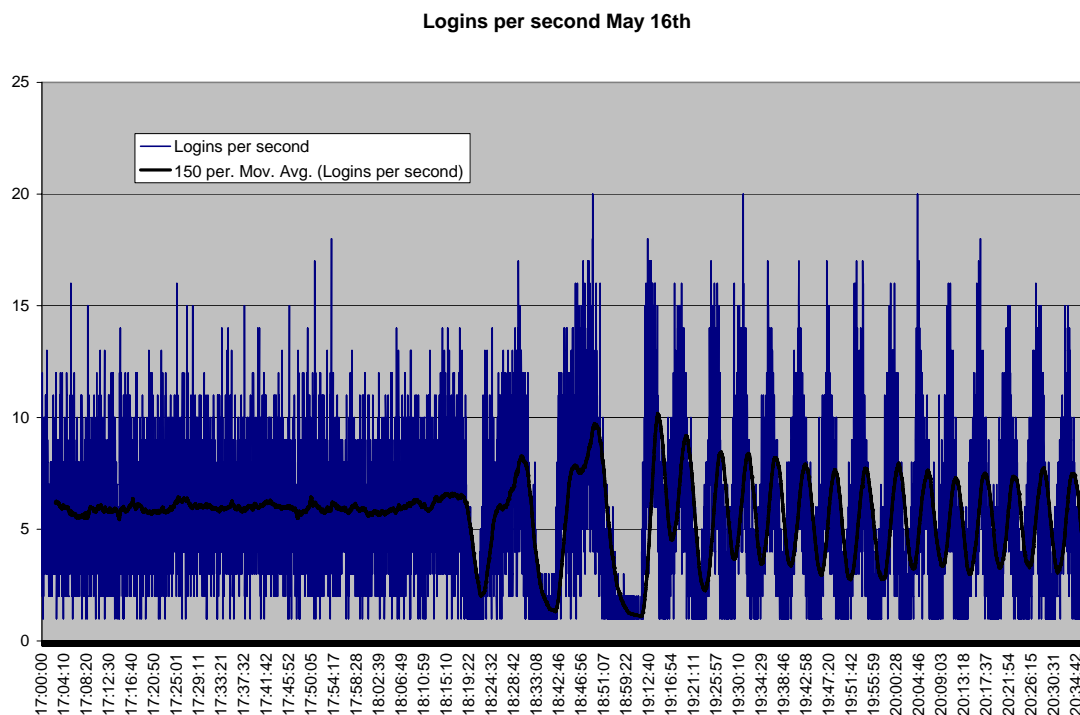
24. Our total target response rate over the Internet was 20% of households that respond prior to non-response follow-up (approximately 2 million questionnaires). In 2006, prior to the start of the non-response follow-up, about 22% of responses were received via the Internet. Given that the non-response follow-up stage is almost exclusively conducted with paper questionnaires, by the completion of data collection operations, some 2.26 million or 18.5% of the total households in Canada used the Internet response mode.

25. The actual use of the Internet option was somewhat lower than anticipated in the two weeks leading up to Census Day but higher than anticipated in the few days prior to Census Day, on Census Day itself, and over the weeks following the Census Day. The planned versus actual pattern of use of the Internet option is illustrated in the following graph.

2006 Census Internet responses Plan and Actual



26. While our targets were established for the entire Census collection period (some 9 weeks from start to finish), we had to specify and build technical capacity to sufficiently handle a finite number of sessions at any one time. With very little experience, and only a theoretical model on which to estimate infrastructure demand, we developed a system of close monitoring, oversight, and escalation and introduced the notion of “graceful deferral” to balance load, performance, cost and potential risk. The notion of graceful deferral, allowed a pre-determined number of respondents already on the system to continue uninterrupted at an acceptable level of performance, and prevented any additional users from accessing the system to the point where the overall infrastructure could no longer be sustained. We faced enormous demand on Census night, especially between 6 and 11 pm, and invoked graceful deferral repeatedly during this peak period. It is estimated that approximately 150,000 users were asked to re-try the application at another time. The pattern of use on Census Day is shown on the following graph.



27. Although information provided to households indicated that they could respond online as soon as they received the questionnaire rather than wait for Census Day, the media focussed on the message that the questionnaire should be completed on Census Day. It is hoped that a mail-out strategy that depends on waves of mailing along with a better communications campaign would result in a more even distribution of census returns over the two weeks preceding Census Day and the week following Census Day. This may help reduce the load experienced on Census Day itself.

C. Higher Quality Responses

28. In general responses received from the Internet were more complete than those received from paper. In our census test, we observed that the incidence of item non-response was many times lower for Internet submissions than for paper. Specifically, for short Census of Population forms the item non-response rate was 0.01% for Internet responses and 2.54% for paper responses. For long Census of Population forms the item non-response rate was 1.80% for Internet responses and 6.97% for paper responses.

29. This observation held true for the actual Census in 2006. The item non-response rate for the short form on paper was 102 times higher than on the short forms received through the online response mode. For the long form the item non-response rate was 10 times higher on the paper forms as compared to the forms submitted via the internet.

30. There are a number of factors that contribute to the improved quality observed with on line responses. By design, a number of online edits were built into the application to prompt respondents when they left questions blank or entered conflicting information. Automated skips guided respondents past questions that were not applicable, and questions were personalized with respondents' names thus reducing the likelihood that one person's responses were inadvertently recorded in another's response area. Also, no data capture errors were introduced during processing. The use of radio buttons in certain questions eliminated the possibility of conflicting responses and the use of drop down menus where appropriate, assisted users in providing appropriate responses. Finally, there was a general perception on the part of users, confirmed by a survey conducted with a sample of Internet respondents after the census test, that the Internet form was easy to use and quick to complete. This is likely at least part of the reason behind the observation that more respondents completed the whole form online (in comparison to paper) rather than getting discouraged and quitting before the end.

D. Cost Savings

31. The potential savings opportunities are in the areas of reduced paper-handling costs, mail-back postage costs, and reduced follow-up costs for missing or inconsistent information. For 2006, we estimated that a 16% Internet return rate, would result in savings that approximated the cost of developing and supporting the Internet application, and thus place us in a near break-even position. At 18.5% the 2006 Census achieved and slightly surpassed this objective.

32. Our Census experience indicated a much higher rate of item response and very low edit-failure rates for the Internet questionnaires. A higher take-up rate would concurrently result in lower follow-up and imputation rates, resulting in lower costs as well as increased quality. The online application also holds prospects for presenting a wider range of options to select from for certain responses, such as those in the labour or education categories (occupations, industries, educational attainment), thus reducing manual coding effort.

D. Reduced Reliance on Field Staff

33. The new methodology introduced in 2006 promised a reduction in the number of enumerators and other field staff needed to complete the Census. Canada is a large country geographically with enormous tracts of sparsely populated land. With the aging of the population and a booming economy in the western provinces due to increased oil production, the need to find a way to reduce our dependence on field staff is critical to our ability to complete future Census projects successfully. Increased use of the online collection option is expected to further reduce the size of the Census work force in the future.

VI. FUTURE OPPORTUNITIES

34. Predicated on a number of basic assumptions such as the continued acceptance and use of the Internet by most citizens, the Internet option holds prospects for long-term cost savings, quality gains, and reduced respondent burden.

35. In geographic areas with a very high Internet penetration rate, simply mailing to households a card providing the URL and an access code for the online option (the Push Strategy) will increase the internet response rate and may also significantly reduce the printing and delivery costs of the paper questionnaire. In Canada this may present proportionally higher savings, as in most cases we have to print and deliver a questionnaire in both English and in French. The challenge is in being able to deliver a paper questionnaire in the language of choice in a timely fashion to those who cannot or do not wish to use the online application (through a phone call to our help-line).

36. As part of the 2006 Census, a study was initiated to test the viability of the Push Strategy. A preliminary sample of 40,000 households in mail-out areas was selected for this study. This sample was split randomly into two groups of 20,000 households each in order to create a control group. The control sample received a paper questionnaire while the Push Strategy sample received the letter. The results indicated that this strategy would likely double the use of the internet option but that the collection methodology would need to properly service the households who choose to request a paper questionnaire.

37. Our Census experience indicated a much higher rate of item response and very low edit-failure rates for the Internet questionnaires. A higher take-up rate would concurrently result in lower follow-up and imputation rates, resulting in lower costs as well as increased quality. The online application also holds prospects for presenting a wider range of options to select from for certain responses, such as those in the labour or education categories (occupations, industries, educational attainment), thus reducing manual coding effort.

38. With an ever-increasing diversity in the Canadian population, the Internet makes it possible to present the census questionnaires in multiple languages at modest additional cost. Even if the respondent chooses to use the paper questionnaire, the Internet can be used to provide translation of the questions in multiple languages. Overall, these could hold the prospect of increased response and lower follow-up costs. Similarly, the online application provides some real alternatives to people with disabilities such as those who are blind to self-enumerate rather than rely on an enumerator.

VII. CONCLUSION

39. Statistics Canada's experience with the development of an online Census application for the 2006 Census has been positive. The execution of a sizable and live Census test in 2004 provided us an opportunity to

exercise an innovative design with respondents in an integrated collection and processing environment. A number of factors favoured the introduction of the online census application for 2006, such as the relatively high connectivity rates and comfort with transacting over the web by a majority of the Canadian population.

40. Our census collection methodology of mailing-out/dropping off questionnaires for respondents to self-enumerate was also conducive to the introduction of the online application. An integrated processing design was developed whereby all modes of collection culminate in the same downstream editing, follow-up, and imputation systems and processes. Our contracting approach yielded a competent and experienced vendor, well positioned to deliver a robust, integrated and user-friendly application. Our in-house online survey experience allowed us to test and specify relevant and priority requirements. The timely development of the Government of Canada's secure channel initiative and their flexibility in adapting it to suit our needs aided us significantly in presenting a secure infrastructure. We instituted a robust governance and project management structure in order to ensure that design changes and costs were closely monitored and tightly controlled.

41. All this meant that we were well positioned to meet the expectations of Canadians for a secure, efficient and user-friendly online application in 2006. Our successful experience in 2006 has given us the confidence to target an online response rate of 40% for the 2011 Census through the use of the "push" strategy (sending only online invitation cards to a large percentage of households).