

Article

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

Building a Business Case for Census Internet Data Collection

by Tracy Moore, Lorna Bailie and Graeme Gilmour

2009



Building a Business Case for Census Internet Data Collection

Tracy Moore¹, Lorna Bailie and Graeme Gilmour²

Abstract

Census developers and social researchers are at a critical juncture in determining collection modes of the future. Internet data collection is technically feasible, but the initial investment in hardware and software is costly. Given the great divide in computer knowledge and access, internet data collection is viable for some, but not for all. Therefore internet cannot fully replace the existing paper questionnaire – at least not in the near future.

Canada, Australia and New Zealand are pioneers in internet data collection as an option for completing the census. This paper studies four driving forces behind this collection mode: 1) responding to social/public expectations; 2) longer term economic benefits; 3) improved data quality; and 4) improved coverage.

Issues currently being faced are: 1) estimating internet uptake and maximizing benefits without undue risk; 2) designing a questionnaire for multiple modes; 3) producing multiple public communication approaches; and 4) gaining positive public reaction and trust in using the internet.

This paper summarizes the countries' collective thinking and experiences on the benefits and limitation of internet data collection for a census of population and dwellings. It also provides an outline of where countries are heading in terms of internet data collection in the future.

Key Words: Census, Internet data collection, Electronic data reporting.

1. Introduction

1.1 Introduction

Data collection methods in population censuses are continually being reviewed in response to changes in the environment, culture and opportunities arising from new technology. The internet is one of the newer frontiers in data collection, which has recently been used for census of populations.

Many countries have offered, or are in the process of offering, respondents the option of completing their Census information via the internet. A key driver for adopting the internet is the potential to improve the way censuses are carried out. Utilizing the internet provides benefits including:

- improving/maintaining census participation (coverage) in an environment where response rates are dropping
- improving data quality
- long term reduction in costs and/or opportunities to redirect resource efficiencies, and
- responding to social/public expectations.

However, including an internet response option in addition to other response modes (such as telephone, face-to-face or mail-back), involves challenges and can cause increased complexity in the field. Some important points to consider include:

- estimating take up of each mode and planning manual collection/follow-up and system capacity
- ensuring seamless links for questionnaire registration and control between the multiple modes
- designing a questionnaire for multiple modes

¹ Tracy Moore, Statistics New Zealand

² Lorna Bailie and Graeme Gilmour, Statistics Canada

- producing multiple public communication approaches
- ensuring privacy and security of the internet completed information.

This paper summarizes various countries’ collective thinking and experiences on the benefits and obstacles to internet data collection for census of populations. It also provides an outline of where countries are heading in terms of using internet data collection in the future.

1.2 Contextual setting: International approaches to the use of internet data collection

Census developers and social researchers are at a critical juncture in determining collection modes of the future. Internet data collection is technically feasible, but the initial investment in hardware and software is very costly. Given the great divide in computer knowledge and access, internet data collection is viable for some, but not for all. Therefore internet can not fully replace the existing paper questionnaire – at least not in the near future. Census takers who wish to offer an internet option have to invest at least initially in supporting multiple modes.

1.3 Previous experience

A number of statistical agencies have recently used the internet as a census data collection method to varying degrees. For their respective 2006 Census cycles, Canada, Australia, and New Zealand offered an internet option as part of their collection strategy. The U.S. 2005 National Census Test conducted extensive testing of two different Internet application designs. This leading research compared a person-based design against a topic-based design to determine which would work better in terms of data quality and ease of use. All countries indicated positive results and viewed the online option as an important approach towards achieving longer-term benefits. A summary of take up rates and future plans by country are summarized in Table 1-1.

Table 1-1
Summary of internet experience and future plans for internet data collection

Country	Internet experience	Future plans
Canada	Started in 2006 with a take up rate of 18 percent	Plan for 2011 is to increase the take up rate to 35-40 percent. Explicit strategy to target internet users. Reuse 2006 system where possible
Australia	Started in 2006 with a take up rate of 9 percent	Plan for 2011 is to increase use. Explicit push at the doorstep (if Internet acceptable, paper form won’t be provided)
New Zealand	Started in 2006 with a take up rate of 7 percent	Plan for 2011 is to increase the take up rate to 20-30 percent. Increased awareness via publicity. Provide internet ePIN to all households
England and Wales	Studied the feasibility	Plan to implement for the first time in 2011 and have estimated a 20 percent take up rate.
USA	Internet option tested on four occasions 2005 test take up rate of 7.2 percent	Will not be offered in 2010

1.4 Future plans

Canada is planning to implement an ‘internet promotion strategy’ for the 2011 Census to stimulate more Canadians to respond via the Internet. Based on knowledge gained from the 2006 Census, dwellings most likely to respond via the Internet will be targeted to receive an internet promotion letter and no questionnaire. The letter will have a URL and a secure access code, as well as a number to call if respondents would prefer a paper questionnaire. It is

expected that approximately 60 percent of dwellings will be in this group. The remaining dwellings (40 percent), will either be mailed or dropped off a paper questionnaire which will also have an Internet access code should the respondent wish to use the Internet.

Australia will be running a drop-off, pick-up/internet Census. Internet will be the preferred return option, so respondents will be actively encouraged to respond via internet. Based on a 2006 staff survey, one of the main reasons identified for opting for the paper form was that the paper form was provided. Therefore developments will focus on providing access to the internet option without the paper form.

New Zealand will provide all households with paper forms, an internet PIN and internet promotional material. The internet option will also be actively promoted for the first time in order to maximize take up. New Zealand are currently investigating ways to maximize internet take up including how the internet form should look, access to the website, communications, and the role that the collectors play in promoting internet.

The Office of National Statistics (ONS), which is responsible for producing a census of England and Wales, will introduce internet collection as part of their 2011 Census. Even though they recognize this will mean increased costs and significant security challenges, their decision is based on the need to respond to the expectations of government and the public.

The U.S. Census Bureau conducted a number of studies to test an internet option for 2010. Currently a decision has been made to eliminate the internet option from the Decennial Response Integration System (DRIS) contract, which provides response functionality for the 2010 Census.

2. Benefits of internet data collection

2.1 Improving/Maintaining participation (coverage)

Most countries report coverage issues for the young mobile generation and inner city apartment dwellers. New Zealand has also identified the country's growing Asian population as a hard to enumerate group primarily due to their diversity in terms of language and culture. With many countries' response rates on the decline, targeted initiatives are required to focus resources and develop new ways to improve the participation of these groups. The internet, as a response option, is seen as making participating more attractive to these groups due to its use of modern technology, convenience and ease of completion.

Currently there is little evidence proving an internet option will improve coverage of these hard to reach groups, but one wonders what the impact will be on future coverage if internet data collection is not an option.

2.2 Improving aspects of data quality

Studies conducted by Canada, Australia and New Zealand have indicated that questionnaires returned by the internet are more complete than paper forms. More work is required to determine how much this apparent improvement in quality is attributable to the characteristics of people most likely to respond using the Internet, versus how much is due to the technology itself. There is a distinct characteristic profile of those most likely to respond using Internet which includes higher levels of education.

The Canadian Census encouraged more complete internet responses by using "soft" edits within the application as well as automated transitions to the relevant questions. The quality in terms of completeness and range of values was higher than those received on paper.

Similarly, New Zealand's internet option automatically determined whether certain fields (sex, age, date of birth, usual residence, census night address, and ethnicity), had been completed by the respondent. These variables were mandatory on the internet option and therefore there was no item non-response for these fields for internet

submissions. Data quality for the internet form was therefore considered of a higher quality than that of mail or pick-up. Australia also reported improved completeness of data from internet forms.

All countries have stated that processing of internet responses is quicker than paper responses since the data are already in electronic format. It also improves the confidentiality of information provided by internet respondents through less personal handling.

With respondents completing the information on-line there is no need for subsequent data capture. Data quality is improved by eliminating data capture errors, which in the case of data capture of Canadian paper forms, ranged from 0.4 - 1.3 percent.

Further in-depth mode effect studies are required to confirm the suggested increase in data quality due to internet data collection.

2.3 Long term reduction in costs and freeing of resources

Canada, Australia, and New Zealand will build on their 2006 technical developments but must realize increased internet data collection for 2011 in order to see an even greater return on this investment. Countries are looking at ways to reuse the technology but also realize further investment is necessary for upgrading to new technology and expanding internet take-up capacity.

For its 2011 Census, Canada will build on the infrastructure investment in 2006 and positive public experience and expects cost savings with an increased take up rate of 35-40 percent. The 2011 investment in the internet application and security infrastructure is off set against estimates of reduced collection and processing costs. More specifically:

- Reduced postage
- Reduced mail-back and registration of paper questionnaires
- Reduced labour costs for processing paper questionnaires
- Reduced item non response follow-up

ONS have built a similar case predicting the benefits of internet-based functions will be:

- Reduced calls to the contact centre for help
- Savings in paper scanning/capture
- Speedier capture of data

Australia and New Zealand censuses are still heavily collector driven due to the absence of an address register. Large cost savings can only be achieved if internet take-up is high enough to allow field staff to be reduced. (Both countries see internet data collection as a way to reduce costs over the longer term, primarily through the reduction of field staff numbers).

Given the rapid advancements in technology it is not feasible to assume the same technology can be reused 5-10 years down the road without at least some upgrades. The upcoming censuses will be critical towards validating original business case assumptions of longer term economic benefits. Clearly the take up rate is the key towards realizing any economic return on investment.

2.4 Responding to social/public expectations

Most countries who have added or plan to add an Internet option as a census collection option indicate doing so in response to public demand and expectation. There has been substantial growth in internet access and usage over the past five years and this seems set to continue in the future. A strong incentive for offering an internet option is that the census must keep in step with public expectations on how citizens want to interact with government. Citizens increasingly are filing income tax electronically, banking via the internet and locating information via the internet. There is already increased reliance on technology by younger generations to perform everyday functions. Will Internet collection be the best and perhaps only way to get their attention and ultimate cooperation? Some countries

have legislative obligations, for example Australia has a digital strategy which requires the Australian Bureau of Statistics to provide an internet option for respondents.

The internet provides respondents with more choice on how they can complete/return their census information. In New Zealand, market research has shown that not having to see or wait for collectors to return is an incentive for respondents to complete their form via the internet. This medium allows respondents to return their forms at their own convenience. Also, the growing concerns about the environment demand alternatives to be found to reduce paper and reduce transport costs for collectors.

A 2007 ONS consultancy's finding predict a very high on-line take-up for their upcoming Census. The increase in environmental concerns, the acceptance of government online submissions and new technology are expected to work in favour of online completion.

3. Limitations of internet data collection

3.1 Estimating take up of each mode, planning manual collection/follow-up and system capacity

All countries are having difficulty estimating on-line take up for future censuses. Generally this is hard to predict and therefore countries are being cautious with their estimates (as seen in Table 1-1). The current levels achieved are not able to produce major economic benefits from the internet. More precise predictions of the geographical clustering of internet response have the potential for savings in collection and field follow-up. Field resources could be reallocated from the internet clusters to areas requiring the manual intervention. The key questions are can it be assumed that areas of high take up in 2006 will translate into high take up areas for 2011, or that the pattern of internet take up will be the same?

Testing take up levels can be misleading. For example, it is difficult to gauge the impact of nationwide communication and promotional campaign, as these can not be tested effectively in smaller tests. As well, the impact of the mandatory nature of the census is difficult to estimate. For example in New Zealand, take up levels reached 15 percent during the Dress Rehearsal in 2005, yet during the Census operation only 7 percent were achieved. While in Australia their Dress Rehearsal reached 7.9 percent take up, and increased to 9 percent during Census proper. Canada achieved a 9% take up rate during its 2004 census test, and this doubled to 18% during the 2006 census.

Barriers to internet take up by respondents are not well understood. Currently it is widely thought that when respondents are given the choice of paper and the internet, the majority are likely to choose paper. Paper is a deterrent to getting the public responding via the internet. New Zealand is still required to deliver paper (legal opinion – Statistics Act (New Zealand 1975)), yet Canada has an active campaign to not deliver paper questionnaires to encourage internet take up.

3.2 Designing a questionnaire for multiple modes

Until more recently, it has been thought that the internet form should mirror the paper version in terms of wording, instructions and presentation of response choice. The reason behind this is to minimize the mode effect (i.e. differences in responses due to the method used) and to facilitate the integration of data received from different response channels.

However, qualitative research undertaken by New Zealand has shown that internet users now expect a more interactive experience when using the internet. They expect the form to be well designed and laid out, not basic and rudimentary, and to look fun and interesting. Ease of completion is also important, as is being able to complete the form as quickly as possible. The use of auto fill, being able to populate common information across the questionnaire and household, to avoid repetition was also mentioned as adding to the experience.

A goal of the U.S. 2005 National Census Test was to further advance the Internet technology while maintaining and improving data quality. Two designs were trialed; topic based and person based questionnaires. The most important differences centered on Internet break-offs, which are records where the respondent never clicked the 'Submit' button at the end of the applications. The conclusion reached was that while the topic-based design seemed to be better at producing the accurate number of person records, the break-off cases resulted in increased levels of item non-response. No firm conclusions were drawn from the test but it was believed that the design depended on mode consistency guidance, policy guidance and relative priorities.

Currently there is resistance to an Internet questionnaire design that differs from paper. Subject matter experts and methodologists worry about the modal impacts and how this could affect comparability of reported data. Further investigation is required to establish a balance between the users experience and potential modal bias.

3.3 Producing multiple public communication approaches

Offering internet data collection in addition to replying by telephone, face-to-face, or mail-back challenges current field collection practices and public communications. However there are strong indications that this is a necessary process towards ensuring long term respondent cooperation.

In order to maximize internet take up, a nationwide communications campaign is required to gain awareness from the public. The ultimate aim of the campaign is to attract people to the internet completion website. In 2006 New Zealand deliberately adopted a low promotion approach, and focused publicity through news media and advertising on the web. They also relied heavily on the collectors informing the respondents of this option on the doorstep. These mechanisms resulted in take up of only 7 percent. For 2011 the importance of a nationwide publicity campaign is well recognized as a major contributor for maximizing internet take up. This begs the question who best to target? To maximize the levels, New Zealand has identified the need to bring mainstream respondents on board.

For those countries that still have collectors knocking on doors this is an excellent opportunity to persuade or encourage the public to take part in the Census and can also encourage internet use on the doorstep. However collectors can also put people off the internet if their own opinion of the internet in general is negative. It is important that there are consistent messages from across the field force encouraging respondents to take part. This starts with ensuring effective and consistent training is undertaken--a challenge unto itself.

3.4 Security and gaining trust in the internet

Internet data collection must feature a secure channel that protects privacy and confidentiality. Any real or perceived breach in this regard would be disastrous to achieving a successful census.

Internet collection can enhance confidentiality through reducing the amount of human paper handling. This is especially true for countries doing face-to-face collection. However secure electronic data transmission must be in place to circumvent any attempts of unauthorized access.

Given the data are already in electronic format when it reaches the processing phase, there is less operator intervention, reducing the risk of a breach in the confidentiality of information provided by respondents.

Respondents expect the internet option to have a plausible, and easily understood security structure to allow them to feel safe about providing their information.

4. Conclusion

There are challenges that must be overcome in order to offer an internet option for a census and these are clearly worth tackling in light of the numerous and important benefits as well as the options that become available in the

future once a statistical agency develops a level of public support as well as experience in internet data collection. With a sufficiently high response rate, an internet option can reduce collection and processing costs, but perhaps even more importantly it can be part of the solution to dealing with declining response rates particularly among specific, homogeneous populations.

Looking towards the future, census internet collection could be used to provide recent immigrants whose first language is not an official one the opportunity to respond using a questionnaire in their native language. The internet already offers respondents with visual disabilities the ability to use adaptive technology to complete their census questionnaires.

Statistical organizations that offer internet questionnaires will have the option of adding additional survey content at much lower cost and much more efficiently than is currently possible using other modes of data collection. Samples for post-census surveys could be selected while the census internet questionnaire is being completed, and additional internet questionnaires offered to screened-in respondents immediately after submitting their census forms. Similarly respondents from specific geographic areas could be offered supplemental surveys to gather information to fill specific data gaps.

In summary, the business case for census internet collection is strong and getting stronger with every census cycle.