Efficient Multi Mode Data Collection in a Census Context

David Dolson, Statistics Canada

Abstract

This paper explains the multimode data collection process used in the 2006 Canadian Census. The ways in which a central Master Control System, using a dynamic frame of dwellings and questionnaire tracking, resolves duplicate responses and ensures efficient control and coordination of all operations whether locally or centrally run are explained. Some initial evaluations are given.

Data collection via five response channels must be coordinated. Questionnaires were delivered either by dropoff, based on field listing of dwellings, or by mail using a register based frame of addresses. Respondents had a choice to respond by either Internet or mail. Some data were collected by personal or CATI interviews. As well, respondents to the long questionnaire could either reply to the income questions or give Statistics Canada permission to link to their tax records to obtain these data.

Keywords: Census, Internet survey, coverage, data processing

1. Introduction

Following Canada's 2001 Census of Population, Statistics Canada identified a number of drivers for significant change in its census methodology. The extent and quality of its address register had improved sufficiently that mailout delivery of census questionnaires had become a viable option. Internet use – including use of high speed connections -- by households was growing rapidly, suggesting that Internet collection of census data was potentially viable. In a closely related development, it had become government policy that departments were to move towards ensuring that services were offered to Canadians electronically. Increasing public concerns about privacy of information suggested a move towards a methodology that no longer required collection and editing of data by local enumerators. Difficulty had been experienced in hiring the large temporary workforce needed to conduct the 2001 Census and this was only expected to get more challenging for 2006. Key entry resources that had previously been used for data capture were not expected to be available in 2006 and automated methods of data capture by image recognition had become very reliable.

So for 2006 the Canadian Census of Population has implemented its most significant changes in over 30 years. With increased automation at many steps, significant centralization of many processing steps, a much smaller labour force, implementation of a mailout methodology and availability of an Internet response channel, Statistics Canada has been able to implement a more cost efficient census – reduced cost per dwelling – that is more responsive to needs and concerns of Canadians and which will yield data quality equivalent to or superior to that in the past.

In doing so Statistics Canada has made use of multiple modes of questionnaire delivery and provided for multiple means of questionnaire return. Although this responded to the drivers for change it also created new challenges in controlling and managing many interrelated census operations. This paper focuses on how this was done through use of a central Master Control System (MCS) that served as the dwelling frame and which controlled and tracked questionnaires as they moved through the steps of collection and processing.

Sections 1.1 and 1.2 provide brief outlines of the methodologies of the 2001 and 2006 Censuses. Sections 2 through 6 expand on a series of major steps in the 2006 Census and the role of the MCS in each one. Some preliminary evaluations are given in section 7 followed by a short conclusion in section 8.

This paper limits itself to consideration of the population in private households.

1.1 2001 Census Methodology Overview

The 2001 Census used an area frame methodology in which each of about 43,000 enumerators was responsible for an Enumeration Area (EA) of 300-600 dwellings. In the two weeks prior to census day (May 15, 2001) each enumerator listed the dwellings in his or her EA in a Visitation Record (VR) – the control document for each EA -- and dropped off the questionnaires, one dwelling in five receiving the long questionnaire. In areas covered by Statistics Canada's Address Register (AR) -- about 70% of addresses in Canada -- enumerators reconciled their lists of dwellings in the VRs with corresponding lists of addresses from the Address Register and made any necessary updates. Completed questionnaires were returned by mail to local offices and then forwarded back to the local enumerators who were to record their

receipt in the VR. In addition respondents could call a Census Help Line (CHL) if they had special circumstances or concerns or needed an additional questionnaire or did not receive one. As necessary, the respondent could either request that a questionnaire be delivered to them or provide their data directly to the CHL operator who would complete a paper questionnaire. In these cases the local enumerator was informed of the response so that this could be recorded in the VR. Editing for coverage and completeness was carried out by the enumerator and follow-up as required was done first by telephone and then in person if necessary. Using the VR as the control, non-response follow-up (NRFU) was conducted in person by the enumerator.

In parts of downtown areas of large cities, remote and Northern areas and Indian reserves, questionnaires –all long– were completed by personal interview (canvasser methodology).

When work was complete for each EA, a box containing all the questionnaires and the VR was forwarded to a regional office where data capture was done by key entry. The electronic data as well as the paper documents were then forwarded to a central Data Processing Centre (DPC) for subsequent processing.

So the process was largely decentralized and questionnaires were processed in batches (EAs) from the start. About 98% of questionnaires were delivered by list leave (dropoff) methodology with the remainder by canvasser. Some 80% of questionnaires were returned by mail, 1.9% were completed by canvasser, 0.5% by responses to CHL and about 15.6% as a result of non-response follow-up, leaving about 2% as residual non-response. Every questionnaire was paper (with the exception of a very small test of Internet data collection in London, Ontario and parts of Alberta).

1.2 2006 Census Methodology Overview

In areas where the AR had an acceptable quality list of dwelling addresses, questionnaires were delivered by mail; this accounted for about 70% of dwellings. In most of the remainder of the country a list leave methodology, generally similar to that of 2001, was used for questionnaire delivery. Canvasser methodology was again used in some parts of the country. As in the last few censuses, 20% of dwellings received the long questionnaire except in Canvassing areas where all households got the long form. A list (frame) of all dwellings was constructed and maintained on the Master Control System and then used to track and control the flow of questionnaires through all subsequent steps.

In contrast to 2001 when questionnaires proceeded through collection and processing steps EA by EA, a flow processing approach was used where from the point of delivery forward each questionnaire moved individually through all subsequent steps of processing.

Except in the canvasser parts of the country, respondents had the choice of responding by Internet or by mail. In either case the response was directed to the central DPC in Ottawa. As in prior censuses, respondents had the option of calling CHL to provide their data; for 2006 these data were collected by CATI.

The use of mailout delivery, flow processing and the availability of the Internet response channel responded to a number of the drivers for change. A much smaller number of enumerators – about 20,000 – was needed, not only for questionnaire delivery but also for follow-up activities. The use of flow processing and the Internet response option created an important degree of flexibility and efficiency in operations. This required development of the Master Control System to ensure proper control under these operational conditions.

Upon receipt at the DPC questionnaires, whether paper or electronic, were first registered on the MCS as received. Next, paper questionnaires were scanned and data capture was done using image recognition technology with data repair via key entry. Paper questionnaires and their images were then stored. Subsequent processing, whether automated or involving human intervention, proceeded using only the data taken from the questionnaires. The paper questionnaires were not further used although their images were referred to on an exception basis.

The requirements of the multiple channels of questionnaire delivery and response as well as the requirements for scanning and image recognition data capture raised many new issues of questionnaire design and mode effects. These are discussed by Roy (2006).

Although human intervention was still needed, editing was much more automated than in the past. Data for failed edit questionnaires were forwarded to regional call centres (the CHL sites) for telephone follow-up via a CATI application. With this centralization, local enumerators were no longer used for editing and follow-up for edit failures thus responding to respondent concerns about having their data seen by a local enumerator. When completed the data were returned to the DPC. This process of dispatch and receipt for Failed Edit Follow-up (FEFU) again was controlled using the MCS.

Non-response follow-up was initiated about ten days after census day (May 16, 2006). Lists of non-responding dwellings were extracted from the MCS and forwarded electronically to local offices. Non-response follow-up was conducted first by telephone and then if necessary in person. Completed questionnaires were forwarded to the DPC by way of the local office, checked in on the MCS and then proceeded through data capture and subsequent processing like other questionnaires.

The MCS then served to track the data for each individual dwelling through the remaining processing steps. As questionnaire data passed through all stages of editing and coding, they were added progressively to a Response Data Base (RDB). After all data had been loaded to the RDB, they were transferred to a final database where imputation and weighting will take place.

The process just outlined in these few paragraphs is described in more detail in Statistics Canada (2005).

2. Dwelling Frame and Master Control System

By the summer of 2005, Statistics Canada's Address Register included about 89% of residential addresses as of 2001 in Canada. It also included a high but unknown percentage of new addresses since that time. Of this about 70 percentage points were usable for mailout. The unused portion consisted mainly of areas where the dwelling coverage was considered inadequate or where too many of the addresses were not valid for use by Canada Post. In preparation for mailout a number of additional steps were taken.

A process of address validation for the mailout area was undertaken. Registers of known addresses were printed and, in fall 2005, a process called Block Canvassing was conducted by enumerators whose job was to update these lists. Each was responsible for an area including about 7,000 dwellings. The task consisted of matching the dwellings listed in the register to dwellings in the field and making any necessary additions, deletions (deleted dwellings were not actually deleted but were flagged as inactive) and corrections to attributes and geocoding. A second step of Late Block Canvassing was conducted in February/March 2006 in areas identified during regular Block Canvassing as being subject to high rates of growth. Dwelling lists were updated to include not only actual dwellings but also dwellings under construction that had the potential to be available for occupancy by Census day.

These updated lists, including the inactive records, became the dwelling frame for the mailout area. Each address was loaded onto the MCS and given a unique identification number called a frame-id.

For the list/leave and canvasser areas, the dwelling frame was initialized in a different fashion. Each collection unit (CU, called an EA in 2001) was provided with a suitable quantity of frame-ids on the MCS, each corresponding to a specific line in the Visitation Record for the CU. After the enumerator listed all the dwellings in the VR at the time of questionnaire delivery, he or she reported back to the MCS the specific line numbers that had been used so that remaining ones could be flagged as inactive.

The MCS then contained a complete initial dwelling frame that could be dynamically updated during operations and which was used throughout for centralized control of the entire process. It was expected to consist of about 91% occupied dwellings, 7% unoccupied dwellings and 2% non-dwellings (e.g., demolished, unfinished construction, unoccupied marginal dwellings, buildings previously misclassified as dwellings).

3. Questionnaire Delivery and Response Channels

Using this initial dwelling frame on the Master Control System, the printer prepared mailout packages for every mailout address. Each questionnaire was preprinted with the address, frame-id, a corresponding barcode and a unique Internet access code. These were shipped to Canada Post for delivery between May 1 and May 5, 2006.

Questionnaires for list/leave and canvasser areas were also printed using information taken from the MCS. A questionnaire was printed for every frame-id for every CU in these areas and each one was preprinted with the frame-id, corresponding barcode, the CU number and the Visitation Record line number to which it was to correspond as well as its unique Internet access code. They were shipped to the local census offices for use by enumerators. List/leave took place over the first two weeks of May while canvasser data collection was spread over a more extended period. Each dwelling in the enumerator's CU was to be listed in the VR. It was critically important for enumerator to then drop off at each address the questionnaire corresponding to the Visitation Record line on which its address was written. Completion of the VR established the link between each address and the frame-id.

All dwellings outside the canvasser area were informed they could respond by Internet or by completing and

mailing the paper questionnaire; it was their choice. However, a strong focus of all communications materials was to encourage response via Internet. Completed questionnaires from canvasser areas were delivered by enumerators first to the local census offices from where they were then shipped to the DPC in Ottawa.

The Internet response channel was made available for use effective May 1, 2006. Dwellings choosing to respond by Internet entered their Internet access code and then, after establishment of a secure connection and confirmation that their computer met the necessary technical requirements, the specific questionnaire (short or long) as assigned to the dwelling could be completed.

In addition to canvasser, mail back and Internet, the CHL provided an important fourth response channel; it served several purposes in this regard. First, for large households responding on the paper questionnaire, CHL could be called for delivery of an additional questionnaire. Second, respondents needing the questionnaire in the other official language or in an alternate format could call CHL. Third, as noted earlier, households with special concerns about privacy of their data could call CHL to provide their data directly via CATI. Fourth, individuals not wanting to be included on the household questionnaire could call CHL for delivery of an individual questionnaire or to provide their data directly via CATI. Fifth, dwellings that either did not receive or had lost their questionnaire could call to request a paper questionnaire, an Internet access code or to provide their data directly.

In every one of these cases, an important concern was to ensure correct linkage of the questionnaire to the dwelling and the MCS. For callers who could provide the frame-id from their original questionnaire this was straight forward. All others were first asked to provide address or location information and, using a simple address search tool, the CHL operator attempted to find the corresponding frame-id. If the dwelling address could be easily located on the MCS by the CHL operator then the questionnaire was successfully linked. These responses could be submitted by direct response to the CHL operator (CATI) or by Internet using a proxy access code linked to the frame-id or on a paper questionnaire. When completed, these linked questionnaires would all proceed to the DPC where receipt and registration would take place.

However, in cases where the dwelling was either not in the mailout area or had been missed or was new, the address would not be on the MCS and no link would

exist. In cases where the address information provided was vague or otherwise of poor quality no link would be found even though one might exist. In such situations a temporary or proxy-id was assigned to the address and was used to accept the response. If the response was to be by Internet then a proxy Internet access code would also be issued. In such cases, linkage was to be determined later via the process outlined in section 4.2.

Finally, although not really a response channel per se but still constituting a fifth means of data collection, respondents to the long questionnaire had the option of giving Statistics Canada permission to link to their income tax data rather than responding to the set of income questions.

4. Questionnaire Receipt and Registration

To be considered as usable by the census, each questionnaire received at the DPC had to be directly associated –"linked"-- with a frame-id on the MCS dwelling frame. As each questionnaire, whether electronic or paper, with a valid frame-id was received at the DPC it was checked in and flagged as such on the MCS thus making this link. For Internet and CATI responses this occurred automatically as the electronic data arrived at the DPC. For paper questionnaires the barcode on the questionnaire was scanned in order to identify and check it in. Those questionnaires with a proxy-id or which used a proxy Internet access code would not immediately link to the MCS and proceeded to an unlinked questionnaire queue to await resolution; see section 4.2. In some cases, multiple questionnaires were received for the same frame-id. Such situations were tracked not only via the MCS but also via a questionnaire control system (QCS) used at the DPC and proceeded to a special process for their resolution; see section 4.1.

After receipt and registration all paper questionnaires, tracked by the MCS, proceeded to data capture. Most were scanned and data were captured via image recognition. Data capture for those that had been damaged or which otherwise could not be scanned was done by key entry. The quality of each questionnaire image and of the capture of each field was assessed automatically and those which failed went to a step of key entry data repair from the paper questionnaire, its image or image snippets, as appropriate, was done.

4.1 Resolution of Multiple Questionnaires

Via the Questionnaire Control System, the quantity of questionnaires registered for each frame-id was tracked. For most dwellings there was only one

questionnaire. However, for some there were more and often this was valid. Examples include dwellings with large households where more than one questionnaire was needed and dwellings with at least one person responding using an individual questionnaire. There were also a number of invalid situations such as: duplicate questionnaires providing information for the same persons; partial duplicates where the smaller questionnaire reports only for a subset of the persons on the larger one; and multiples where an unlinked questionnaire has been linked to an incorrect frame-id.

Most valid multiples were identified automatically and required no special processing. For remaining multiples the resolution process was a computer assisted clerical operation requiring reference to data from all questionnaires for the frame-id concurrently.

Valid multiple questionnaires usually arrived at the DPC together and were logically combined for subsequent processing. Occasionally however, whether valid or not, they did not arrive at the same time. Once a second questionnaire was registered to a frame-id, all related questionnaires were sent to the holding queue either immediately or after scanning if not already scanned. When all had arrived, the resolution process could be initiated.

For invalid situations there are a number of possible resolutions. In the case of duplicate questionnaires, including partial duplicates, usually the one with the better quality data will be retained and the other one deleted (it is not physically deleted but is flagged as such). When one of the questionnaires is determined to be an incorrect linkage, it is unlinked and sent to the process for unlinked questionnaires while the remaining questionnaire is retained.

In every case the presence of the multiple questionnaires and the resolution of the situation are indicated on the MCS.

4.2 Resolution of Unlinked Questionnaires

As explained in section 3 some calls to the CHL resulted in unlinked questionnaires where a proxy-id or proxy Internet access code had to be used. In this situation the CHL operator could take the response via CATI using the proxy-id or, to respond via Internet, the respondent could be provided with a proxy Internet access code. Last, the respondent could be provided with a paper questionnaire. In each case the address and/or location information and the proxy–id were sent to a linkage process to determine the appropriate frame-id.

For all three response modes, completed questionnaires proceeded to the DPC and were placed in an unlinked questionnaire queue to await the linkage outcome.

In a first step of this process, appropriate local enumerators were given the address information and had the responsibility for resolution of the case. They were to use Visitation Records, non-response follow-up lists, other records and field visits as needed. Possible outcomes included confirmation that the dwellings was already on the frame and identification of the linkage to the frame-id or confirmation that the dwelling should be birthed to the frame or a conclusion that the case could not be resolved and so would remain unlinked.

Once these proxy-id to frame-id links and births were communicated to the MCS, the questionnaires in the unlinked queue at the DPC could be released to the registration process noted earlier at the beginning of section 4.

5. Follow-up

5.1 Follow-up for Failed Edits

Following data capture at the DPC, each questionnaire was edited for completeness and coverage. A score function (Bornais and Boudreau, 2003) summarizing the severity of edit failure was used to identify those to be sent for follow-up with respondents. Interviews using CATI were conducted from the CHL sites, thus avoiding problems with privacy encountered in previous censuses with the use of local enumerators at this step. As with other operations, the MCS was used for tracking and control. Edit failure, dispatch for failed edit follow-up and receipt at the DPC of completed follow-up data were all monitored via the MCS.

5.2 Follow-up for Non-response

Non-response follow-up (NRFU) was initiated about 10 days after census day. A list of frame units for which a response had not yet been registered was extracted from the MCS and made available to the NRFU control system. This list in turn was used at local census offices for dynamic creation of assignment lists for enumerators. Unlike past censuses, supervisors were not constrained by CU boundaries for NRFU; assignments could be constructed more optimally and completed questionnaires could proceed immediately for processing without waiting for completion of the entire CU. Each case was to be resolved by identifying whether, on census day, it was unoccupied or not a census dwelling and by obtaining

a completed questionnaire if it was occupied. In a first step enumerators attempted telephone contact with these dwellings and then, as necessary, made field visits.

The first phase of NRFU was to determine the status of each non-respondent dwelling as being occupied, unoccupied or not a census dwelling. The MCS and the NFFU control systems were updated to indicate as such.

The second phase was of course to obtain a completed questionnaire from each occupied dwelling. These responses could again be returned to the DPC – where they would be registered as received on the MCS – by three means: mail or Internet by the respondent, or by shipment of enumerator completed questionnaires. The NRFU control system was to be updated to show these outcomes: response expected (i.e., the respondent had promised to submit a paper or Internet questionnaire), interviewer completed questionnaire, no contact, etc.. Concurrently some responses were submitted independently without reminder by a NRFU enumerator.

An essential part of controlling the NRFU operation was feedback from the MCS to NFRU enumerators about cases that were on the non-response follow-up list for which a questionnaire had now been received. This was particularly important for questionnaires not completed by NRFU enumerators – "response expected" cases and independent responses. The non-response follow-up system was updated to indicate receipt of such questionnaires so that field work on these cases could be stopped. Then at local offices it was necessary to re-issue assignment lists or manually update existing assignment lists.

6. Subsequent Steps

The MCS was used as well for tracking of each questionnaire through typical subsequent census processing steps. This included coding of several variables and the final steps of editing where failures were dealt with via imputation. Variables requiring coding were on the long questionnaire only and included a number of socio-cultural variables, place of work, industry and occupation. Coding for each variable involved a step of automated coding followed by computer assisted clerical coding for the residual. As a final step of processing tracked by the MCS, upon completion of all processing steps all finalized data as well as the original response data and a substantial amount of related paradata were moved from the processing environment onto a response database. Later, all finalized data for all active frame-ids (those

flagged on the MCS as corresponding to a census dwelling whether occupied or vacant) will be copied to another database where imputation (Bankier et al. 1999) and weighting (Bankier et al, 1997) for the 20% long form sample will be conducted.

7. Some Results and Initial Evaluations

In this section some initial evaluations and figures are given on the use and impact of the various response channels and the quality of these responses. However the reader is cautioned though that all figures cited here are preliminary, as they are based on a process that is not yet complete and whose evaluation is only just beginning at the time of writing this article. Some brief comments are given on the success of the MCS in maintaining coordinated control of operations.

For the most part the multimode data collection for the 2006 Canadian Census of Population has worked well. Acceptance of response by Internet was excellent. Compared to mail, responses were received faster and were of better quality. CATI systems were used for both the CHL and for FEFU and are expected to yield quality improvements compared to previous paper and pencil methods.

The MCS was successful as an easily updated dwelling frame. Feedback from enumerators following list/leave and canvasser operations was simple and, in most cases, timely. The process for feeding back to the MCS identification of unoccupied dwellings and non-dwellings identified during NRFU worked well technically but feedback from enumerators was often slower than expected. Enumerators appeared to delay completion and return of the related forms in favour of focusing on obtaining completed questionnaires. Additions to the frame on the MCS – previously missed dwellings – usually arose first as a subset of the unlinked questionnaires. The planned process for resolution of these unlinked questionnaires, outlined in section 4.2, did not work as well as expected and a number of unplanned initiatives were undertaken that in the end were more successful in resolving the unlinked questionnaires.

As a tool for coordination and control the MCS was very successful. Operations at the Data Processing Centre worked very smoothly in part due to the coordinating role of the MCS. A specific example is multiple questionnaire situations where the process described in section 4.1 worked very well and the MCS without fail identified these situations and played a key role in their rapid resolution -- important since many of these were not valid. There were multiple questionnaires for about 1.9% of dwellings. Of these

about 58% were valid – dwellings with large households or with one or more individual questionnaires. Another 38% were invalid because the questionnaires for the frame-id were duplicates. Finally, about 4% were invalid due to incorrectly linked questionnaires.

A second important example is in the resolution of unlinked questionnaires. Of responses arising from calls to the CHL, many were linked immediately by CHL operators as part of the planned linking process, involving direct reference to the MCS dwelling frame. However, many more – in the order of 250,000, far more than expected – resulted in unlinked questionnaires. Remaining steps in the planned process were neither as effective nor as timely as expected in resolving these. Instead, a number of initiatives were undertaken at head office. These involved a combination of both automated and clerical processes which matched the respondent supplied addresses to those on the MCS and to other geographic files. To identify possible duplicate responses data from these unlinked questionnaires were matched to data for all responses that had already been linked. Depending on the quality of the address information and the response data various resolutions were possible: linkage to a specific frame-id, confirmation as a birth to the frame, linkage to a CU, linkage to a set of CUs, identification as a duplicate response or no linkage identifiable. Over 70% were in the first category and thus could be added to the census database; those in the second category could also be added to the database. There were much smaller proportions in the other categories. At the time of writing this article, strategies were being investigated for addition to the database of the cases that could be linked only down to a CU or set of CUs.

The MCS identified and provided the initial list of dwellings requiring NRFU to local census offices. Following this, the smooth flow of information between NFRU systems located in these offices and the MCS was critical and this went very well technically. However the flow of this information, particularly that concerning identification of unoccupied dwellings and non-dwellings, from NRFU enumerators to local offices and the NRFU control system was much slower than planned. Further, the flow of information to local non-response follow-up enumerators concerning receipt of questionnaires appears to have too often been too slow. This may have contributed to the 38% of multiple questionnaire situations being duplicates, a much higher proportion than had been expected. Slow feedback to enumerators of information concerning receipt of questionnaires may have contributed to them sometimes collecting

information from dwellings for which a questionnaire had already been received .

At the start of NRFU, questionnaires had been received and registered on the MCS for about 74% of the expected number of occupied dwellings. This was somewhat below the planned figure of 78% leaving a considerably increased workload for NRFU beyond that planned. Of these responses about 21% were received via Internet, exceeding the planned rate of 20%, a major success.

Although during NRFU most responses were field completed, the availability of multiple response channels proved valuable as responses continued to be received via all of them. Their distribution by response channel is shown in the following table. Field completed includes both responses from canvasser areas as well as enumerator completed questionnaires during NRFU.

Response Channel	Share of Responses (%)			Percent long form
	Pre-NRFU	NRFU	All	
CHL	0.3	5.7	2.0	20.7
Field completed	0.3	56.3	17.0	18.6
Internet	21.2	9.3	17.9	20.5
Mail	78.2	28.7	63.1	19.4

As in past censuses, non-response was slightly higher for the long questionnaire than the short. The former should make up 20% of responses but in fact account for only 19.5%. The above table shows this percentage by response channel. The differences are small and are influenced by both respondent and interviewer behaviours, but they may suggest that dwellings with the long form were more likely to respond by Internet than those with the short form.

In terms of failure rates for failed edit follow-up, data quality is better than planned. As expected, that for Internet response is much better than that for mail responses. In fact, the extent to which it is better is substantially greater than expected. However, more studies would be needed to know how much of the increment in quality is directly related to the collection medium.

Failure Rates for FEFU (%)				
Questionnaire Type	Actual		Planned	
	Mail	Internet	Mail	Internet
Short	5.7	2.5	6.8	4.0
Long	39.1	5.4	45.0	12.0

Overall about 85% of respondents (over 14 years old) receiving the long questionnaire granted permission to access their tax records rather than respond to the income questions. This should result in a substantial improvement in the quality of the income data. Rates by response channel are shown in the following table and are slightly higher for questionnaires completed via the CHL. Permission rates show very little variability by geography or sociodemographic characteristics although rates are slightly lower for persons who are very young, never married, living alone, or whose mother tongue is neither English nor French.

Response Channel	Tax permission=yes (%)
CHL	89.3
Field completed	85.0
Internet	85.7
Mail	84.7

8. Conclusion

The data collection methodology for the 2006 Canadian Census used multiple methods of questionnaire delivery and multiple methods of questionnaire return for initial response from dwellings and again for non-response follow-up. Compared to 2001 many activities were centralized and automated. In this environment, the Master Control System played a critical and successful central role in coordinating and controlling all census operations from construction of the initial frame through to finalization of the data. It facilitated a more efficient and flexible collection methodology that was no longer constrained to work CU by CU but rather allowed each questionnaire to proceed largely independently through data collection and processing.

Acknowledgements

References

Bankier, M., Houle, A.M. and Luc, M. (1997), "Calibration Estimation in the 1991 and 1996 Canadian Censuses", Proceedings of the Survey Research Methods Section, American Statistical Association, 66-75.

Bankier, M., Lachance, M. and Poirier, P. (1999), "A Generic Implementation of the New Imputation Methodology", Proceedings of the Survey Research Methods Section, American Statistical Association, 548-553.

Bornais K., Boudreau J.R. (2003), "Conclusions of the Edit and Follow-up Study" Internal Report. Social Survey Methods Division. Statistics Canada.

Roy, L. (2006), "The Internet Response Method: Impact on the Canadian Census of Population data", Proceedings of the Survey Research Methods Section, American Statistical Association.

Statistics Canada (2005), 2006 Census Process: Methodology. Census Management Office working paper.