

Hanging in the Balance: Using Technology Without Losing Touch in the Canadian Census of Agriculture

**Claire Bradshaw, Processing, Census of Agriculture
Catherine Cromey, Manager, Census of Agriculture**

Agriculture Division, Statistics Canada

Abstract: In recent censuses, the Canadian Census of Agriculture has adopted new technologies in data collection and processing. While these technological innovations have made collecting and processing the agricultural census faster and have provided many tools to ensure data quality, they carry a price. The push to more cost-effective collection methodologies has coincided with increased demands for privacy and confidentiality; in combination, the two have decreased the quantity and quality of personal contact with respondents.

Each five-year census presents new challenges in ensuring that the agriculture census addresses the important issues of a very rapidly changing agriculture industry. The collection process, although still largely paper-based, has incorporated Internet and computer-assisted telephone interviewing, limiting opportunities for ad hoc feedback and face-to-face contact. This paper presents the issues and solutions for balancing respondents' privacy and confidentiality, and the growing adoption of more impersonal and fragmented data collection options.

1. Introduction to the Census of Agriculture

The Census of Agriculture is a comprehensive stock-taking of Canada's agricultural industry every five years. The five-year cycle allows us to track the evolution of farms and the ways they are adapting in order to stay profitable. It also allows for the timely identification of trends in a rapidly changing industry, while the small area data provide the flexibility and precision that make this possible.

The Census of Agriculture retains a strong time series on core questions while maintaining some flexibility for adding and revising questions that reflect current issues in the industry. The census collects data on farm operators, crops, livestock, land management practices, farm finances, machinery, and computer use. The last census in 2001 counted approximately 247,000 farms, 167 million acres of total farmland, 101 million acres in crops, 15.5 million cattle, 14 million pigs, and 126 million hens and chickens.

For the past 50 years the Census of Agriculture and Census of Population have conducted a joint data collection operation. Since the majority of agricultural operations are household-based, they can be effectively and accurately enumerated using the same methodology. Preparations are well underway for the next census in May of 2006, which will continue this tradition. The joint collection also facilitates a link between the

agriculture and population databases that provides a detailed socio-economic profile of the farm population.

This paper focuses on the changes in census collection methodology and coverage erosion that have compelled the Census of Agriculture to find additional avenues for collection as well as new strategies to enhance that collection. The role that technology plays, the third component in this balancing act, is discussed at the end of the paper.

The paper is divided into five sections: this section has set the background, while the second explains the joint census collection process and describes the changes for 2006. Section 3 explains past changes to collection procedures for the Census of Agriculture and the impact of those for 2006. Section 4 describes the various strategies that will be put in place to compensate for coverage erosion and looks at the role technology plays. The fifth section offers concluding comments.

2. The joint collection process for agriculture and population

Understanding the census collection process used in the past and the extent of the changes coming in 2006 is critical to explaining some of the challenges the Census of Agriculture is facing.

Between 1971, with the start of self enumeration, and 2001, the census data collection methods for the joint agriculture and population censuses changed little. Most households (about 98%) were enumerated by the drop-off and mail-back methodology. A Census of Population questionnaire was dropped off at each household in the assignment area and whenever the census enumerator determined that a farm operator resided in the household, a Census of Agriculture questionnaire was also delivered. Respondents were instructed to complete the questionnaires and forward them in the enclosed postage-paid envelope. A census enumerator was responsible for delivering, receiving, editing, following up, and meeting quality standards for an assigned area.

The mailed-back questionnaires were returned to the same census enumerator that had delivered them, who then edited the forms and conducted a telephone follow-up for any questionnaires that failed edit. When dropped-off forms were not returned by mail within a certain period of time, the enumerator was required to follow up. After all questionnaires for the enumerator's area, both population and agriculture, were completed, the entire assignment was moved up the line. Once quality checks were concluded, the Census of Agriculture questionnaires were removed and sent to processing operations at agriculture head office for imaging, automated data capture, and further processing.

The 2006 Census collection represents the largest change in data collection methodology in more than 30 years. Several reasons are behind these changes, but topping the list are concerns for respondent privacy and confidentiality as well as the potential for cost savings. The availability of new technologies, such as the Internet, has also played an important role in the decision.

While most rural households will still have questionnaires dropped off at the door in 2006 as in previous censuses, the enumerator's edit and follow-up for edit failures will be eliminated. As well, having all completed questionnaires mailed directly to a central data processing centre will restrict the local enumerator's involvement to delivering the questionnaires, a significant departure from the previous census. Having a neighbour see a completed form has been a long-standing issue among rural respondents.

Once the mailed-back form has arrived at the data processing centre, the unedited questionnaire will be imaged, and the data captured using intelligent character recognition technology. This marks the first time the two censuses will share data capture operations. At this point the data and questionnaire images will be sent to the agriculture head office to start processing. While eliminating the enumerator edit and follow-up means that questionnaire data will have more missing entries and consistency errors, the forms will arrive faster than ever before.

3. Collection and coverage challenges in 2006

The Census of Agriculture had already seen coverage erosion in previous censuses, and the new 2006 Census collection process will add new coverage challenges.

3.1 Changes to Census of Agriculture collection prior to 2006

The Census of Agriculture first initiated a change in the standard collection process as a result of the trend to corporate farms. A special collection, with contact at the corporate business address, was formally implemented in 1986. This one-time special collection has since evolved. Today a unit of people dedicated to this task continually updates and maintains profiles on large corporate farms for both the Census of Agriculture and the regular survey program. These operations are contacted and profiled prior to Census Day and arrangements are made to collect census information. This unit's work has grown from about 25 businesses in 1986 to over 300 corporate contacts.

In the 1996 Census, changes to collection dates and a new reference date implemented by the Census of Population meant another change in agriculture's collection procedures. The traditional reference date of the first Tuesday in June was moved to the second Tuesday in May, so that questionnaires could be delivered and completed within the same month. Although a benefit for the Census of Population, (households that move, usually at the end or beginning of a month, would not be as easily missed), this change had an impact on the field crop areas reported by farm operators.

Much of the field crop seeding across Canada typically occurs before the first of June. Depending on the spring weather, a large portion of the crop may not be seeded when respondents complete their forms in May. In response, the Census of Agriculture implemented the Progress of Seeding (POS) Survey. The first post-collection POS survey to verify and update crop data was conducted from Statistics Canada's Regional Offices in 1996 as a computer-assisted telephone interview (CATI) survey. It involved

following up with operators who reported less than 90% of their field crops seeded when they completed their forms. In 1996, a late spring in many parts of Canada, approximately 115,000 were contacted; in 2001, a relatively early spring, only 44,700 farm operators were surveyed. The resulting updates were later re-integrated to the database.

The next collection change was a result of difficulties census enumerators were experiencing during collection. Making contact at the door at drop-off had become less frequent as more farm operators increasingly worked off the farm. Moreover, increasing numbers of farm operators were living off their operations, making it more difficult for the census enumerator to determine that an agriculture questionnaire was required. Trends such as the declining number of farms, the declining percentage of farm operators among the rural population, and an increasing number of farms having a non-traditional appearance have all contributed to this undercoverage.

Starting with the 1996 Census, a missing farms follow-up survey was implemented to collect data for farms that were missed by census enumerators. Once agriculture questionnaires were received from field collection, they were matched to a database of existing farms and larger farms considered missing were contacted through a CATI survey. The first missing farms follow-up in 1996 found an additional 3,000 farms. In 2001, about 5,000 farms were added. Their updated information was then integrated into the census data before publication.

3.2 The 2006 Census collection picture — more fragmented than ever before

The elimination of census enumerator field edits for 2006 has added an additional layer to the coverage challenge. In the 2001 scenario, the census enumerator was responsible for dropping off all population and agriculture forms as appropriate, then editing the forms and following up with respondents for failed edit and non-response — all to ensure the complete enumeration of “their part of Canada.” Now the Census of Agriculture has had to find a different way to complete this work. Two separate edit operations are affected by the change.

One is the edit to the agriculture operator screening question on the Census of Population questionnaire. This agriculture screening question allows farm operators to self-identify on the population census form to ensure that farm operations missed at drop-off receive a questionnaire. As a field aid to identify farm operations, the screening question was a very effective tool for identifying undercoverage since the enumerator could quickly deliver any missed agriculture forms. It also worked well to eliminate overcoverage, as local enumerators generally knew the people in their area and were able to effectively weed out “false-positive” responses and, more importantly, confirm and cancel agriculture forms dropped off in error. The scope of these overcoverage adjustments reduced the number of questionnaires by about 30,000 across all of Canada.

In 2006, the two-part screening question on all population forms asks, "Is anyone (in the dwelling) a farm operator who produces at least one agricultural product intended for sale?" and "Does this farm operator make the day-to-day management decisions related to the farm?" The second part of the question was added for 2006 to reduce the false-positive responses and the cost of follow-up. Potential new operators identified by this question but who do not return an agriculture questionnaire will be followed up at the same time as the missing farms follow-up survey. Now that this edit will be centralized and automated, sophisticated logic trails will target those most likely to be farm operators, for example those in rural areas, and weed out false-positives even before follow-up begins. We expect the 2006 operation to contact 15,000 population respondents who indicated that they operated a farm and identify about 1,500 new farms.

The second editing process concerns the series of edits completed by the census enumerator on the agriculture form. In the past, these edits were usually completed within several weeks of the respondent completing the form. Generally the edit ensured completeness of the questionnaire and allowed a cleaner data capture. Over 55% of the agriculture questionnaires failed the completeness edit and required a follow-up call by the census enumerator. In 2006, the questionnaire will arrive at a central processing centre just as they were completed by the respondent.

The 2006 version of the former enumerator edit of the agriculture questionnaire will also be conducted as a centralized follow-up CATI survey. The move to head office presents new opportunities to make the process more efficient: the edits previously done by the census enumerator will now be automated, meaning the edit can be restricted to the most important or most severe edits and focus more on the larger farm operations, while including a sample of medium and small farms. The Progress of Seeding survey can also be incorporated into the failed edit follow-up. The switch from a large decentralized field staff in 2001 to a smaller centralized and specialized interview staff means we can move more complex edits that would have been performed by agriculture specialists in later edit and validation processing to an earlier process. Questionnaires not selected for follow-up will go to imputation for edit failure corrections. All these features will also reduce respondent burden.

One drawback to the new design is timing. The 2006 CATI failed edit follow-up for agriculture cannot be completed until after the data are captured, edited for basic logic errors, and an automated edit completed to identify candidates for the follow-up survey. Although farm operators will be re-contacted for failed edit as soon as possible, a delay of at least one month and up to two months will be needed to complete all steps necessary to identify those respondents requiring follow-up.

The end result for the Census of Agriculture will be a data collection process that is more fragmented than ever before. The single agriculture census collection vehicle used in the 1971 Census has been replaced by a complex, multi-mode collection operation. Besides basic collection through a paper questionnaire or the Internet, in 2006 it will include four independent collection operations: a special universe collection

for corporate and large farms, and three post-census CATI follow-up surveys — the missing farms follow-up, the follow-up to the population agriculture operator screening question, and a failed-edit follow-up for completed census questionnaires that fail automated edits.

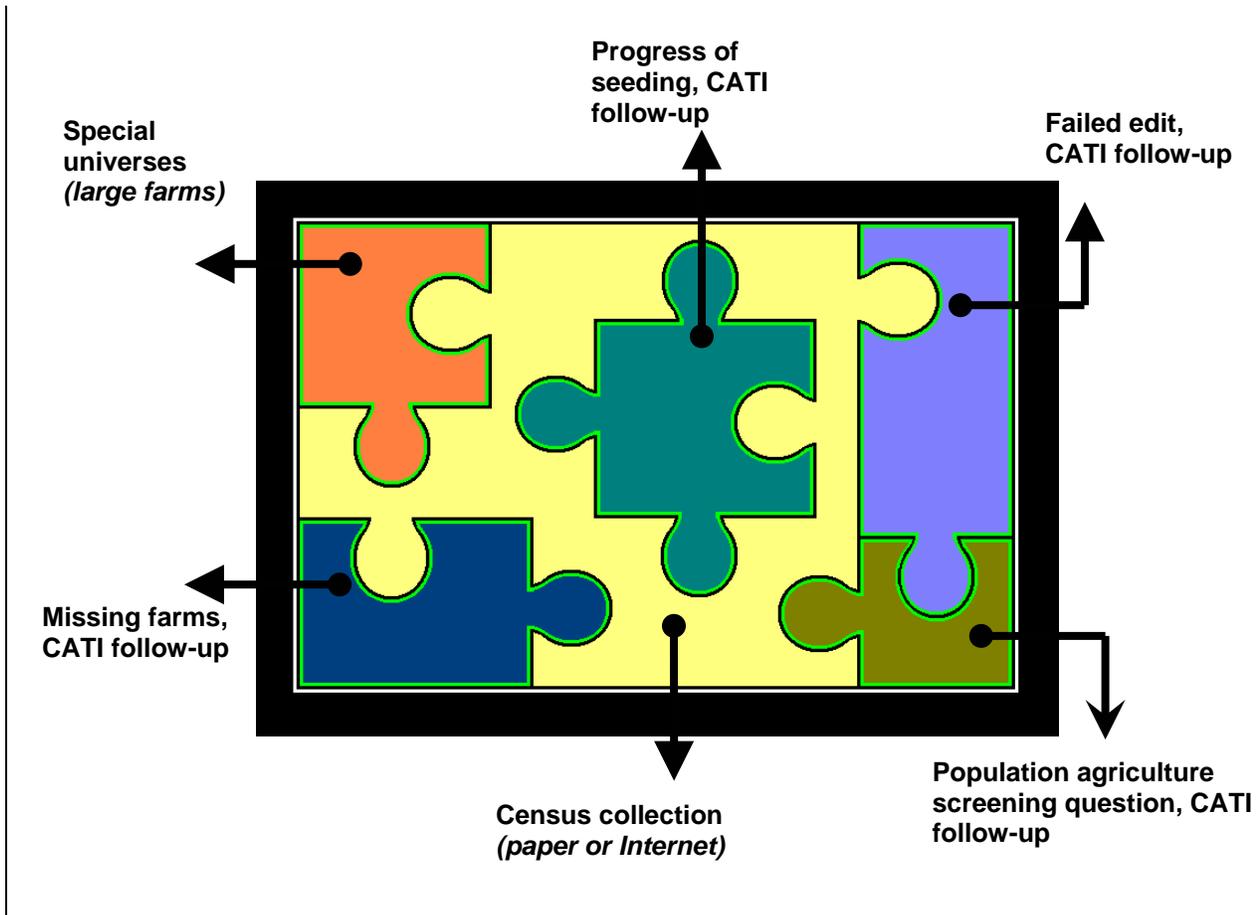


Figure 1: Collection is more fragmented than ever before.

4. The 2006 challenge — integrating the collection processes

Faced with the reality of increasingly fragmented data collection schedules and a variety of collection modes, the Census of Agriculture looked for ways to integrate these new processes while maintaining coverage, data quality and release dates. With data for both population and agriculture to be captured earlier and faster, processing flows were examined to see how the new work could be incorporated and the process made more efficient. This section highlights the new processes and strategies that the Census of Agriculture has adopted to answer the collection challenges, particularly the increasing dependence on the Farm Register and the two new edit follow-up surveys. Finally, the various technologies that have made a difference are discussed.

4.1 Increased dependence on the Farm Register

The Farm Register, a central database that stores administrative data on all farms in Canada, has become a key coverage tool for the census and plays an important role in processing the data. After each census, the Farm Register undergoes a complete update and in the years between censuses receives updates according to information received from agriculture surveys and tax data.

At the outset of the Farm Register's development the census was an important data source. While this relationship remains today, it has evolved to a point where the Farm Register is also an important list frame source for the Census of Agriculture. Investments in maintaining and updating the Farm Register have become very important as its role in census collection and processing increases.

Census records are matched continuously to the Farm Register as the data files arrive in head office. Various processing operations use the matched information, as these examples illustrate:

- A list of farms from the Farm Register is pre-ranked by size according to criteria such as land area or farm receipts. The ranked list is used by the follow-up surveys to ensure complete coverage of the largest farms and a representative sample of the smaller farms. This also allows for a more cost-effective follow-up.
- The imputation and validation processes use matched data to confirm and assess the accuracy of information collected on the Census of Agriculture.

Several activities are underway for the 2006 Census to ensure that the Farm Register is as up-to-date as possible and free from duplicate operations and those no longer in business. The annual Farm Update Survey calls approximately 10,000 operations that have not been sampled in recent surveys to confirm their status. Another 1,000 records identified by income tax data as "likely" farms are also included. In 2005, a list of agricultural producers collected from various external sources not already on the Farm Register will be added to the Farm Update sample.

4.2 Technology and the Census of Agriculture

For those who have witnessed many censuses, the improvements generated by technical innovations at each five-year census interval have been dramatic. The range of tools available for data processing and data validation has made the work in each successive census proceed faster and with more accuracy. Within the balancing act of the various collection activities, these technologies allow us to control and monitor all components, keep an exact trail of all changes, make sure the best information is there when needed, automate processes to meet tight timeframes, and focus our efforts on the most important errors or contributors. Some key technology tools that have served us well are discussed in this section, including our central processing system, imaging technology, automated data capture systems, and data validation tools.

In 2006, the base for co-ordinating and integrating all processing is the Central Processing System (CPS). The system consists of a network of processes through which all records must pass. Unlike the fragmented data collection process, the single processing environment provides a unifying central control. To support the various processes, scanned questionnaire images, Farm Register data, 2001 Census data, electronic maps, and a number of other reference sources are available online. All are quickly available through the CPS interface on each employee's desktop.

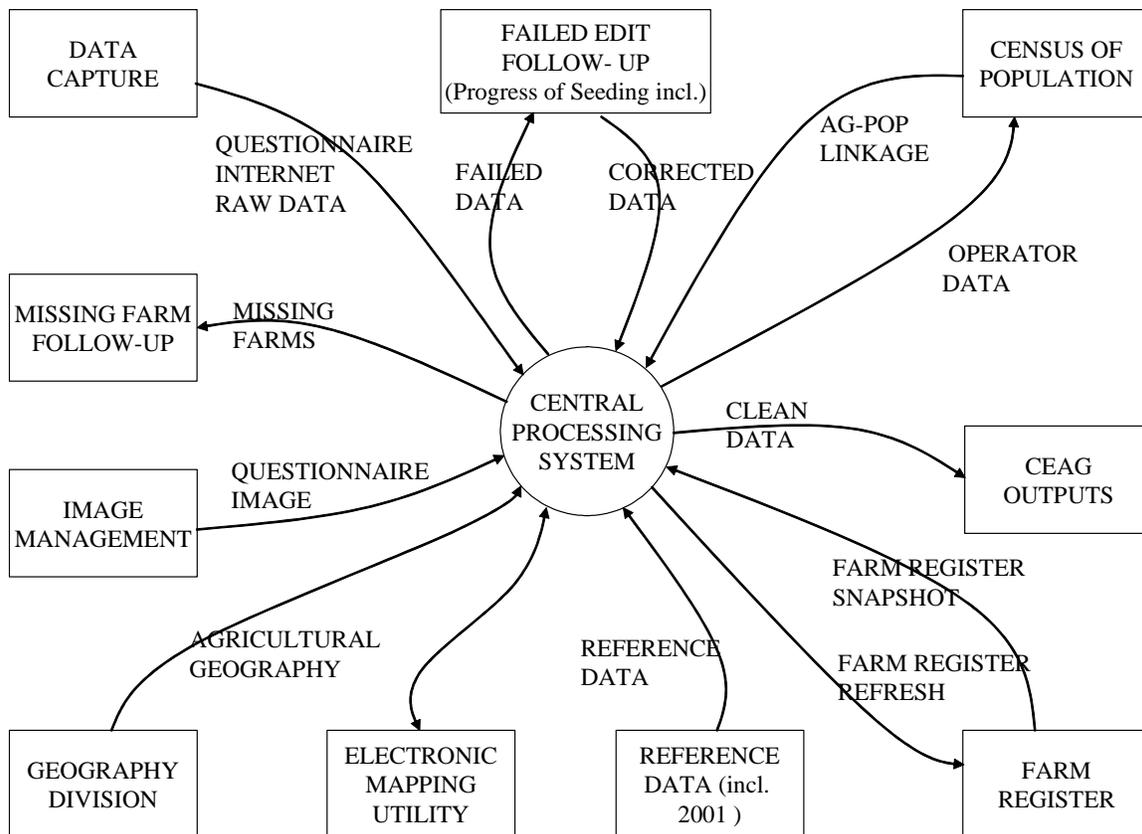


Figure 2: The Central Processing System unifies all data sources and supporting systems throughout the production cycle.

The addition of questionnaire images in the 1996 Census was a milestone in census processing in terms of major improvements in efficiency and accuracy. Questionnaire images were made available to all the processes that had required access to paper questionnaires in the past. The 1996 innovation had many benefits, including greatly reducing the amount of paper handled and completely eliminating the need to track and control questionnaires. Several processing operations witnessed considerable time savings and quality gains by having images quickly available. Most importantly, imaging allowed more than one analyst to view a questionnaire image simultaneously. In agriculture large farming operations make significant contributions to many data variables and these questionnaires are in frequent demand. Again in 2006, the process

will permit instant access to an electronic image of the questionnaire in order to check the accuracy of capture, review forms that fail automated edits and look for any explanatory notes the farm operator may have written.

Since imaging is the enabling technology for automated data capture, this step was implemented in the 2001 Census. The questionnaire was redesigned for automated data capture technology with segmented boxes to encourage respondents to keep characters clearly delineated. Automated data capture was completed using 65% fewer employees than manual data capture in the same amount of time as the previous census. Overall the quality of the data capture was comparable to error rates found in traditional "heads-down" data capture, although the type of error was quite different.

For 2006, the automated data capture operation will be conducted with the Census of Population. With the benefit of economies of scale, the data capture is expected to be completed in half the time of the preceding census. Adjusting the edit strategy to detect automated data capture errors is one of the new initiatives in the 2006 edit system.

Finally, the data validation process has used efficient computerized tools (reports, graphs, maps) to make analysis quick, efficient and thorough. Some specific examples include:

- A standard query library using an inexpensive "off-the-shelf" software package
- A custom-built online report system that shows top contributors within specific geographic regions, comparisons to previous census data at several levels of geography and impact-of-processing tables that show the impact of imputation and validation. (These reports are generated in an Excel format so that validators can easily add value to them with new formulas or by sorting.)
- A custom-built form for updating the database that automatically checks updates for accuracy and consistency
- A custom-built graphical mapping utility that can produce charts and graphs to illustrate patterns and changes in variables.

5. Conclusion

Many of the new questions raised by changes in collection and edit procedures will be answered by the outcome to the 2006 Census of Agriculture. Specifically, we are looking for answers to these questions:

- How effective is the two-part screening question on the population questionnaire at identifying new or missed farm operations?
- Can a centralized CATI failed-edit follow-up that targets more severe edits and larger operations reduce respondent burden and provide data quality similar to the former field edit?
- Can the fragmented data collection overcome additional coverage challenges?

- Will improvements to the Farm Register make it a more effective tool for identifying missing farms and aiding in processing operations?

The answers to these and other questions will direct decisions for the 2011 Census of Agriculture. In the 2011 Census we expect the Census of Population to extend the mail-out/mail-back collection methodology to most rural areas in Canada, putting increased pressure on an accurate and complete Farm Register. We expect to find fewer but larger farms in terms of numbers, land area, and sales. We also expect farms to be less household-based and have more complex operating arrangements. All of these expectations will further complicate the task of completing an accurate and comprehensive census. These changes will increase the impersonal nature of the data collection (mail-out/mail-back), adding even more distance from our respondents. The balancing act will continue as we look to strengthen strategies to maintain coverage and data quality and release the data results faster than before.

References

Duggan, J., Bradshaw, C., Julien, C., Villani, R. (2001) Data Collection Methods for the 2001 Census of Agriculture, *Symposium 2001 Achieving Data Quality in a Statistical Agency: A Methodological Perspective*, Statistics Canada.

Julien, C. (2001) Using Administrative Data for Census Coverage, *Conference on Agricultural and Environmental Statistical Applications in Rome*, Volume II, 303-312.