Using Paradata to Actively Manage Data Collection Survey Process

François Laflamme[1], Mike Maydan[2], Andrew Miller[3],
[1] Statistics Canada, Business Survey Methods Division, francois.laflamme@statcan.gc.ca
[2] Statistics Canada, Collection Planning and Management Division, mike.maydan @statcan.gc.ca
[3] Statistics Canada, Collection Planning and Management Division, andrew.miller@statcan.gc.ca

Abstract
With the increasing complexity of survey design, survey managers need better tools to assess the data collection survey process. The development of data collection computer-assisted methods has opened the door to a wide scope of process data ("paradata") that can be used to evaluate the data collection survey process status in a timely manner. Active Management at Statistics Canada is a set of plans and tools developed to manage survey data collection while it is in progress. This includes monitoring progress, timely analysis of these indicators, identifying problems, implementing and communicating corrective actions, and evaluating success. The paper begins with an introduction to the key components of Active Management and discusses a series of practical examples currently used to monitor issues and data quality indicators at Statistics Canada.

Key Words: Paradata, call history, active management, management information system

1. Introduction

With the increasing complexity of survey data collection procedures, survey managers need better tools to assess the survey data collection process. The development of computer-assisted data collection methods has opened the door to a wide range of process data ("paradata") (Scheuren, 2005) that can be used to evaluate the data collection status in a timely manner.

Statistics Canada's Active Management program is a set of processes and tools used to manage survey data collection while it is in progress. Active Management has two complementary objectives. The first objective is to identify problems as early as possible and correct them before collection has finished, while the second is to make effective use of collection resources to find the most appropriate trade-offs between data quality (including timeliness) and survey costs. Active Management activities include monitoring progress, performing timely analysis, identifying problems, implementing and communicating corrective actions, and evaluating success.

This paper begins with a description of the reasons behind the development of the Active Management program for social and agricultural surveys at Statistics Canada, including an overview of the data sources used by this new initiative. The next sections describe the data collection management framework of the Agency and present the Active Management program and its key components. A subsequent section examines the series of standard and customized reports and data quality indicators used to actively monitor data collection. The final sections discuss the current status of the Active Management program with regards to the Data Collection Methodology Framework and present a summary of Active Management initiatives.

2. Reasons behind the Active Management initiative

Requirements for timely data collection progress information, increasing complexity of survey data collection procedures and the emergence of the idea of responsive design are among the main reasons that have led to the development and implementation of the Active Management program at Statistics Canada.

2.1 Timely Data Collection Progress Information
While our current monitoring reports meet the needs at aggregated levels, survey data collection managers typically lack sufficient detailed "real-time" information about the in-progress cases as well as about some key features of the data collection process. The current data collection reports are no longer able to meet the increasing demands for timely data from users with varying needs and requirements.

2.2 Complexity of Surveys
Other pressures come from the increasing complexity of the survey data collection process and its procedures. The introduction of new data collection features such as call scheduler, time slice and cap on calls, as well as the increasing complexity of data collection procedures, has made survey monitoring more and more difficult over time since these factors often have an impact on each other. In that regard, it is important to constantly monitor survey progress with the most appropriate tools available.

2.3 Responsive Design

Highlights of previous research (Laflamme and Mohl 2007, Laflamme 2008) identify the necessity to develop an adaptive and responsive data collection strategy (Groves and Heeringa, 2006), to make more efficient use of the attempts allowed under the cap on calls policy that limits the maximum number of attempts that can be made on a given case. In particular, the collection strategy needs to respond and evolve all the way through the collection cycle in a timely and consistent manner for the units that remain in-progress. Active Management is a key element in the decision-making process of any responsive design since timely changes need to be implemented during the course of the data collection period based on current observations (Hunter and Carbonneau, 2005).

The next section provides an overview of the data sources used by the Active Management program, in particular, Data Integration and Production Planning (DIPP) warehouse which consolidates various sources of paradata information into a central and standard repository.

3. Data Integration and Production Planning

This section provides an overview of Data Integration and Production Planning (DIPP) warehouse and describes its data sources as well as the content of this data warehouse for the Active Management perspective.

3.1 Overview

Data Integration and Production Planning (DIPP) was created as a Management Information System (MIS) to help collection managers with both daily operations and long-term planning as well as to assist senior management in their decision making. In particular, DIPP aims to:

- Improve accessibility to administrative data and paradata;
- Standardize and consolidate data and reports across multiple surveys and survey types;
- Provide relevant data to monitor survey progress and interviewer performance; and
- Provide standardized direct access to raw data for further detailed analysis including Active Management.

3.2 Data Source and Contents

DIPP was developed in Survey Operations Division (SOD) to integrate and standardize operational data from three different sources: 1) Blaise Transaction Files (BTH) for CATI surveys, 2) Case Management (Caseman) for CAPI surveys, and 3) the Survey Operations Pay System (SOPS) for all surveys including non-standard surveys (e.g., the paper and pencil surveys).

DIPP has maintained a database of most of Statistics Canada's regionally collected surveys since 2003, which is updated daily with the most recent information from active surveys. In addition to post-collection and historical studies, this standardized database is also extensively used to perform timely, coherent and comparative analyses of various types of surveys conducted in regional offices across the country. Essentially, most of the CATI survey monitoring performed for Active Management purposes uses the Blaise transaction file which contains a data record for each time that a case is opened, either for data collection or other purposes. More precisely, this record contains relevant detailed information about each call or attempt made to contact each selected unit during the data collection process. It includes information on the amount of time that the case was open, the person who worked on it, the result of the contact attempt and other data.

As part of its key features, DIPP also provides an exporting tool that gives direct daily access to the cumulative call transactions made since the beginning of the data collection period for any survey. In practice, the most recent call transaction information becomes available the day after a given transaction took place when the DIPP database is updated with the most recent available information of the following day.

4. Data Collection Management

Data collection is a key element of the survey process because it often has a direct impact on the quality of the Agency's information and is therefore an important feature of any statistical program.

The management of data collection at Statistics Canada occurs typically within a project team framework that includes subject matter, head office and regional survey operations. Even though subject matter areas are involved in the collection decision-making process at the higher level, their main responsibility remains data production and analysis. Collection Planning and Management Division in head office and the regional offices share the responsibility for social survey data collection. In general, head office is responsible for the front-end work which comprises data collection planning, application and training development as well as the liaison with subject matter divisions. On the other hand, regional offices are mainly responsible for building and maintaining the data collection organization, managing interviewers including interviewer training, maintaining administrative operations and

managing field data collection. Regional offices work with head office to maintain the data collection and administrative capacity and to report on survey progress and costs.

A number of aspects are associated with the management of data collection operations. At its most basic level collection has three principle elements: cost, quality and timeliness. A management information framework must take these elements into account to provide quality (Statistics Canada, 2002) information to users. This management framework is illustrated in Figure 1 using a triangle with these elements as its points. Too much emphasis toward any one element creates a state if imbalance.

In reality, there are additional dimensions that impact on management information needs and outputs. The organizational structure has various levels of management, supervision and oversight and they have different requirements in the depth and breadth of their information needs. Figure 1 illustrates the regional office levels as points along an arrow. At the higher levels there is a need for a wider view that reflects the various surveys active within the manager's span of control along with comparison points such as established targets and national measures for the same surveys. In many ways this view is in line with the needs of head office survey operations managers, while client focus is narrowed to a specific survey or set of surveys. Moving further down the arrow, closer to the collection activities themselves the focus narrows again but increases in depth, with details on individuals or groups of interviewers measuring progress and other indicators against more specific targets on a daily or even hourly frequency. There is an increased need for "real-time" low level data. The management information system needs to be able to build up from this lowest level of aggregation.

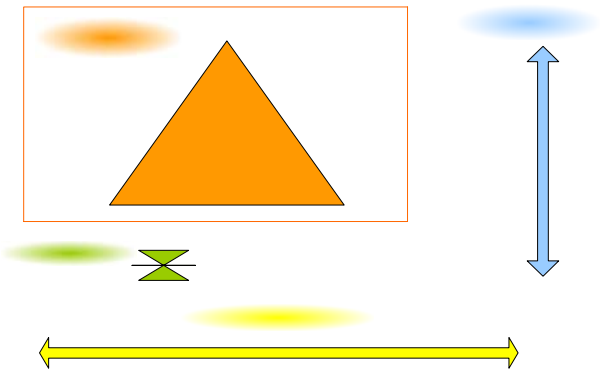

Figure 1

Another dimension is temporal, reflecting the point of time along the data collection continuum. Data needs will vary depending on the point in time. This refers not only to the frequency of refreshing data and providing outputs, but also to the nature of the data elements of interest themselves. For instance if there is a cap on the number of calls per unit on a CATI survey, this is not of primary importance in the early phase of collection with a plentiful sample and a scheduler delivering cases. It does, however, become increasingly important in the middle and late stages of the data collection period. In the early stage, measurement of progress on untouched cases and spreading call attempts across time slices is critical. Early detection of potential cases requiring attention or special follow-up is important. Close cost measurement is essential throughout the survey, but it becomes critical in the last days when scarce resources need to be applied as efficiently as possible to maximize survey response.

Survey managers tend to focus on either interviewers or production units depending on the circumstances, and the management information system needs to encompass the relevant details of both. For instance, staff scheduling and allocation centers on the interviewer, and a manager requires indicators about the individual to make the best choices to optimize cost, quality and timeliness (Laflamme, 2008). These measures also help in detecting problems and launching remedial actions. The unit focus requires data on the cost and quality of completed units as well as narrower measures of completion or response at different levels of aggregation.

5. Active Management

The purpose of the Active Management program is to provide timely, topical and relevant data on survey performance, so that problems with collection are identified early and decisions about how to correct problems are based on cost, effort, and data quality in addition to response rates. Active Management includes:

- Planning
- Monitoring
- Identifying problems and finding solutions
- Taking appropriate corrective actions

- Communicating
- Evaluating and documenting

5.1 Planning

First, Active Management requires planning at the time of application design and development. The plan should identify the most appropriate progress indicators for that survey, the type of analysis to be performed, the types of actions that could be taken, and the strategy to deal with hard-to-reach cases. Planning could also include strategies to follow-up non-response cases as well as to move cases from one mode to another. In summary, planning involves all aspects of data collection at all stages that are required to meet the survey objectives.

5.2 Monitoring

Monitoring survey progress is typically done through reports. The reporting plan for a survey should identify the type of information needed as well as any variables that are required for reporting at various levels of aggregation and ensure that this information is available on a timely basis during the data collection period. The plan should also identify which indicators are most important for monitoring progress and consequently which special reports might be required for a given survey. The next section, which represents the core of this paper, describes the survey monitoring framework used to timely monitor the survey progress against the expectations during the data collection period.

5.3 Identifying Collection Problems

The idea is to find the problem when it occurs, or even before it happens, not when collection has ended. This requires a timely analysis plan and well-trained staff to quickly identify the nature of the problem. The monitoring reports do not presume to give all the answers to issues but are designed to identify possible problems and the reasons that could explain the source of this problem. From that perspective, Active Management also provides mechanisms for handling unanticipated or emerging problems. This includes the ability to generate ad hoc reports to investigate situations as they develop.

5.4 Taking Appropriate Corrective Actions

The objective at this stage is to determine if action is required to correct the problem and, if so, to decide what steps to take. The decision-making strategy and its implementation should have been planned at the initial planning stage. However, since the ability to take corrective action may depend on the design of the application and on operational constraints that might reduce the possible range of actions, it is important to include some contingency planning in the overall collection management plan.

5.5 Communicating

The communication strategy should clearly identify who needs to be contacted in the event of a problem, who has the authority to make decisions, and what the process should be for taking corrective action. The strategy should also identify who is responsible for monitoring different aspects of the collection process. A communication strategy that covers the collection phase of a survey can ensure that survey clients are informed about any decisions or discussions impacting their work.

5.6 Evaluating and Documenting

In order to establish best practices and share experiences so other surveys can avoid repeating a problem, an evaluation of the collection process should be done. This evaluation should document the types of problems that were identified, the decisions that were taken to correct the problems, and the impact of these actions on the data collection indicators.

Active Management versus Quality Assurance

Active Management is very similar in spirit to Quality Assurance since both approaches essentially pursue the same objectives. As with Quality Assurance, Active Management attempts to move "quality" upstream by anticipating problems before they occur and aims to ensure quality via the use of prevention and monitoring strategies. Active Management, like Quality Assurance, uses all available information for generating data collection improvements and is mostly applicable at the planning stage (Statistics Canada, 2002).

6. Survey Cost and Progress Monitoring

This section focuses on the survey monitoring process that is currently used as part of the Active Management program in Collection Planning and Management Division (CPMD) at Statistics Canada. In order to work efficiently and not over-burden head office and regional collection managers with too much information, it has been suggested that three levels of survey monitoring reports be considered in CPMD: standard reports, customized and ad hoc reports.

Standard reports are common to all surveys. These reports are available directly through DIPP in a timely manner to monitor an agreed-upon set of indicators. Customized reports provide information to monitor specific issues and a predefined set of indicators at a level of aggregation not available in standard reports. Standard and customized reports are complementary in providing relevant and timely information at different levels of detail. Finally, ad hoc reports are created when situations arise that are not covered by standard and customized reports. For example, once a problem has been identified, more details on the cause of the problem can be found through the use of ad hoc reports.

It should be noted that this paper only discusses the reports produced by CPMD as part of the new Active Management initiative. It does not include, for example, reports produced by the regional offices or any other Statistics Canada divisions.

6.1 Standard Reports

These reports are standardized across all types of surveys (i.e. CATI/CAPI social, agricultural and business surveys). They are directly available through DIPP and provide consolidated information on the data collection process and cost. The MIS reports or features currently available are described in the following.

6.1.1 Survey Performance Report – Brings together production and financial data to monitor collection indicators such as time per unit, cost per unit, completion rate and response rate at various levels of aggregation.

6.1.2 Survey Progress Report – Evaluates collection progress in terms of cases resolved against the prescribed completion targets or response rates. It also provides information in terms of cases finalized versus cases still in progress.

6.1.3 Interviewer Profile Report – Documents the performance, associated costs and employment information of each interviewer. In particular, the performance indicators include the following information for each interviewer and survey: a) the sample size broken down by out-of-scope, b) response, c) non-response at detailed levels, and d) system time. It also includes financial information recorded through Survey Operations Pay System (SOPS) for each interviewer.

6.1.4 Survey Summary Report – Used to evaluate the overall survey results and costs. In particular, the report compares the recorded interviewing hours on BTH/CASE and the reported (claimed) interviewing hours on SOPS.

6.1.5 Calls versus Claims Report – Allows managers to visually compare an interviewer's current production time and all call outcomes from the BTH and case event files against the interviewer's SOPS pay claims at various levels of aggregation.

6.1.6 Exporting Interviewer Calls and Claims – Gives users access to the collection activity records contained in the Blaise transaction history (BTH) and Case Management case event files or in SOPS. This is the most important source of information used by the customized reports.

6.2 Customized Reports

Like the standard reports, customized reports rely upon DIPP information for consistency purposes. Customized reports are usually tailored to specific needs of survey. In particular, these reports use the Export Interviewer Calls feature to extract records of call information for each monitored survey. In addition, the customized reports often use sample design information as stratification variables to enhance the analytical relevance of the reports by producing reports at a more detailed level of aggregation than available via the standard reports. These reports allow operations staff in the regions and headquarters to jointly monitor collection and make informed decisions. Beyond direct survey monitoring, they also allow operations to advise survey client managers of impending issues and to quickly respond to their concerns during the course of data collection. For example, reports could easily be used to show survey clients the amount of time spent on their survey each day as well as the amount of effort devoted to specific types of sample units (such as refusal and tracing cases). In the following, the series of customized reports produced as part of the Active Management is described in more details.

Description of the Customized Reports

This section begins with the presentation of customized reports currently produced for many CATI social and agricultural surveys and then gives an overview of the reports in-progress used to monitor issues not covered by either standard or customized reports.

6.2.1 System time across surveys

On any given day, regional offices conduct many CATI surveys which compete in some way for the available data collection resources. For example, 9 distinct surveys were conducted in Sturgeon Falls in November 2006. Figure 2 shows the distribution of system time by survey and day during this month. The system time corresponds to the total duration of all calls during the reference period. This information is very useful from both the head office and regional office perspectives to compare and understand the relative contribution of each survey on a daily basis. Another national survey report provides a high-level overview of progress for

all active surveys during a specific reference period. This report focuses on resolution status versus resolution targets and previous reference points, providing a break down of units by not-started, in-progress and finalized across all surveys.

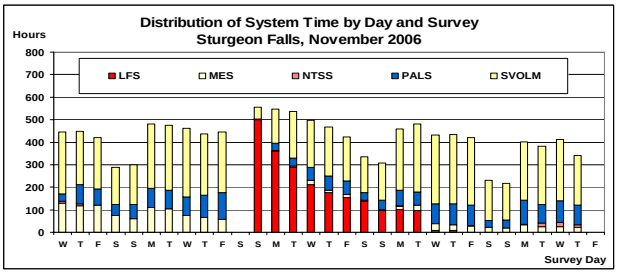

Figure 2

6.2.2 System time by survey

The next customized report provides, for a given survey, the distribution of system time over different periods throughout the course of the data collection period. This report provides managers with a tool to monitor the number of hours worked in the peak period (e.g., evening periods for social surveys) against the achieved number of hours. Generally, over the data collection period for social surveys, data collection managers expect a gradual shift of interviewing effort from the morning and day time through to the evening. This trend could be monitored through this report. In addition, the report also provides the cumulative distribution of system time hours as well as the number of interviewers assigned in each time slice.

6.2.3 Progress Reports

The next report monitors the progress of a survey by regional office, date and any type of variable available on the sample file. For practical reasons, the report is divided into three parts although the parts are linked to each other.

The first part of the report compares the number of cases accessed through the browser1 with the total number of attempts made that day. The second part of the report provides information about the in-progress cases available during each data collection day. In particular, the report gives the distribution of total number of attempts for each in-progress unit at the end of the data collection day. The last part of the report provides the number of completed questionnaires by day since the beginning of the data collection period, as well as the cumulative number of resolved cases at the end of each data collection day by final status of the case i.e. respondents, non-respondents, out-of-scope and capped records. The entire in-progress report not only gives a snapshot of current survey progress status, but also provides the daily progress situation at the end of each data collection day. In combining the progress, the system time report and the upcoming performance reports, the data collection managers have a very good overview of the amount of effort versus success put into the survey on a daily basis.

6.2.4 Progress at sub-population level

The next report provides detailed survey information progress by domain of interest, through taking advantage of sample design information available at the beginning of the data collection period. In particular, it compares the expected responses (or response rates) with those actually achieved at a level of aggregation not available in the standard reports. It thus provides a tool that could be used to prioritize some domains of interest for which progress is below expectations compared to others. For advanced users, it is also possible to compute timely approximate coefficient of variation to respond to the needs of some specific survey managers and users.

6.2.5 Refusal and tracing indicators

The following series of tables aims to provide a dynamic tool to separately monitor cases that pass through at least one refusal or tracing occurrence. In other words, the issue is to monitor the proportion of these cases still in progress, the average number of attempts made on both finalized and in-progress cases, as well as the number of cases converted as complete response. In addition, the reports show the distribution of total number of attempts made after the first refusal or tracing occurrence. This could also be seen as a measure of effort made by the regional office on the hard-to-reach (tracing) and hard-to interview (refusals) cases. Refusal cases are of particular concern since Statistics Canada limits the number of follow-up contacts with a respondent once he/she states their unwillingness to cooperate.

6.2.6 Cut-off

At any point in time, it is possible to monitor the distribution of cases among the different groups available throughout the call scheduler to assess the workload available to various interviewer groups depending on skills and experience. For example, cases in the interviewers' refusal and tracing groups, as well as those in the senior interviewers group, are only available to experienced

[1] The browser facility allows the interviewer to manually scroll and select a case, thus "skipping" the call scheduler's automatic delivery.

interviewers. This report could also be used to quickly identify cases with a low number of attempts particularly when nearing the end of the data collection period which could require specific attention.

6.2.7 Staffing levels and workload
Finally, some research was undertaken to better understand the relationship between expended interviewing effort and expected in-progress workload throughout the data collection cycle, especially in the second half of the data collection period.

Figure 4 shows that cases are likely to be called more often during a single day in the second half of the data collection period. During the first five days of collection only about 4% of all cases were called three or more times on any one day. However, by the ninth five-day period this number had risen to 25%. This suggests that at this point interviewer staffing was larger than sample workload. In other words, more interviewing hours were scheduled for this period of time than reasonably required. Making multiple futile attempts on a single day can be damaging to surveys with a cap on the number of calls, especially for longitudinal surveys or others that may require tracing. Better tools that would allow the collection centres to more effectively schedule the interviewers' time would reduce this risk.

6.2.8 Special reports
Additional customized reports are currently in development, in particular, time zone and productivity and performance reports.

A time zone report would likely monitor attempts by comparing regional office time to respondent time (i.e., time at the origin and destination of the call). This is of particular interest for agricultural surveys where the whole sample is allocated to a few regional offices. This means that regional offices have to manage their sample according to many time zones across the country.

As part of the Active Management, the development and implementation of productivity indicators was identified as one of the important areas of research to monitor survey performance. The main objective of such a report would be to provide a general framework for monitoring and assessing the data collection performance of individual surveys throughout the data collection cycle. This report would contain the following data collection information by hour slice: total number of attempts, total number of contacts, total number of completed interviews as well as total system time and system time spent on completed interviews. This could be done by day or data collection intervals depending on the duration of the data collection period.

6.3 Ad Hoc Reports

The survey monitoring reports do not aim to provide answers to all possible issues. However they are built in such a way as to highlight key elements that could either identify the source of the problem, or point to the need for more in-depth investigation through the use of ad hoc reports. These reports could be produced either directly by the user through EXCEL pivot table utilities, or by special request to the Active Management group when more sophisticated processing tools are required.

7. Data Collection Methodology Framework

The common data collection methodology framework could take the form of responsive design (Groves 2006) or adaptive collection design (Laflamme and Mohl 2007). The main idea is to constantly assess the progression of data collection (Active Management) using the most recent information, available, and adjust data collection strategies in order to make the most efficient use of available resources including cost (Adaptive Collection).

The responsive or adaptive collection design could be seen as a multi-phase or phase-in strategy (i.e., collection phases that evolve as collection progresses). Each phase represents a period of time during the collection period in which the same data collection strategy is used. For example, the first phase would likely make better use of the information available prior to the beginning of the data collection. The second phase would benefit from using the sequence of calls information built for each specific case, in order to increase the likelihood of contacting and interviewing a given case based on its own call history. A third phase could give priority to certain cases during data collection as the score function does for a business survey. Finally, the last phase might consider the possibility of sub-sampling non-respondents to reduce the amount of time and resources dedicated to cases which have a low probability of responding.

This common data collection methodology framework should provide the data collection environment architecture that fully supports multi-mode (CATI\CAPI) and multi-site surveys. This would also give the flexibility to better manage workloads, and would be tied into the Data Collection Modernisation by allowing managers at all levels to access common, central reporting tools so that they can better monitor collection progress.

Future Work

Initially, the focus of the work on Active Management activities was essentially put on the CATI surveys. However, Statistics Canada is currently developing Active Management tools for CAPI surveys that are very similar to those used for CATI surveys. The main focus of CAPI reports is to recreate the hierarchical dynamic of field work between the various levels of management. In other words, it would be possible to link interviewers, senior interviewers, regional office managers and field workloads to have a better and timelier picture of survey progress at any point during data collection progress.

In the future, it would be interesting to consolidate transaction and audit trail information in a unique system such as the US Census Bureau's Performance and Data Analysis tool (PANDA). This could improve the data collection monitoring framework by providing for example a more efficient decision-making tool for the timely identification of cases requiring a transfer from mode to mode.

Conclusion

Data collection is definitively a key element of the survey process of any statistical program. The manager's challenge with regards to data collection management is to make the appropriate trade-offs between quality, timeliness and cost in order to meet the clients' needs. With the increasing complexity of data collection procedures, managers need better tools to assess the survey data collection process. The main objective of the Active Management program is then to provide timely information in order to closely monitor survey data collection progress. Currently, Active Management focuses on timely survey monitoring of social surveys through the use of standard and customized reports.

The customized reports produced from the DIPP environment provide common and consistent tools to head office and regional office data collection managers to closely monitor survey progress in a different way than the standard reports do. Generally, each customized report focuses on one or two specific issues in a more detailed manner and are often linked to each other for analytical purposes. The customized reports constitute a flexible monitoring framework that could easily be adapted to specific survey needs. However, they should be seen as prototypes requiring some fine-tuning before being considered for integration into the standard reports series. The ad hoc reports, by their nature, require particular attention.

Active Management is a key component of the data collection methodology framework integrating a wide variety of components and processes that have different levels of importance to different management levels at different points in time in the collection cycle. The effectiveness of this framework depends not on any one process but on the collective effect of many interdependent and often complex measures. As other parts of the survey progress, data collection efficiency depends on the ability of the Agency to adapt and evolve.

To do this, Statistics Canada needs to continue its support of analysis, innovation and experimentation in order to maintain a professional operational infrastructure. Staff must not only focus on technical knowledge but also be aware of data collection issues and be able to monitor and analyse the data collection process in order to be recognized as data collection experts.

Acknowledgements

The authors wish to thank the staff in the Collection Planning and Management Division and the Regional Office Branch of Statistics Canada, especially Duncan Wrighte, Yves Béland, Lecily Hunter, Chad Mackinnon, Teresa Bax, Lise Rivais and Tom Thibault for their advice, contribution and comments to this research.

References

Groves, R.M. and Heeringa, S.G. (2006), "Responsive design for household surveys: Tools for actively controlling survey errors and costs". Journal of the Royal Statistical Society Series A. Volume 169, Part 3.

Hunter, L. and Carbonneau, J.-F. (2005), "An Active Management Approach to Survey Collection". Proceedings from the 2005 Statistics Canada International Symposium on Methodological Issues.

Laflamme, F. and Mohl , C. (2007), "Research and Responsive Design Options for Survey Data Collection at Statistics Canada". 2007 American Statistical Association, Proceedings of the Section on Survey Research Methods.

Laflamme, F., (2008), "Understanding Survey Data Collection through the Analysis of Paradata at Statistics Canada". American Association for Public Opinion Research 63rd Annual Conference., 2008 American Statistical Association, Proceedings of the Section on Survey Research Methods.

Scheuren, F. (2005). "Paradata from Concept to Completion". Proceedings from the 2005 Statistics Canada International Symposium on Methodological Issues.

Statistics Canada (2002), Statistic Canada's Quality Assurance Framework, No. 12-586-XIE.