

Survey Management for the United Enterprise Statistics Programme at Statistics Canada

**Armin Braslins, Operations and Research Development Division, Statistics Canada
March 2000**

Background

In October 1996, Statistics Canada was presented with one of its largest and most complex initiatives of the decade. It was a direct result of the decision taken by the Federal government and the Provincial governments of Nova Scotia, New Brunswick and Newfoundland and Labrador. These four governments decided to consolidate the Federal Sales Tax and the Provincial Sales Tax into a new combined Harmonized Sales Tax. The main objective of this initiative was to reduce the administrative burden (and costs) of administering four separate tax collection systems. They agreed on a complex formula to distribute the tax revenues among the four governments. Statistics Canada was asked, indeed challenged, to provide the detailed economic data needed for the formula. The result of this challenge was the creation of a new programme to improve the statistics gathered for the business sector. This new programme became known as the "Project to Improve Provincial Economic Statistics" (PIPES).

The PIPES programme has many goals and objectives affecting all aspects of the survey collection process. In general terms, the main objective of PIPES is to supply the data needed to support the agreed revenue distribution formula. Gathering the necessary raw data meant a very substantial increase in the amount of available provincial economic statistics. The data had to measure the final sales of goods and services, on an annual, calendar year basis, by province and industry, by commodity and class of respondent. The end result being a transformation of these data into a set of Provincial Economic Accounts.

The Provincial Economic Accounts would collect additional data on both the expenditures and the incomes of households and businesses. To increase the data for the expenditure side, improvements would include larger and more frequent surveys to a number of personal surveys such as the Household Spending Survey and the Repairs and Renovations Survey.

But, the large majority of the changes were needed on the income side. The major changes to these surveys would occur with the development of a new integrated programme that would eventually replace all the existing programme of annual business surveys. This new programme within the PIPES initiative is known as the United Enterprise Statistics Programme (UESP).

United Enterprise Statistics Programme

The UESP would reform all the processes associated with conducting any business survey. This includes such basic components as sample selection, questionnaire development, data collection, data editing, imputation, analysis and dissemination. The UESP as a departmental programme has many goals and some of the basic principles that are guiding its evolution include:

- business surveys would be enterprise based;
- all data for an enterprise and its establishments will be collected and analysed together;
- special attention would be placed on the complex enterprises, which, although small in number, have a large impact on all economic production;
- reduce the response burden, particularly for the small business, by making use of administrative (Tax) data wherever possible; and,
- harmonize the concepts, definitions and questionnaires used in business surveys.

The UESP consists of four parts (figure 1) depending on the type of enterprise (simple or complex) and the type of data required (enterprise-level or establishment-level). Part 1 would be a census of the complex enterprises collecting primarily financial data. Part 2 would be a census of the establishments belonging to the complex enterprise collecting financial and non-financial data. Part 3 will be a sample survey collecting primarily non-financial data at the establishment level for simple enterprises. Part 4 would be the use of Tax data for simple enterprises.

Figure 1. Components of the UESP

	Enterprise Data	Establishment Data
Complex Enterprises	Part 1	Part 2
Simple Enterprises	Tax	Part 3

The development of the UESP would mean that the methodologies related to business surveys would change in many aspects. One of the biggest and one of the first to be encountered involved the survey questionnaire. A major objective is to harmonize and simplify the number of questionnaires sent to businesses. Statistics Canada currently has about 100 annual business surveys involving more than 700 different questionnaires. This is a result of the long established traditional vertically integrated method of survey collection.

The UESP approach is to have one consolidated questionnaire at the enterprise level, supplemented with industry specific "Schedules" at the establishment level. Standard concepts and question wording would be used across industries wherever possible. The questionnaires would also be simplified by using concepts and terms familiar to the respondent. And eventually, personalized questionnaires would be created showing only their previous responses.

The emphasis on the enterprise or enterprise-centric approach to the UESP will also affect sampling and data collection. The foremost of the concerns being the response burden, especially with small businesses, some, of which may not even be able to provide the requested data. In the past, data at the enterprise and establishment level have been collected by separate surveys with different sample designs. The UESP approach, which combines them into one vehicle, requires a new approach. Collection of data for both the enterprise and establishments in a coherent manner would be a challenge with many questions.

UESP Data Collection

Now in its third year, the UESP is a work in progress and it continues to evolve in all aspects. This paper will concentrate on the data collection process.

The PIPES programme developed a plan for the integration of surveys into the UESP process. For the first year, 14 surveys were identified for inclusion. This meant that these surveys had to be developed and the data collected in a very short time period. Most of these surveys had to be developed from the beginning because the surveys collected data for new industries not previously sampled. These surveys varied in size from 9 pages to 30 pages (including annexes). Sample sizes also varied from 12,000 for the Wholesale Survey to only a few for the Banking Survey (90).

The primary collection tool was the CASES software. The development methods and survey management processes of the existing annual and monthly collection cycles were used. This software and management system was not built with the UESP model in mind. While the first year proved workable, it would not serve as the model for the future.

The search for a replacement collection tool began in earnest. New data collection software had been the subject of much discussion within our collection groups for some time. There were several factors influencing which software would meet Statistics Canada's requirements. We did not want to build it; the implementation time frame of PIPES would not permit this. We knew that we had to move to a Windows environment in order to make efficient use of our existing base of computers. DOS, as an operating system was nearing the end of its life cycle. It was not flexible enough nor was it stable enough to keep up with the ever increasing demands of the surveys. Survey requirements were becoming more and more complex. At the same time, our capi surveys were experiencing growing pains and limitations as well. They too were interested in finding a Windows based collection tool for use on the field interviewers laptop computers. Unix was not seen as a viable tool for use on laptops and seen only as a tool available to a central collection facility. A Windows based collection software was the only solution.

Blaise happened to arrive on the scene with its new release of Blaise for Windows. At first appearances, it looked very promising. The timing could not have been more fortunate from the perspective of Westat and Statistics Netherlands. The timely release and initial demonstration of the Blaise for Windows capability clearly started the department thinking of its possibilities.

From our perspective, some of the main features of Blaise that worked in its favour were:

- a production version of the software was available in Windows including documentation;
- the data model could present the questions in more than one language;
- program code was easily reusable in different data models;
- it was easy to switch between multiple data collection modes (data entry, cati, capi) using the same program;
- ease of programming (modern programming concepts);
- speed of the collection application;
- built in cati scheduler;
- the appearance of being able to quickly develop a collection application; and,
- access to training and support from a reliable organization.

While Blaise has many of the essential elements of an ideal data collection tool, it did and still does have its limitations. At that time, some of the more important missing features, from Statistics Canada's perspective, included:

- no data capture keying verification for a high speed data capture operations;
- no capability to directly read and write to an external database, a beta version of read only is now available;
- data editing, since upgraded in a recent release with the addition of user defined edit types and edit masks;
- capability to easily change the default screen layout, since upgraded with a release of Emily;
- edit tracking (keeping a record of which edit failed); and ,
- a facility to produce management information reports.

But, it was the opportune emergence of Blaise for Windows and its many other desirable features that resulted in it being selected as the standard data collection software for Statistics Canada.

Year two was a new beginning for the UESP. It was a year of transition from the old to the new. Conversion from one software collection tool to another is never an easy task. But, the UESP had a distinct advantage; their requirements were new requirements. New software, new concepts, new approaches, new development. A huge advantage over trying to fit a new data collection system into existing methods and operations.

During the second year, the original 14 surveys were rewritten and 18 new surveys were added. With new surveys being introduced each year into the UESP model, being able to quickly add a new survey is very important. The PIPES strategy has all of Statistics Canada's business surveys being folded into this same model.

The process model used for the first year of UESP identified areas where improvements would have to be made. A new process model was developed. Briefly stated, it consists of:

- a mail out of a paper questionnaire;
- data capture of the completed paper questionnaire;
- follow-up of late returns;
- follow-up of edit failures from the data capture operation;
- on-line capture of identified cati respondents; and,
- feedback to the business register.

This process model may not appear to be any different than the process model used by many other statistical organizations but what is different is the data, the collection entities and the inter-relationships demanded by the UESP. Another major difference is the annual collection cycle for UESP. PIPES requires that the data be collected for a calendar year and as near the end of the year as possible. This meant that a mail out and collection operation had to take place over a relatively short time period. A staggered mail-out throughout the year is not an option for the large majority of these surveys.

For this cycle, a new survey (case) management system was also developed to better manage the centralized data collection operation. Again, for this process, good survey management at Statistics Canada is no different than good survey management at other statistical agencies. Information is required on completion status, who is outstanding, ensure timely non-response follow-up; in simple terms, what has been completed and what remains to be done.

UESP year three brings the introduction of 20 additional surveys, including the conversion of some of our bigger, more complex surveys. And, further refinements to the survey management system. Components to control what data editors are allowed to work on, information on the status of the survey's environment and who is using which application. Included also, is an option to report data electronically. For surveys that report their data electronically, the data is re-integrated into the Blaise data model for editing and follow-up.

Survey Management System

Survey management starts with the mail out of questionnaires and is followed by their return, in various states of completeness, to the mailroom. Questionnaires are logged as received and sent to data capture and editing. Non-responding units are identified for follow-up and reminders sent. Other units are scheduled for cati collection. Probably very similar and familiar to many organizations. What makes this different, is the tools that are used and the operational constraints of the surveys. Under the constraints imposed by the UESP previously mentioned, you can imagine that the return of many thousands of questionnaires for many surveys could mean total confusion. Knowing the processing status of a particular questionnaire could be very difficult to determine. And, this status must be known as soon as possible particularly if a non-response follow-up is scheduled. One would not want to have a questionnaire being identified for a non-response follow-up when that completed questionnaire is in fact being data captured. One central repository had to exist to manage the many questionnaires in their varying stages of processing. A central database would be a natural choice. But, with the lack of a facility within Blaise to directly access an external database, a Blaise database was used instead. The resulting system actually consists of four Blaise data bases supported by a number of Maniplus and Manipula programs.

This management system is a series of menus and sub-menus presented in the familiar Blaise format that provides information relative to each individual survey and each respondent within the survey. Data is stored that records information such as the status of the survey environment (internal testing, client testing, production), access privileges (interviewer, supervisor, administrator), and management information (complete, partial, non-response, refusals, etc) and permits access to the central document control, survey collection functions and administrative utilities. This whole system of managing the surveys uses Blaise tools only.

In these menus are two very important functions that deserve special mention. The first is the central document control. This menu includes the ability to:

- log-in incoming questionnaires sent through the mail;
- update mailing label information;
- view, add to or create mail groups;
- view, add to or create combined reports;
- produce labels for re-mail; and,
- produce Fax follow -up.

The second important menu is called active surveys. It is through this menu that the actual survey collection takes place; again assisted by a number of sub-menus. Some of these functions include:

- selecting a respondent by specifying its unique identification number;
- creating a day batch;
- starting the data entry program;
- starting the cati call scheduler;
- starting cati management; and,
- producing status reports.

Each time that an interviewer enters and leaves the central document control or the active surveys, status information is collected. This information can be stored on one of four Blaise databases. As you can appreciate, since these four databases are used throughout the system, it is extremely important that the databases are kept synchronized. Before an interviewer enters one of these functions, the databases are synchronized by a Manipula job. Every night, another Manipula job is run to further ensure that the databases are synchronized and readied for the creation of the next day batch.

This process is not without its problems. The problems that arise are not from its functionality but rather in the performance of some of its functions. Most functions execute quickly, but some are slow. Analysis and system monitoring is being done to try and pin point the exact cause, but that has not yet been determined. Some problems may be related to our system design and solving this problem remains a top priority for us. However, for an organization that is relatively new in its experience with Blaise and particularly Blaise for Windows we feel that we have made considerable progress. We certainly have not discovered all of its quirks. Perhaps our expectations for the product are too high and it may become necessary to use tools outside the set of Blaise software to meet some of our requirements. We very much wanted to use Blaise as it comes "right out of the box" without any extra software.

What lies ahead for case management within the UESP is more evolution. The development of a personalized questionnaire for a respondent has completed the analysis phase and is progressing to development. Its design includes the use of an Oracle database and XML. While this does not directly use Blaise, it is a highly desirable feature within our survey collection process. Other enhancements currently under construction include a component to handle the mail out of questionnaires, a component to manage all the various survey outputs and a component to move a survey unit which is "out of scope" for one survey to being "in scope" for a different survey.

Conclusions

As mentioned earlier, Blaise is not without its limitations and the good news is that additional features are being developed. We feel that having a direct read and write capability to an external data base would go a long way to addressing our performance issues. A beta release that permits a read is available and a write capability is planned. We are looking forward to its release.

The ability to keep a record of which edit has failed within a data model is also very desirable. This information would be used to analyse how often and under which conditions edits are being triggered. This would be used to define better edits in the future.

Also, we need the capability to perform data entry verification. We still receive a large portion of our questionnaires by return mail. If a completed questionnaire is returned by mail, it remains a very cost efficient method of data collection. But, in this day of e-commerce and the explosive use of the Internet, traditional methods of data collection may become a thing of the past. Some respondents are currently returning their completed questionnaires electronically and this is expected to grow. How will it be done in the future? A very good question but no final answers. We are actively working on collecting and editing data electronically, particularly using the Internet. A proto-type for use in our 2001 census of population is being developed but a lot of challenges remain. That work, for us, is very much at the research and experimentation stage and clearly the topic for a future conference.

Back to the present, Blaise is proving to live up to its billing even though production problems continue to creep into our operations. These problems are more a result of our application design, programming errors and improper operational instructions than problems with the software itself. We are continuing to build our Blaise expertise and increase the number of developers. Recent changes within Statistics Canada, which re-organized all the computer assisted interviewing applications development into one division, has re-enforced the original decision to select Blaise. A long process of conversion and re-writing all of our existing collection applications now lies ahead.

References:

1. An Overview of the Project to Improve Provincial Economic Statistics, November 1997, by George Beelen, Francine Hardy and Don Royce
2. PIPES Information Package - Outline, Updated October 1998, Statistics Canada
3. Unified Enterprise Survey Information Package, March 1999, Statistics Canada