Integration of CAPI and CATI infrastructures at Statistics Canada

Jacqueline Mayda & Pamela Ford, Statistics Canada

1.Background and motivation for the integration

Statistics Canada's numerous social surveys have been conducted for many years now using Computer Assisted Interviewing applications, some in person (Computer Assisted Personal Interviewing, CAPI) and others by telephone (Computer Assisted Telephone Interviewing, CATI). The telephone component can be conducted by the field staff from their homes using the same application as used for CAPI, or can be conducted in the call center in each Regional Office. The Blaise software is currently being used to author all of these collection applications. Workload planning and allocation is done differently depending on the mode of collection. Field interviewers have specific cases assigned to them through case-management software that was developed at Statistics Canada and interacts with the Blaise survey applications. Telephone interviewing by field staff from their home does not require the same functionality as a full CATI environment in a Regional office where the Blaise call scheduler allocates the cases amongst many CATI interviewers depending on the case criteria, respecting appointment times and the like.

A few years ago the introduction of a new large-scale survey (the Canadian Community Health Survey) dramatically increased the number of personal interviews. In order to meet the capacity requirements, a decision was made to consolidate the telephone interviewing portion of all surveys into the existing Regional Offices throughout the country. Due to the short time frame for implementation of this initiative, a formal redesign of the software was not possible. As an interim measure, survey applications previously used solely by field staff were quickly set up in a pseudo-CATI environment in the workstations in the call centers, with the software functioning exactly as it did on the laptops. Consequently this did not allow for any of the features of a CATI environment, such as the use of the Blaise call scheduler.

Management problems in the Regional Offices (in terms of assigning cases to interviewers and workload planning) arose as a result of not having a call scheduler facility in this environment. Pseudo-CATI and CATI (call scheduler) applications share workstations. When a workstation is being used for CATI surveys, its pseudo-CATI cases are completely unavailable. When a workstation is being used for pseudo-CATI, it cannot be used for CATI surveys. Managers must print lists of cases with appointments and ensure that both workstations and interviewers are available at the appropriate times. In addition, transferring cases from Regional Office servers to laptops for current full-CATI (call scheduler) surveys is not possible due to the difference in software configuration; we can copy files to LAN servers for CATI, but there is a rich mix of software and files involved in CAPI. To change from CAPI to CATI or vice versa would require a significant investment in additional development time.

Given all of these issues it was felt by all stakeholders that the best approach would be to undertake a redesign of the infrastructure for the CAPI and CATI collection applications with the goal of one system that could be used in both environments, and the ability to transfer cases between the two environments. The full integration option in terms of cost and implementation issues was investigated, but the time

and resources needed to develop the components required for full integration, the impact of a "big bang" approach for clients, the need for a lengthy implementation and testing window, as well as a high cost, made this option less desirable.

Instead, a partial integration approach was adopted. This approach promises to render collection operations more efficient by introducing a true CATI environment with the utilization of a call scheduler starting with the Labour Force Survey (LFS) and its supplementary surveys, and progressively converting each of the other CAPI/CATI surveys to the new model. While the partial integration does not provide for one fully integrated CAPI/CATI environment, it does maintain a transfer process between CAPI and CATI, albeit a manual one. An additional benefit of the integration relates to the programming of the applications. Since the original conversion of CAPI surveys to Blaise, there has been a move towards more coding standards, and the development has become more efficient. These standards can now be applied to the Labour Force and supplementary survey applications as part of the redesigned infrastructure.

2. Description of the Existing System

In the current CAPI environment, cases are stored in individual databases in individual directories and one database is created *per case*, *per laptop*. For surveys that share a sample, such as the LFS and its Supplements, each survey has physically separate cases. Case-management software, called CASEMAN and developed in Sybase by Statistics Canada, is used to send and retrieve data to and from the laptop computers and to workstations in the regional offices. The case management (assignment planning, reporting) is done on one database for <u>all</u> surveys. Survey Collection applications consist of a combination of Blaise and Visual Basic components. Most of the questionnaire modules have been developed in Blaise, with a few left in Visual Basic. All components are controlled by a Visual Basic program manager that has the ability to 'communicate' with CASEMAN. The pseudo-CATI environment is the same as the laptop CAPI environment, except that the applications reside on Network Workstations in a Regional Office.

The full CATI (call scheduler) environment has quite a different configuration than that of the CAPI infrastructure. All survey data are stored in a single database, and there is one database *per survey per CATI Server*. In each region, each survey has its own Blaise call scheduler. Case Management (call scheduling, reporting) is done on the survey database. All components of all full CATI surveys have been developed in Blaise.

3. Objectives of the redesign

To attain the overall goal of providing an environment to effectively manage interviewing resources, respondent contacts, and respondent burden, the project team delineated the following requirements as objectives that must be met and that drive the design of the integrated system:

- There is a need for one collection system useable in both the CAPI and CATI environments.
- All CATI surveys should be administered using the Blaise call scheduler.
- There should be a means of transferring cases between platforms.
- Surveys must be able to share samples, but receive data and have their progress tracked individually. In addition, surveys sharing a sample should be permitted to have varying collection periods.

❑ There should be a mechanism to effectively manage interviewing resources (interviewers, workstations, workloads, etc.…) in both CAPI and CATI environments.

4. Conceptual Model

The existing CASEMAN system already provides functionality for CAPI assignment planning and for transmitting cases to CAPI laptops. As any significant modification or redesign would come with great cost, it was determined that using this software (with the least change) to move cases between platforms should be an additional requirement.

Furthermore, it was determined that there would be additional savings in using the call scheduler included in the Blaise software package which is already in use for existing full CATI surveys, rather than adding this functionality to CASEMAN. This last requirement implies that the existing CATI Survey application model, or one similar to it, must be implemented on the laptops. As the existing CATI model is not suitable for surveys sharing a sample, such as the LFS and its supplements, a new model has been proposed.

A model had to be devised for survey collection applications so that they could be implemented in both CAPI and CATI environments. The approach currently in use on the laptops is not appropriate for call scheduling; a call scheduler would have to be implemented into CASEMAN. The approach currently in use on CATI (call scheduler) workstations can be used on laptops, but the interviewing application would be independent from CASEMAN and an interface between the two systems would have to be developed if CASEMAN were to be used for transmitting cases.

Furthermore, surveys that share a sample must be treated as a single entity to enable the use of a call scheduler. For LFS and its supplements, this means that a single entity must be created which encompasses the parent (LFS) and supplement client records for the sample unit. This entity is being referred to as the master sample entity (MSE). The CATI call scheduler will schedule MSE cases rather than individual LFS or supplement cases.

Until such time that an interface between CASEMAN and the new survey application model can be developed and implemented, it will be necessary to maintain the existing survey application model in the CAPI environment. It will nevertheless be possible and advisable to implement the actual Blaise questionnaire components being used in the CATI application into this model. Figure 1 shows the conceptual model of the CAPI/CATI environment:

Figure 1: Conceptual Model CAPI/CATI environment
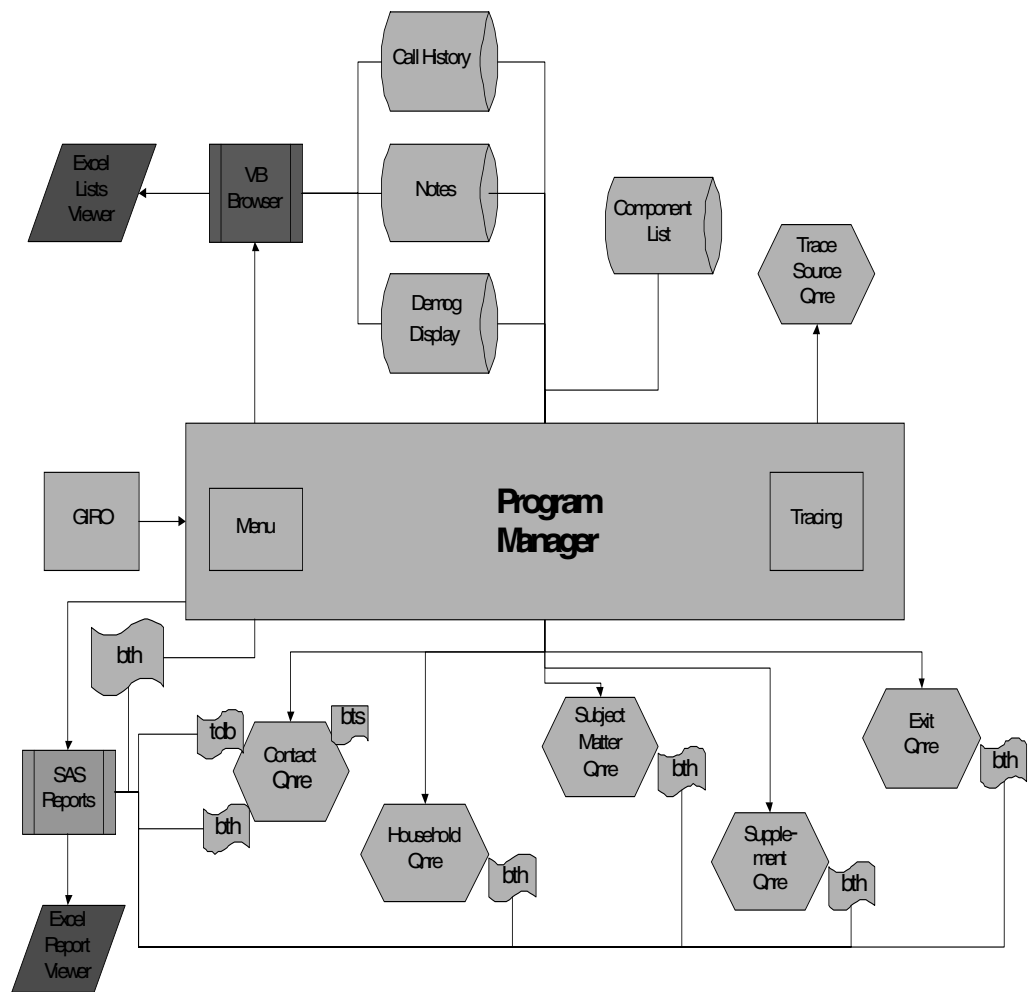
4.1 Program Managers

A program manager is the driver of a collection system. It controls the flow between questionnaire components and launches the questionnaire components when appropriate.

In the CAPI environment, each survey will have its own Visual Basic program manager with supplement program managers being launched at times by the parent program manger. CAPI program managers also consist of the functions required to send and receive information to and from the Sybase CASEMAN database. When collection of the parent survey is complete, interviewers access each of the supplement surveys directly through their individual program managers.

In the CATI environment there will be one Maniplus program manager that controls all parent and supplement questionnaire components.

For both program managers, much of the code surrounding the launching of the questionnaires themselves is generic and does not require modification from survey to survey. The benefits of this design are obvious.

4.2 Questionnaire Components

There are four types of questionnaire components. The same components will be used in each of the CAPI and CATI environments.

The <u>Contact</u> questionnaire has three roles. The first is to record whether contact has been made at the dwelling or telephone number level and to indicate which treatment should be applied in the event of non-contact. The second role is to guide the interviewer in making contact with an appropriate respondent for the survey. Finally the contact questionnaire is also used to provide introductory information for the respondent, to obtain the respondent's language of preference, and to inform them, when applicable, that the interview may be monitored.

The <u>Household</u> questionnaire is used to collect household-level data. Normally the information collected is used to determine which subject matter detail should be collected and who should be interviewed further.

<u>Subject matter</u> questionnaires are used to collect the main content of a survey. The Labour Force Survey has two subject matter components, with various supplements each contributing additional components when required. For multi-datamodel surveys such as LFS, subject matter questionnaires are normally accessed via a component list.

The <u>Exit</u> questionnaire is used to wrap up an interview session. It provides interviewers with a means of assigning an outcome code to the survey and includes text used to thank respondents for their participation and to inform them of any potential future contacts they may have with Statistics Canada as a result of their participation.

4.3 Sample Files

Clients will prepare a sample file for each environment that serves as input to the case management and collection systems. The file for the CAPI environment consists of two parts, the first part containing information used solely by the CASEMAN case management system and the second part used solely by the survey collection application. The structure of the input to the CAPI collection system will be identical to that of the CATI collection system.

4.4 Output Files

The structure of the output files will be the same for both CAPI and CATI environments. Records from the CAPI environment will be returned in individual files per case, while the CATI environment will return multi-case files on a daily basis.

4.5 Reports

Being able to report on the progress of individual surveys is a key requirement. In the CAPI environment, collection reports are associated with the CASEMAN case management system and survey-level reports already exist.

In the CATI environment, standard reports are created from historical case event data that exists at the scheduled case level. It will be necessary, then, to ensure that historical case event information is kept at the individual survey level as well, so that the required reports can be produced. Keeping historical case event information for each questionnaire component will facilitate this.

5. Implementation plans, testing and other considerations

As this project has been foreseen for months, many steps towards implementation were initiated behind the scenes in preparation for the redesign. A repository of standard question blocks (used for all Social Surveys at Statistics Canada) has been created. The existence of this repository should reduce development time in the Contact, Household, and Exit components of the applications. A prototype of the household roster blocks has been developed, as well as a proof-of-concept prototype of the new multi-datamodel program manager. A generic browser (in Visual Basic) used by interviewers to select specific cases, as well as standard reports (in SAS) have been developed and tested. They will be implemented in production for several surveys before this redesign takes place. Having this preparation work done has allowed us to focus on the user-defined aspects.

In order to manage the project effectively, a detailed list of activities and tasks has been created. This schedule leads to an implementation date of July 2003 for the new CAPI environment, and January 2004 for the CATI implementation. High level requirements were gathered from the key stakeholders, and further discussions within the multi-disciplinary project team have lead to detailed specifications for all components. The programming and testing of key components is well underway. At each step team members ensure that documentation of the process, as well as the product, is not forgotten.

Prior to development, a specifications review process was established whereby a team comprised of members from the survey operations, subject matter area and application developers work together to develop specifications. This has resulted in superior specifications that have helped to expedite development.

To ensure a high quality product, development and testing at progressive levels is promoted for all collection applications and this project is no different. Once the initial authoring is complete, blocks of the collection application are passed to team members from the survey operations area as well as the other stakeholder divisions for initial testing. After at most three rounds of block level testing, the blocks are put together into questionnaire components incorporating the flow logic between the blocks. Next, the questionnaires are integrated together with a program manager and integrated testing of the entire application begins. Once the integrated application has been approved, end-to-end testing takes place in the survey operations area, where a mock-up of Regional Office environments allows the simulation of the actual interviewing environments. This final round of testing involves the input and output files generation, as well as simulating the transfer of data to and from the regional offices.

Another key section of the schedule is the tasks and timeframes allowed for development and training of Regional Office staff. Training and reference material for all levels of personnel will be prepared, and site visits from knowledgeable team members will help to clarify outstanding issues.

A capacity test is planned in one regional office well ahead of the production date, to ensure that there will be no issues with the volume of cases being processed on the Blaise servers, now that a larger volume of true CATI cases will be managed. To allow sufficient time for field tests and the subsequent evaluation, two adjacent months of field testing have been set aside, after which a decision will be made as to whether the project should proceed to phase-in the remaining Regional Offices one at a time over a period of several months.

6. Conclusion

For a project of such a grand scale, and far-reaching implications if not done well, a phased approach with step-by-step implementation is the best scenario. Some of the hardest potential defeats we have come across are a commitment in terms of personnel and monetary resources for the project from all stakeholders. The greatest challenge is to balance new development and production at the same time. Many of the key players are the same, and have conflicting priorities. This has caused delays in a schedule that seems appropriate. Regular team meetings, as well as a steering committee that has the authority to make decisions and re-orient priorities as required, are essential to the success of such a project.

The move towards a true CATI environment using the Blaise call scheduler for the CAPI/CATI surveys should provide a more efficient way of working in the call centers, as compared to the manual selection of cases that occurs at the moment in pseudo-CATI. One concern does remain with the client area and survey operations division: the first survey being converted, the Labour Force Survey, has a 7 day collection window each month. The notion of best time and day to call are very important to this survey, as the collection period is short and the sample size is large. We will be experimenting with new processes surrounding the setting of the call scheduler parameters, as well as simulating time slices during the day in order to try to maximize the number of calls to a respondent during their "best time." This may result in several daybatch updates during each day for the first several days of collection, and different procedures during the last few days of collection. Extensive simulations are being planned, to assure the developers and the clients that the new model, as well as the Blaise call scheduler, will be effective and perform as expected.