Five years' experience with CAI and Blaise

Fred Wensing, Eden Brinkley, Australian Bureau of Statistics

Abstract

This paper summarises five years of experience by the Australian Bureau of Statistics (ABS) with Computer Assisted Interviewing (CAI) and the use of Blaise.

The paper will give a brief history of the use of CAI in household surveys at the ABS, commencing with a small scale trial in 1994 and progressing through its use in 5 major surveys. The paper will discuss the issues considered in the business case for CAI and some of the concerns which led to it not being implemented as widely as first planned.

The ABS decided to use Blaise for CAI in household surveys after a detailed evaluation of the software was made in 1994 and in recognition of the plans that were being made to further develop the product. The ABS has used version III of Blaise since the early beta test versions became available in 1995. Our experience with Blaise software has been varied and much has been learned about building instruments and applications. In recent months some of our Blaise facilities have also been converted to make use of Blaise 4 Windows. The paper will give a summary of those experiences and highlight some of the lessons learned.

Five years' experience with CAI and Blaise

Fred Wensing, Eden Brinkley, Australian Bureau of Statistics

Introduction

Computer assisted interviewing (CAI) has been employed as a collection method for a selection of household surveys at the Australian Bureau of Statistics (ABS) since 1994. During that time we have maintained a stock of 230 notebook computers, conducted over 17 pilot tests and 6 major surveys involving more than 250,000 CAPI interviews, 400 interviewers and 150 office staff.

This paper summarises the ABS experience and highlights the major issues associated with the decisions to make use of CAI. Some of the difficulties experienced are also discussed.

When CAI was first employed by the ABS in 1994 version III of Blaise software (for DOS) was just becoming available in a beta test version. Our use of Blaise has progressed through a number of enhancements of the software and we are now commencing to make use of Blaise 4 Windows.

This paper also describes our experiences with the development of the infrastructure associated with CAI, in particular the key role played by the Blaise software suite.

Brief history of CAI in the ABS

The use of CAI in ABS household surveys was examined in 1993 and a case was prepared for its introduction subject to a comprehensive test being conducted to ensure that it was viable. Previous considerations of CAI had concluded that systems were not advanced enough. But the situation had changed, with a number of overseas agencies adopting CAI.

A small scale trial of CAI was conducted in July 1994 using the Household Expenditure Survey (HES). The HES was deliberately chosen as it was a large survey questionnaire involving more than 800 questions. The trial was conducted on a sample of 450 households using 10 interviewers who were each equipped with a 486 notebook computer (monochrome screen). The test also included the transmission of data via E-mail.

On the basis of the success of this first trial of CAI a proposal was developed and approved in 1994 for the introduction of CAI to all household surveys by 1997, including the monthly Labour Force survey.

From the end of 1994 and through 1995 work proceeded on the infrastructure required for CAI, such as field systems, office systems, transmission systems and instrument design. During that time a further four tests were conducted to check various aspects of the infrastructure, office and field procedures, and training material.

In order to ensure that the systems could support the requirements of a monthly survey with a tight schedule, a major trial of CAI was conducted in the first six months of 1996. This trial made use of 80 interviewers and a rotating sample of 4,000 households per month which was conducted in parallel with the existing monthly Labour Force survey (which used a pen and paper collection methodology) to enable direct comparisons to be made. This major test provided valuable experience which contributed further to the development of systems.

At the same time, indications were emerging that the original cost estimates for CAI were likely to be significantly exceeded particularly in the area of hardware. This led to a management decision taken in late 1996 to pull back from full implementation in all household surveys. Given that the stock of notebook computers had reached 230 by then (in preparation for the first production use in an ABS survey), it was agreed that CAI should continue to be used for a number of major stand-alone surveys but that plans to employ CAI for the monthly Labour Force survey should be abandoned for the time being.

The first production application of CAI at the ABS occurred with the second wave of the longitudinal Survey of Employment and Unemployment Patterns, conducted in September 1996 with a sample size of 8,500 and interviewer work force of around 200. The most significant feature of this survey was the use of dependent interviewing in which information from the previous wave of the survey was used within the CAI instrument to guide the conduct of the interview.

CAI went on to be used in the following major surveys conducted by the ABS:

- Survey of Mental Health and Wellbeing (1997) using a client-provided CAI instrument;
- Third wave of the longitudinal Survey of Employment and Unemployment Patterns (1997);
- Survey of Disability, Ageing and Carers (1998);
- Household Expenditure Survey (1998/99);

- Australian Housing Survey (1999).

With the stock of computers coming to the end of its useful life in 1999, it was necessary to develop a new case for continuation of CAI. With careful structuring of the survey program and due consideration of the costs and benefits, a successful case has now been presented to management to enable CAI to be continued for major surveys in the future. The details of this case are discussed later.

Initial justification for CAI at the ABS

The case for CAI at the ABS was based on expectations of improvement in data quality, survey timeliness and cost effectiveness. At the time of preparation of the initial case for CAI, these benefits were described in the writings of a number of authors (Weeks 1992, Manners 1992, Martin, O'Muircheartaigh, Curtice 1993). The attractions of CAI could be summarised as:

Improvements in data quality - Interviewers are directed automatically by the computer to the next relevant question, eliminating any sequencing errors. Editing of responses can be done during the interview, allowing queries to be answered directly and immediately by the respondent. Introducing questioning based on responses given in previous interviews (dependent interviewing), can provide greater consistency and accuracy of data over time.

Improvements in timeliness - Timeliness of survey processing can be improved as data is available in electronic form at the end of the interview. Time spent on handling paper forms and entering data into the office computing system is eliminated. Potentially, the data delivered by the survey instrument could be a clean unit record which would be fed directly into the survey output system. In practice, for a complex survey, some post-field processing may still be required to clean records.

Reduction in survey operating costs - CAI offers savings in a number of areas of survey operations. Printing of paper forms is no longer required. Physical handling of forms (bundling, processing, storage) is also removed. Office editing and coding is greatly reduced. Clerical data entry is eliminated. It is also possible that costs of survey management can be reduced. Any savings in these areas need to be counterbalanced with the additional costs of hardware, software and development time accruing to CAI implementation. As these costs are greatest at the outset, savings are likely to be realised only in the longer term.

It was recognised that the cost reductions mentioned above may not be as great at the ABS as some other agencies because clerical data entry had been replaced with optical mark reading (OMR) of forms for household surveys in 1989. The remaining benefits were accepted by ABS management, subject to confirmation through the testing programme.

The small scale trial, which was conducted in July 1994, concluded that CAI was indeed viable and could deliver quality and timeliness improvements with possibly some cost savings. The trial also showed that it was relatively easy to introduce CAI to an interviewer workforce who have had very little computing experience. Respondents were not at all concerned by the use of computers in the field. It was on that basis that continued use of CAI was approved at the time.

Major issues which received attention

Choice of software

One area to receive early attention was the selection of suitable CAI software. Due to the special nature of CAI it was preferable to use an existing software package rather than to build one of our own. Evaluation of available software was carried out against a set of defined criteria which can be grouped under the following headings:

Facilities provided - the software must provide adequate functionality for the survey business and primarily for the interviewing part. Features should include support for editing, routing, coding, derivation, complex data structures, error recovery, management and analysis.

Metadata - this area of concern related primarily to how the product was to fit into the ABS corporate infrastructure which placed a significant emphasis on metadata. Features would include the ease with which metadata could be used by or supplied to other systems.

Usability of the software - this related to how efficient and easy the product is to use. The software was assessed on such characteristics as the ease of learning, speed of questionnaire development, provision of development tools, documentation and help facilities, usability by questionnaire designers who are not necessarily IT professionals, maintainability and readability, reusability of code between surveys and training required for all types of staff involved.

Usability of the resulting collection instruments in the field - this related to how usable the resulting collection instruments are in areas such as layout and readability, performance and efficiency (including response time for all functions), on-line

help for all functions, robustness and error recovery (including protection against data loss) and overall acceptance by interviewers.

Post-field processing - this addressed how further processing of data is supported, including derivation, weighting, tabulation and ease of linking to alternative post-field processing products.

Other ABS computing environment issues - these requirements were to ensure that the product would fit satisfactorily into the ABS computing environment, and covered such considerations as compatibility with other software, file management, communications and security.

Future directions - this addressed relevance of future directions as perceived by the ABS, in areas such as transition to Windows, links to external systems and adaptablity.

A brief assessment of the alternatives was made and two serious possibilities (one of which was Blaise) were then given closer scrutiny and testing. In the end, Blaise (version III) was selected as the most appropriate software for CAI in the ABS due to its language characteristics (simple syntax and block structure), its metadata-centered approach (shared by all the tools) and the availability of data and metadata conversions to interface with the ABS environment. While there were some limitations in the software at the time (being 1994) it was known that these were being addressed by the software developers. It was further recognised that the developers of Blaise, Statistics Netherlands, had very similar goals and priorities to the ABS and was pursuing developments in accord with those at the ABS.

Selection of hardware

The minimal technical requirements for notebook computers required to operate Blaise III software in 1994 were determined to be:

- Intel 486sx processor (this was later changed to 486dx in anticipation of additional demand by the software)
- DOS 6.2 operating system
- 4 megabytes of RAM (preferably upgradable)
- 80 megabyte hard disk (allowing for future requirements)
- Full size keyboard
- VGA graphics screen (to handle 80 by 25 characters of text)
- Unit to be connectable to modem, external keyboard and/or monitor

Non-technical requirements included:

- Low physical weight (preferably less than 2.2 Kilograms)
- Screen to be easily readable (colour optional)
- Minimum of four hours continuous battery life
- Unit that is easily to carry and use while sitting or standing
- Shock resistant case
- Cost
- Future plans for upgrades and support

Weightings were assigned to the functional requirements of the notebooks with non-technical issues being given a high level of consideration in the selection/evaluation process. The physical and ergonomic features/characteristics of the notebooks were considered very important. Consequently, the following criteria were given high weightings as mandatory requirements:

- low physical weight
- high quality display screen
- long battery life
- reliable battery life indicator

Tenders were received from 14 suppliers which were then shortlisted on the extent to which each tender complied with the mandatory functional requirements specified. The tendered notebooks were then ranked on their value for money, calculated as a ratio of the cost over their score on technical worth. Six shortlisted tenderers were asked to provide the ABS with appropriately configured hardware to allow performance tests to be conducted.

The shortlisted notebooks underwent comparative performance testing using the CAI application. A physical evaluation of these notebooks was also conducted examining ergonomic issues such as keyboard feel and screen clarity. Some interviewers were also invited to assess the equipment.

Following the detailed assessment described above, and further discussions of price and future plans, the decision was made to purchase the IBM Thinkpad 340 (486dx 75Mhz colour with 4Mb RAM).

Purchase and delivery of the notebook computers were made in three lots with a total of 230 being purchased by 1996 when the decision to curtail CAI was made. By the time the third lot of computers (150 in number) was ordered the 486dx processor was no longer available and the IBM Thinkpad 560 (pentium 75Mhz colour with 8Mb RAM) was substituted.

At about the time of the placement of the third order for notebook computers (mid 1996) it was quite noticeable that 4Mb of RAM was inadequate for good performance of version III of Blaise and RAM upgrades were purchased for all the existing stock of 80 computers. This additional outlay was quite expensive.

While the notebook computers were certainly adequate for the operation of CAI applications (all in DOS) when purchased, the speed of change in software was such that by the time they were two years old they were inadequate for operating anything else. In particular, they were not suitable for Windows applications which restricted the possibilities for using Blaise 4 Windows.

The final price of notebook computers in 1996, including an encrypting modem and a three year service arrangement was almost 100% more than the estimate prepared in 1994. The higher than expected costs were a significant factor in the decision to pull back from full implementation of CAI in all household surveys

Occupational health and safety

A consultant was hired to conduct research into feasible and practical ways of minimizing the risks associated with the use of notebook computers in CAI. A primary objective of the consultancy was to determine whether a suitable ergonomic aid could be designed to support the weight of a notebook computer when used in a standing position.

The development of such an aid to interviewing was considered important because most face-to-face interviewing of the Labour Force survey in Australia, our largest and most important household survey, is conducted at the door. Use of the notebook computer in these circumstances, without a physical support of some kind, was considered to be an occupational risk. To encourage interviewers to go indoors risked increasing the enumeration costs.

The consultants designed a prototype collapsible frame which was capable of supporting a notebook computer with safety. Around 80 of the initial design were manufactured and used in the 6 month trial of CAI in 1996. Following feedback from the interviewers the design was revised to make it more collapsible and more readily carried in the canvas bag used to carry the computer. The design has subsequently been patented.

Apart from the ergonomic stand, revised field procedures have also been implemented for the major surveys undertaken using CAI to date. In particular, respondents are requested to make available a suitable table at which the interview can be conducted. Relevant literature and training is also routinely given to all interviewers undertaking CAI surveys about the health and safety aspects of the use of notebook computers.

Security

Security of respondent data is an important issue and a number of measures were employed to ensure that the data was safe both on the notebook and in transmission. The following measures were employed:

No floppy disk access - notebook computers were configured without an internal floppy disk drive to safeguard against unauthorised copying to disk.

Screen saver – screen saver software was included to prevent access to the data if the notebook was unattended for any period of time (such as during transmission).

Encryption and passwords - the notebook hard disk was encrypted (using Blockit software) so that data access was not possible without a password. The encryption was also linked to the configuration of the notebook which made it impossible to access the data on the hard disk if the configuration has changed. Passwords need to be updated by the interviewer on a monthly basis.

Encrypted transmission - transmissions were carried out via an encrypting modem which ensured that the transmitted data cannot be interpreted. While encrypting modems are much more expensive it was justified to protect the data.

Secure logon procedures - logon to the FTP computers in the office (for transmission) was done through a system which requires the entry of a 6-digit pin number obtained from the readout on a smartcard (or key fob) which provides a sequence of 6-digit pin numbers which changes once every minute. The FTP computer in the office matches the generated pin number with the card number and registered owner (interviewer) before further access to the server is granted.

While this kind of system was more complicated to establish and the procedures took a while to master, there was a general acceptance that our duty towards respondent privacy warranted it.

Infrastructure built for CAI

The development of CAI facilities happened at the same time as other significant developments were taking place at the ABS. These were:

> *Development of a corporate data warehouse* which provides a central store of all released and releasable statistical data along with comprehensive metadata and access facilities. By policy all major systems in the ABS are required to link with this facility and eventually deliver data and metadata to it.
>
> *Redevelopment of processing facilities for household surveys* to overcome concerns about the viability of existing systems (which use legacy software). Developments were also aimed at taking advantage of new technologies and linkage to other systems such as the corporate data warehouse.

When CAI was first proposed there were virtually no support facilities in existence in the ABS. The plans for CAI therefore included the development of infrastructure to support it. The main components which were developed were:

- Survey Development Tool to facilitate preparation and specification of the survey concepts, data items, question modules and derivations to feed the CAI instrument preparation
- A field management system called the Interviewer Workload Management System
- An Office Management System
- Transmission facilities

These are described briefly below.

Survey Development Tool

Development of a survey questionnaire or instrument commences with the preparation of detailed data item specifications. Before CAI the specifications had been prepared in the form of documents or spreadsheet tables which described the data items in detail, including relevant respondent populations, definitions and response categories. The specifications provided in this way were then converted to paper questionnaire designs by survey staff using publishing software. The focus of specification in this case was the design of questions with little attention being paid to the edits, derivations and output items at this time.

The advent of CAI brought with it the possibility of applying edits and derivations in the instrument. It also opened up new ways for designing simpler more modular questionnaires, as the number of pages (and hence the need for more convoluted sequencing) was no longer a constraint. A way of specifying the questionnaire for more optimal Blaise coding and questionnaire validation was also required. These factors significantly increased the complexity of the specification process. In addition, because CAI requires the involvement of more people (eg. programmers), there was a need to make the specifications more widely accessible within our network environment. At the same time, the emergence of the corporate data warehouse was also placing pressure on areas to make their metadata available for linking with other collections.

As a result of these pressures it was decided to develop a special tool called the Survey Development Tool (SDT) which could be used to store and manage the specifications (or metadata) for all household surveys and particularly the CAI surveys.

The design objectives of the SDT were to:

- have a single centralised store of survey metadata (or specifications);
- accessible to all who needed to specify, examine, modify or use the information;
- assist with the management of the specification process;
- assist with the instrument development and validation processes;
- have the possibility of generating Blaise code;
- link with the corporate data warehouse; and
- be easy to use.

The SDT was developed as a Lotus Notes database which contained various forms to record all the specification details, and structured views to assist with locating and reviewing the contents. Lotus Notes was chosen because it was, and still is, the standard group-ware used by all staff at the ABS and was therefore considered to be the most user-friendly option.

The basic building unit of a questionnaire specification under this tool is a module of related questions. Such a module may measure a single concept or group of related concepts and will generally be applicable to a defined population in the survey. Earlier prototypes using a question as the building unit gave too many elements and were found to be difficult to manage. For convenience a module can be seen as containing up to 10 questions although provision was made for larger modules. A module has the advantage of having closely related questions kept together and could be small enough to treat as a single unit

that can be stored and reused for other surveys (without the likelihood of breaking apart). The module concept was also similar to the block construct used in Blaise.

The SDT was designed to be linked with the Data Warehouse and a cycle of metadata exchanges were envisaged between them to ensure that data items and concepts were aligned with those of other collections. Control of the development and collection processes were to be linked with other corporate management facilities of the Warehouse. Integration of the SDT with Data Warehouse is described in detail in a separate paper (Colledge et al 1996).

A production prototype SDT was constructed and used for the first few CAI surveys. The initial functionality was limited to the development and storage of specifications, with some export functions to assist with other aspects of survey work, including an early attempt at Blaise code generation. Unfortunately, the decision to pull back from full implementation of CAI in all household surveys in 1996 meant that further work on this tool was also stopped. Nevertheless, the SDT has continued to be used for all CAI surveys and is still in use today.

Further development of the SDT will occur as part of future work on the CAI infrastructure with particular attention being paid to the links with other systems.

Interviewer Workload Management System

Soon after the commencement of work on CAI it was realised that a system was required to manage the interviewer work in the field. Since there was no generic facility available from other agencies using Blaise (particularly not in Blaise III) it was decided that one should be built.

The first prototype system for field management was built in the ABS using Pascal. This facility simply fed a list of addresses in a screen to the interviewer who selected them for interview. This was adequate for the first few tests using CAI.

As soon as Maniplus became available (in beta form in late 1995) serious development started on a comprehensive Interviewer Workload Management System (IWMS). Maniplus, being part of the Blaise suite, had the distinct advantage of being able to extract information from Blaise data files for display and then update those same files with interviewer remarks, address changes, appointment information and status codes, all fed from Maniplus dialog boxes and display screens. The compiled facilities of IWMS had a consistent look and feel to the Blaise instrument and made the transition between field operations and interview quite seamless.

The IWMS as developed in the ABS provides the following major functions:

- receipt and transmission of data and programs from the office
- access to training materials
- listings of parcels of work (known as a workload)
- listing the addresses to be visited
- display the names and details of household members
- recording of appointments
- backup facilities
- control the flow of interview between instruments
- ability to update the status of records

The IWMS was intended to be a generic facility capable of managing all surveys in the same way. However, the differences in requirements for complex surveys were such that the IWMS needed to be adjusted for almost every survey. Differences have occurred in information to be displayed, different flow of interviews with different instruments, different recording of status and different appointment details.

The main characteristic of the work provided to interviewers is that it is organised into parcels called workloads. A workload is a group of addresses (usually in the same local geographical area) which are to be administered the same survey. The number of addresses in a workload varies according to the complexity of the survey but is approximately what can be achieved in a one week or two-week interviewing contract. The workload has its origins in the paper systems but was kept as something that interviewers and office staff were used to dealing with.

Supporting the workload concept described above has made the management of cases under CAI a little more complicated. This was particularly so when one or more cases need to be referred to the office, or to another interviewer, or when transmission of some completed cases occurs. Because all cases in a workload are generally kept together, transmissions would involve all records in the workload no matter whether those cases were completed or not. This can result in the same record being transmitted to the office a number of times as the workload becomes progressively more compete. While this was initially thought to be a problem, it turned out to have a positive aspect in that it provided multiple copies that could form

backup versions if the workload became corrupted or a transmission was lost for some reason. The problem with potentially multiple copies of a workload being available, however, was to know where the latest copy of any case was.

The solution which was devised for tracking of individual cases in the field and the office involved the use of a control file which is maintained as a Blaise data file in the office. The control file has a record of every case and an indicator of which interviewer or system component has the most up-to-date details for the case. The operation of this control file is described in more detail in a paper on the Office Management System (Henden et al 1997).

The main design feature of the IWMS is the use of "buttons" and fields that can be updated on the screen. The use of pull-down menus, although possible, has not been employed in the IWMS because most interviewers were not familiar with computers and track balls, making buttons operated by Alt-key operations (or by tab key movement) easier to use.

Some performance problems were encountered in the refreshing of lists on the screen particularly when a case or household record had been updated, or when the interviewer applied a sort operation to the list of addresses. These problems were overcome through the use of indexing on the underlying data files.

The functionality of the IWMS has served our CAI surveys well and the latest version provides a good model for the one which will be written in Maniplus under Windows for future surveys. It is expected that some refinements will be made to make it more generalised and easier to maintain.

Office Management System

An office management system (OMS) was developed to manage the flow of work to and from the interviewers and to enable the received data to be checked and prepared for processing.

The system was expected to provide the same degree of management and flexibility which existed with non-CAI systems in existence at the time.

The OMS was developed during 1996 and now provides the following functionality:

- assign selected addresses to workloads;
- assign workloads to interviewers;
- prepare respondent records for each workload from data collected previously;
- collate workload data for transmission to each interviewer;
- collate associated instrument software for transmission to each interviewer;
- manage transmission to and retrieval of data from interviewers;
- provide for reallocation of workloads or individual respondents from one interviewer to another;
- enable the examination of received data to clean up 'dirty' records or to resolve queries raised by interviewers;
- provide for the coding of some data fields in the office (eg. occupation, industry);
- provide different levels of functionality for central and regional staff;
- monitor survey progress and extract management information; and
- export a clean data file for further processing.

The OMS facilities were built using a mix of Oracle/SQL, Blaise and DOS. The Oracle components were required to interrogate and manage the links with the sample address lists (based in Oracle tables). The Blaise components were required to load and prepare the instruments with records passed to it from the Oracle facilities, then package them and copy them to the FTP server for interviewers to collect (by transmission).

Once the data is returned from the field (by transmission), Blaise based facilities are used to unload the data and present the records to office staff for checking. At that stage additional coding may also be done. Survey data is then collated from groups of interviewers into regional data files which are then sent to the central office for final checking and export to ASCII files for processing by other systems.

A comprehensive description of the OMS is contained in a paper presented at the Fourth International Blaise Users Conference in Paris (Henden et al, 1997).

As mentioned earlier, a key feature of the OMS is the use of a control file to track the location of respondent records between sample systems, office processing and field work. Each time a respondent record is moved to or from any of the processes which form part of the office and field systems, the control file is first checked to determine whether that process has control of the record. If it does then the process is activated and a record is written or modified in the control file to indicate the change in control.

Transmission

Transmission to and from the interviewers is managed through a series of regional FTP (file transfer protocol) servers connected to the LAN and protected with a "firewall" which certifies the user (both internal and external) and controls access to the facilities on the server. Dial-in access to the server is only possible through encrypting modems which ensure that all communications are authorised and protected. Access is further controlled through a technique known as "strong authentication" which requires the interviewer to carry a smart card (or key fob) that generates a single-use password with a limited validity time (a few minutes) that can be validated by the server.

Experience with Blaise

As mentioned earlier, the ABS decided to use Blaise for CAI in household surveys following an evaluation of the software in 1994 (see earlier section). All the Blaise systems and survey instruments mentioned in this paper were developed using version III of Blaise (various releases).

Some of the issues that have arisen in our use of Blaise are discussed below.

Developing software

A succession of updates in the early days of version III of Blaise did provide us with the challenges of testing the software and then convincing our development and field operations staff that the new release would (generally) be better than the old with risks of problems in the field being minimal.

Introducing new releases of any software to production systems does have an element of risk and it is pleasing to note that generally the experience with Blaise upgrades of version III has been uneventful. The software release practices of Statistics Netherlands have also improved considerably with very good versioning and update history now being made available.

Extent of use in ABS

Although Blaise has been approved for use in the ABS for household surveys it has been restricted to data capture and limited office processing functions as described in this paper. The main reason for this has been that Blaise III is a DOS product and many ABS applications, as well as the general office software, had been operating in a Windows environment for some time. A further reason was that many of our processing systems made use of other software (eg SAS or Oracle/SQL) which was where the expertise was and there was no convincing reason to change. Some of components of the Blaise software suite (eg. Abacus, Bascula and CATI call management) have received little attention at the ABS as a result.

The advent of Blaise 4 Windows does open up opportunities for more extensive use of Blaise software in applications other than household surveys. However its use may be limited by the ability for it to fit into the broader ABS corporate infrastructure, in particular, the Data Warehouse and our office group-ware (Lotus Notes).

Some experiences in integration with Blaise III is contained in a paper presented at the Fifth International Blaise Users Conference in Lillehammer (DeMamiel et al, 1998) which discusses the use of Blaise in the data capture of diary information for the Household Expenditure Survey and the exchange of data and messages with office software.

Experience in programming in Blaise

Given the amount of infrastructure that needed to be built we quickly established a fair amount of expertise in the programming of Blaise applications. Most of the training was on-the-job and made use of the manuals and sample code provided with the software.

At the time of peak demand for Blaise experience in the ABS (probably 1996) there were 3 officers who were well experienced in the programming of Maniplus and another 12 officers who were involved in writing the Blaise code for various instruments. With the recent pull back in CAI, however, those numbers have reduced to around 3.

Who should write Blaise code

There are differing opinions on who should write the code for CAI instruments and a fundamental issue is whether the same or different people are responsible for content development and CAI programming (Kinsey and Jewell, 1999).

At the ABS, the writing of Blaise code for instruments has been carried out by survey branch staff rather than programmers, although advice and training has been provided by them. People writing Blaise code at the ABS have tended to be staff who have some understanding and experience in survey processes and who posses an aptitude for or skills in programming work. The specifications for CAI instruments are supplied by development staff or researchers who have training or experience in the design and testing of questions. In order to standardise the specification process there has been a requirement to make use of the Survey Design Tool (SDT) described earlier.

While there is no objection to the suggestion that development staff could write their own instruments there is a feeling at the ABS that people with training and technical skills in question design and testing are better to focus on questionnaire development work, while others with skills in programming are better to focus on the activity of Blaise programming. Given the relatively high staff turnover in our survey infrastructure areas it would be difficult for someone to become highly proficient in both areas of work, but more importantly the creative tension between the question designers and Blaise programmers has been a healthy one for the surveys where we have pushed the limits of what is possible. There is also an expectation that the further development of CAI facilities will result in more code being "generated" by the system which supports the SDT.

The development of Maniplus facilities such as the IWMS and OMS described above has always been considered the function of programmer staff at the ABS.

Software interaction problems at the ABS

The functionality of version III of Blaise has continued to improve over the years although a few problems in the software and its interactions with ABS technical infrastructure remain. These are:

The sticky key problem - for some unknown reason, the Blaise III software occasionally produced a locking of certain keys on the IBM 340 Thinkpad notebooks. Despite sending the computer to the Netherlands for further investigation and upgrading the system BIOS this problem never disappeared. It was tolerated by interviewers who developed ingenious methods for "unsticking" the keys so they could continue interviewing.

The network interaction problem - use of Blaise III produced an occasional problem on the ABS network (Banyan Vines) which resulted in a "locked file" situation on the ABS logo file (in a read-only directory) and terminated the execution. This problem was never resolved but could be avoided by removing the offending logo file from the software set.

The shared file problem - use of shared network files (for code look up) in an office data entry system (for entry of diary data from the Household Expenditure Survey), involving 6 coding staff sharing the same network drive, produced mysterious "crashes" which were only solved by localising both the lookup files and the software. (This solution also produced better performance).

Resolution of some problems is hampered by an inability to run some of the diagnostic software which is only available to the software developers, and the cost of arranging for someone to come halfway around the world to be on site to investigate matters locally.

It is expected that the Windows version of Blaise will be more compatible with other software written for Windows.

Support from the supplier

Blaise software was developed by Statistics Netherlands and, despite the separation of distance, relations between the ABS and Statistics Netherlands in relation to Blaise have always been good. The problem of whom to contact, which was present in the early days of ABS use of Blaise, have largely disappeared with the creation of the blaise@cbs.nl e-mail address. Nevertheless, communications which are separated by time and distance can sometimes be a cause for frustration if systems are dependent on an urgent solution to a problem, no matter how small. The inability to share one's problems face to face with the Blaise developers is also a disappointment, although new possibilities could be opened up with video links over the internet.

The manuals for Blaise software and the sample programs which accompany the software are very useful sources of information about its operations.

Sharing experiences

Given the extent of use of Blaise around the world, there should be an increasing pool of people to share experiences with. The establishment of the Blaise user group with its newsletter and website, the Blaise Corporate Users Board, and series of international conferences provides excellent avenues for the sharing of experiences to occur.

ABS staff have tried to maintain contacts with key people in the Blaise fraternity but it can be time-consuming because most communications are by e-mail which takes time to compose. Furthermore, most people being consulted are heavily committed in their own work areas and not always able to respond in a timely way.

Split datamodel instrument design

One of the advantages that CAI provides is the ability to store survey data collected information from a complete household within one physical record. While this keeps the members of a household together it has the limitation of providing a more complex instrument (essentially an array of person records) with many more fields, leading to a demand on both memory and storage space. A single household record containing all persons also results in a nominal limit being placed on the number of persons that can be included (currently 10 in the ABS).

When problems were encountered with lack of memory on the 4Mb notebooks, the possibility of separating the individual questionnaires from the household was suggested as a way of reducing the demand on memory and improving performance (Wensing, 1996).

This proposal became known as the split datamodel design in that it splits the physical datamodel into two or more parts. Once the household part has been completed the interviewer drops back into the IWMS which presents a dialog box of the personal interviews that remain to be conducted. Each required personal instrument is initialised with corresponding data from the main household instrument. Edits in the personal instrument which need information from the household instrument can still be carried out using relevant Blaise statements that can extract (lookup) the required data from the corresponding household record.

The disability survey 1997 was the first survey to employ the split datamodel design and, while it complicated the management of questionnaire flow in the IWMS, it was considered to be a success and has been used in most CAI surveys since.

The main disadvantage of the split datamodel is that both parts of the datamodel need to be present for any processes (such as interviewing, editing or transferring data) to be applied successfully. On the other hand, a split datamodel enables each part to be treated as an instrument on its own from the point of view of the data within it being considered clean and available for use.

Other instrument design issues

CAI has brought with it a temptation to collect more information because the constraints of a paper form do not exist. Thus, it is easy to add a field here and there, or to collect text responses where previously there may have been a simple tick box. Also, the limits on collecting the details of instances of something (eg. loan details in the HES) can be pushed out beyond those which may have applied to a paper form. This has resulted in much larger record structures and greater volumes of data to be processed, possibly at the expense of efficiency.

Integration with other processing

Use of CAI enables other processes, such as imputation and derivation of new variables, to be brought forward in the survey cycle, thereby saving output time. Ideally some of these processes should happen in the electronic survey instrument and Blaise provides the functionality for that.

While it was the intention of the CAI development team at the ABS to incorporate additional processes in the instruments that have been developed over the past five years, in practice only those items deemed necessary for the conduct of the interview have tended to be derived. In only one CAI survey at the ABS has this integration of derivations into the instrument been achieved, and even then the processing was executed in the office. In all other cases the data has been extracted from Blaise files and processed in other software (usually SAS).

A major problem encountered in transferring data to other software for processing was the need to translate the Blaise field names into a set of unique 8-character names. The Cameleon software component of Blaise provides a function to "generate" these unique 8-character names but the outcome suffers when an existing Blaise field name ends in a numeric character (and the field is used in an array). This led to the preparation of a naming convention which kept those parts of Blaise names both short and devoid of numeric character endings (Wensing, 1999).

The issues of integration of other processes will be looked at more closely with the next round of developments for CAI at the ABS. An emphasis will be placed on translating the questions to as complete and final a data item as possible in the instrument and keeping the re-processing of data in the office to a minimum.

Business case for renewed CAI

Despite the pull back from CAI in 1996, there has remained a positive opinion among survey staff at the ABS about the prospects of continuing to make use of CAI. With the stock of notebook computers approaching the end of their serviceable life, a business case was prepared in 1999 to recommend the continuation of CAI for larger household surveys.

The business case for the re-development of CAI focussed on three key aspects:

Better utilisation of notebook computers – by structuring the schedule of surveys appropriately so that there are no significant peaks and troughs in the work, it is possible obtain more usage from the stock of notebooks than had been obtained previously. This means that less notebooks are needed, saving on cost.

Determination to actively pursue the re-engineering of all processes – CAI has the ability to deliver data which is more complete in that more of the processing operations (eg. coding, derivation) can be done in the field. These gains are only possible if the development and operational processes involved are changed (re-engineered).

Management of the project - a strong commitment to carefully manage the project, so that all of the benefits that the methodology can provide would be realised at some point in the future.

With this approach, it is expected that CAI will produce improvements in quality and timeliness and some small savings in ongoing costs that will be worth the investment in hardware and software development.

Conclusions

What can be seen from the ABS experience is that CAI requires a fair amount of infrastructure and we have built almost all of it ourselves. Given that other agencies using CAI need similar facilities there would be merit in some joint developments occurring between agencies.

The Blaise suite of software has provided the underlying tool set for most of the facilities we have built. It may be appropriate for more infrastructure elements to be built into Blaise software. One example where this has been done is in CATI call management which operates on special blocks that are required to be added to CATI instruments. Maybe features like this could be added to support case management via a generic case management tool (replacing our IWMS).

The ABS has gone from being an early user of Blaise version III to a late starter in Blaise 4 Windows. When Blaise was chosen for CAI in 1994 other developments at the ABS were well advanced in the use of Windows, and office software was about to move to Windows 95. Use of Blaise III for CAI kept those developments in DOS when other staff were becoming used to a Windows environment. This meant that use of Blaise beyond data capture was unlikely to happen at the ABS. This may change with Blaise 4 Windows.

Enhancement of Blaise and other software over time has increased the demand on hardware such that our CAI notebooks are now suffering in performance. Software developers, such as those developing Blaise, need to be aware that users do not always have the latest most powerful hardware and resist the temptation to make their software bigger and faster without due consideration of the equipment implications for their customers.

The ABS has a diverse technical environment and only a few applications have been built in Blaise. With the move of Blaise to the Windows environment it is hoped that there will be more compatibility between Blaise and other applications. That way there will be greater opportunity to develop integrated solutions that draw on the best features of each software. One particular need that arises out of integrated solutions is the need to be able to convert data and metadata more readily from Blaise to other forms (and vice versa).

Finally, much has been learned about CAI at the ABS over the past five years and we are well positioned to launch into another period of development in preparation for the next series of household surveys to use CAI commencing in 2001.

References

Colledge M., Wensing F., and Brinkley E. (1996), "Integrating Metadata with Survey Development in a CAI Environment" *Proceedings of the Bureau of the Census Annual Research Conference and Technology Exchange*, Washington DC: US Bureau of the Census, pp. 1078-1100.

DeMamiel M., Wensing F. and Green, B. (1998), "Diary and Office Processing: Integrating Blaise with other Facilities", *Proceedings of the Fifth International Blaise Users Conference*, Oslo: Statistics Norway, pp. 66-77.

Henden M., Wensing F., Smith K. and Georgopoulos M. (1997), "An Office Management System in Blaise III", *Proceedings of the Fourth International Blaise Users Conference*, Paris: INSEE, pp. 107-122.

Kinsey S. H., Jewell D. M. (1995), "A Systematic Approach to Instrument Development in CAI", *Computer Assisted Survey Information Collection*, New York: Wiley & Sons, pp. 105-123.

Manners T. (1992), "New Developments in Computer Assisted Survey Methodology (CASM) for the British Labour Force Survey and Other OPCS Surveys", *Proceedings of the Bureau of the Census Annual Research Conference*, Washington DC: US Bureau of the Census, pp. 491-500.

Martin J., O'Muircheartaigh C. and Curtice J. (1993), "The Use of CAPI for Attitude Surveys: An Experimental Comparison with Traditional Methods", *Journal of Official Statistics*, **9**, pp. 641-661.

Weeks, M. F. (1992), "Computer-Assisted Survey Information Collection: A Review of CASIC Methods and Their Implications", *Journal of Official Statistics*, **8**, pp. 445-465.

Wensing F. (1996), "Memory Madness" *International BLAISE User Group Newsletter*, Issue number 9, ??

Wensing F. (1999), “Naming for Export: A problem concerning Blaise field names and export” *International BLAISE User Group Newsletter* Issue number 12.