

An Office Management System in Blaise III

*Meryl Henden, Fred Wensing, Ken Smith, Mano Georgopoulos
Australian Bureau of Statistics*

1. Introduction

Collection of data using computer assisted interviewing (CAI) does not stop with the field work. After the data is returned to the office it needs to be downloaded, unpacked, registered, checked and coded before it can be handed over for further processing and output production. The arrival of Maniplus in the tool set for Blaise III has made it possible to develop a comprehensive system to manage all these processes.

This paper describes the Office Management System (OMS) developed in Blaise for the first production application of computer assisted interviewing at the Australian Bureau of Statistics (ABS) and discusses some of the problems that were encountered.

2. Household survey operations in the ABS

The ABS operates a substantial household survey program comprising :

- *Monthly Labour Force Survey* - 30,000 households involving an eight month rotating sample using any-responsible-adult methodology, face-to-face interview in the first month and telephone interview from interviewer's home in subsequent months,
- *Supplementary surveys* - generally conducted as an extension to the Monthly Labour Force Survey, covering a range of short topics,
- *Special supplementary surveys* - 15,000 households, carried out once or twice in a year using personal interview methodology, face-to-face, and covering a range of topics in depth,
- *Quarterly omnibus surveys* - 3,000 households covering a range of user-pays items and generally involving any-responsible-adult methodology,
- *Other user-pays surveys* - varying size, content and methodology depending on demand.

The first two components of the survey program are generally known as the *Monthly Population Survey*.

Most of the surveys in the program are developed and coordinated by central management (located in Canberra) with the field phase being conducted by eight regional offices around Australia. These regional offices are responsible for managing the interviewer work force and carrying out the input processing of survey data (receipt, coding, data entry and editing). Subsequent processing is usually managed by central processing and subject areas.

3. Use of BLAISE in the ABS

Blaise was first used at the ABS in 1994 in a pilot test to assess the viability of computer assisted interviewing in the context of population surveys. That test made use of an early Beta version of Blaise III and concluded that computer assisted interviewing was indeed viable and could deliver quality and timeliness improvements with possibly some cost savings. A technical assessment of the Blaise III software concluded that it was a suitable product, although there were some concerns about integration with other ABS computer facilities given that it is a DOS product.

Blaise software has subsequently been endorsed by ABS management as appropriate for the collection of survey data where that collection involves the use of computers in the field. Given that many other facilities within the ABS make use of Windows and the client-server architecture, the use of Blaise has been restricted to field collection and limited office processing (sufficient to clean up the data and deliver it to other systems).

The test program which was put in place for CAI also included a major test of the *Monthly Population Survey* conducted over a six month period in 1996 which made use of 80 interviewers and a rotating sample of 4,000 households per month. This major test provided significant experience which contributed to the development of systems.

The first production application of CAI at the ABS has been 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, for which the systems described below provided essential facilities.

4. Office Management System Requirements

A new office management system developed to handle CAI was expected to provide, at the very least, the same degree of management and flexibility which is currently experienced with the well established non-CAI systems at the ABS.

Two key features of the old systems which the new developments would have to maintain were :

- ability to manage the work of interviewers from a network of regional offices (ie. the existing organisational arrangements should continue),

- the concept of a batch of work for each interviewer (known as a *workload* and well understood by interviewers).

In addition, the office management system would need to provide 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,
- export a clean data file for further processing.

5. Description of the Office Management System

5.1 Overview

The core of the office management system developed for CAI surveys at the ABS is a series of Blaise facilities which handle three main office activities associated with surveys :

- preparation of workloads for transmission,
- receipt of data from interviewers,
- office processing (clearing of records).

The system also includes Oracle based facilities where selected addresses are allocated to workloads and assigned to interviewers.

The office management facilities are complemented by a system on the interviewers' notebooks (also written in Blaise) which manages the receipt of workloads from the office and the working needs of the interviewer.

Figure 1 shows the Blaise components of the Office Management System and their relationships to the Oracle and notebook processes.

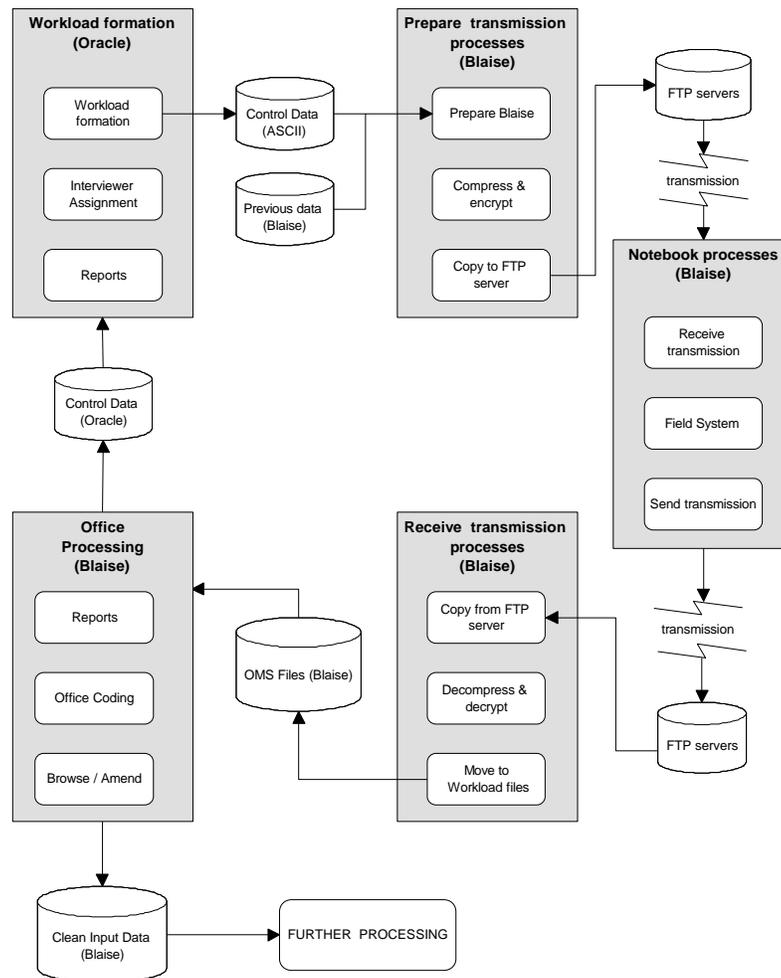


Figure 1. Office Management System showing Blaise components

5.2 Workload formation and interviewer assignment (Oracle)

The underlying feature of this component of the OMS is a series of relational tables stored in Oracle which contain the addresses and other respondent details for the survey sample. For the longitudinal Survey of Employment and Unemployment Patterns these other details included contact information to assist with tracking the respondents over a three year period.

The main reason for developing this component in Oracle (accessed via SQL Windows) is that the functions to be carried out are highly interactive and lend themselves well to that environment and relational tables are a good medium for storage of this kind of data. In addition, Oracle/SQL Windows is the standard client-server environment used at the ABS for most other applications and using this environment gives it a similar look and feel to other applications and position it well to be integrated with other ABS facilities (eg. general sample management facilities) in the future.

Having this component in a client-server environment while other components are in Blaise/DOS does, however, introduce a platform/environment problem. This is discussed further in section 6.

The workload formation facilities take care of the following processes :

- assign selected addresses to workload numbers,
- assign interviewers to workloads,
- transfer records from one workload to another,
- export workloads to ASCII to be prepared in Blaise for transmission to the field,
- update address details for households (as required),
- produce collection control reports including workload and response status summaries.

Once the workloads for a survey have been formed and assigned to interviewers, the Oracle processes produce a series of ASCII files that are used to prime Blaise instruments and prepare transmission files for the interviewer. This is described in the next section.

5.3 Prepare transmission processes (Blaise)

Preparation for transmission involves the following steps :

- prepare Blaise records for each selected addresses in an interviewer's workload using the ASCII data received from Oracle,
- load the Blaise records with any additional data obtained in previous cycles of the survey,
- collate any software required to update the field systems,
- compress and encrypt all the files into a single file (for each survey) to be sent to each interviewer,
- move the transmission files from the LAN to the corresponding FTP server for interviewers to collect,
- check the FTP server to see whether transmission files have been collected.

The transmission preparation processes have been developed using a series of Blaise datamodels, Manipula programs and DOS utilities which are activated from a Maniplus interface. A diagram of the Maniplus menu system to control the transmission preparation process is shown in Figure 2.

Included in this subsystem (and the other Blaise-based subsystems) are collection control files which are maintained as Blaise datamodels. Where possible the system programs have been developed as generic processes which make use of parameters and standard naming conventions to identify each survey and its associated files. To assist with the location of data, metadata and program files, standard directory structures have been set up on both the LAN and the FTP servers.

While the processes may appear simple, some important features were found to be necessary :

- the need know whether software updates were to be included (eg. for a new survey) as additional installation procedures would be needed,
- the need to be aware of whether a transmission was to be an update of a previously transmitted workload (eg. to add a new respondent) in which case a Manipula update procedure would be needed once the transmission was received at the notebook rather than a DOS copy procedure,
- support for one-off transmissions which may be required to fix problems on notebooks in the field (in some cases this may be required for all notebooks in the field),
- the need to keep a record of all preparation events in case problems occur and re-preparation is required,
- the ability to handle multiple transmissions to the same interviewer on the same or a different survey,
- the ability to operate the facilities in both batch and interactive mode.

Some of these features are discussed further in Section 6.

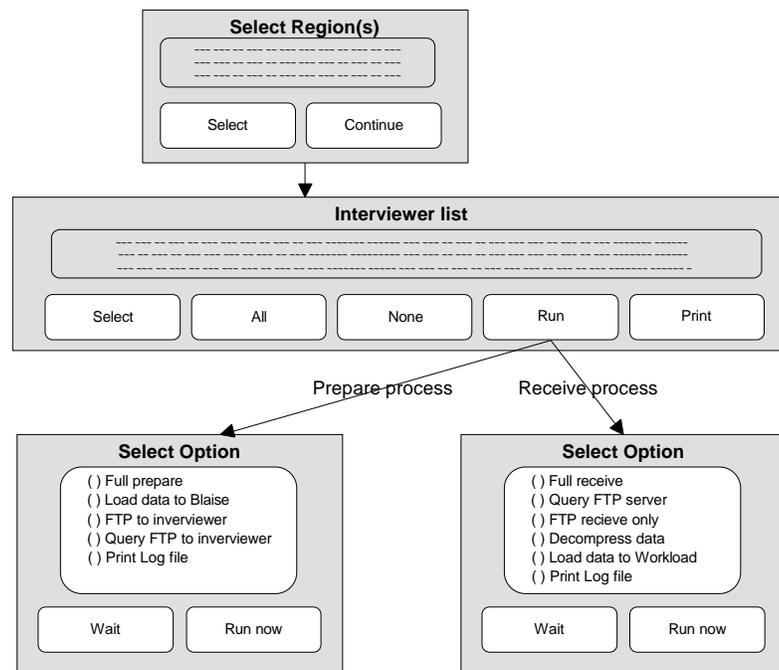


Figure 2. Menus for Preparation and Receive processes

5.4 Receive transmission processes (Blaise)

Receipt of transmissions involves the following steps :

- search the FTP servers for any transmitted data,
- transfer the transmitted data from the FTP servers to the LAN,

- decompress and decrypt the transmitted data for each interviewer for each survey and load the Blaise data into separate workload files.

The receive transmission processes have also been developed using a series of Blaise datamodels, Manipula programs and DOS utilities which are activated from a variation of the same Maniplus interface used for preparation of transmissions (see Figure 2).

Important features that were found to be necessary for receipt of transmissions include :

- the ability to check whether there were any transmissions from an interviewer,
- the ability to handle multiple transmissions from the same interviewer for the same survey,
- the ability to handle transmissions for more than one survey at a time,
- the ability to operate the facilities in both batch and interactive mode.

Some of these features are discussed further in Section 6.

5.5 Transmission (FTP)

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 (see Figure 4). 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" that generates a single-use password with a limited validity time that can be validated by the server.

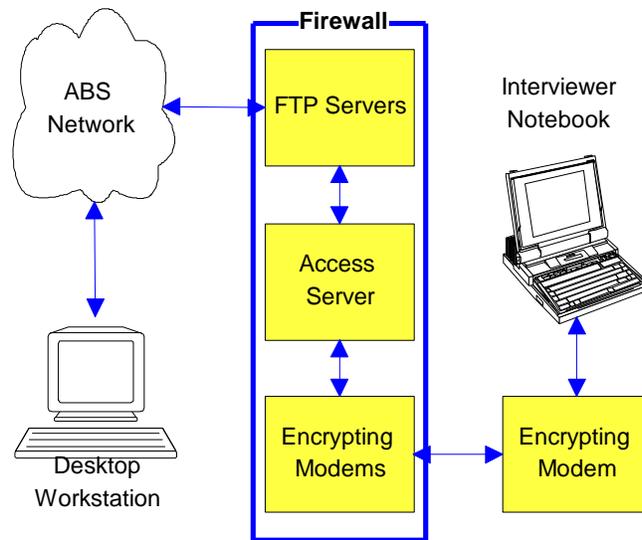


Figure 3. ABS transmission facilities

Each interviewer is generally registered to one of the regional servers and can only access data within his/her own directory on the server. Transmissions are interviewer initiated and make use of a 'pick up' and 'drop off' methodology, in which all external transactions with the FTP server are program controlled from the notebook. Email access is not available to interviewers although this may be considered later.

5.6 Notebook transmission processes (Blaise/DOS)

As mentioned above, transmissions are controlled through programs on the Notebook which handle the following processes :

- prepare response data on the notebook for transmission,
- send response data to the FTP server,
- receive transmission files from the FTP server,
- decompress and decrypt the transmission files,
- execute installation batch files.

The notebook transmission processes have been developed as a series of batch and executable files (containing DOS or FTP commands and Manipula programs) which are called from a Manipulus interface. Interviewers can select whether to send or receive, or both. Facilities are also available through the Manipulus interface to reconfigure the transmission parameters (eg dial-in telephone number) or to switch transmission to floppy disk.

The notebook transmission facilities are completely generic and will act on any or all surveys that are on the notebook. Control of which data files are to be transmitted is managed by including with every survey a preparation batch file (with the same standard name) that contains DOS commands and other instructions to be carried out (to encrypt and compress the data

into a single transmission file). The prepare process simply searches all directories on the notebook for preparation batch files and executes them before transmission. In this way the sending and receiving of data is left in control of the programmers developing the survey since they can include relevant commands in the corresponding batch file that is sent out with the survey instrument.

A significant feature of the transmission processes that have been developed at the ABS is that transmissions to the office will generally contain all records from the interviewer's workload (for a survey) whether interviews are completed or not. This was a deliberate decision, and provides a higher degree of management since transmission files will contain all records for a workload at all times. Transmission of whole workloads also provides a convenient means of data backup (within the notebook as well as from the notebook to the FTP server). There is of course a marginal additional cost associated with longer transmission times.

To handle multiple submissions, a simple naming convention is used for the transmission files. After the respondent data is compressed and encrypted, a single transmission file is produced for each survey and given a name that identifies the survey, followed by a numerical suffix. The suffix number is incremented for successive transmissions using a C program utility. A copy of each transmitted file is kept on the notebook (for backup purposes) and is deleted after the survey is well and truly completed. Recovery of old transmitted files can be made once the notebook is returned to the office. In the office, the latest transmission received for any one interviewer is the file with the highest numbered suffix. Earlier transmissions are kept in the office system and can also be processed if problems arise with the latest transmission file.

Once transmission has been carried out, individual records in the workload which were previously identified as completed by the interviewer (through a "response status" field) are marked as "transmitted" by a Manipula program which is included in the transmission batch file. All records in a workload which are marked as "transmitted" are no longer accessible to be changed by the interviewer, although they can be listed on the screen (for information).

5.7 Office processes (Blaise)

Once the respondent data is successfully transferred from the transmission facilities the first phases of office processing are concerned with :

- following up queries raised by interviewers,
- coding which could not be done in the field,
- clearing up any outstanding data problems identified by the Blaise instrument (eg remaining edit failures or route problems),
- reporting the processing status of workloads for the survey,
- changing the status of individual records,
- combining released workloads into a file for the next stage(s) of processing.

The office processing facilities have been developed using a Maniplus interface, a series of Manipula programs and Blaise datamodels (control files) as well as the main survey datamodel. A diagram of part of the Maniplus menu system to control the processes is shown in Figure 4.

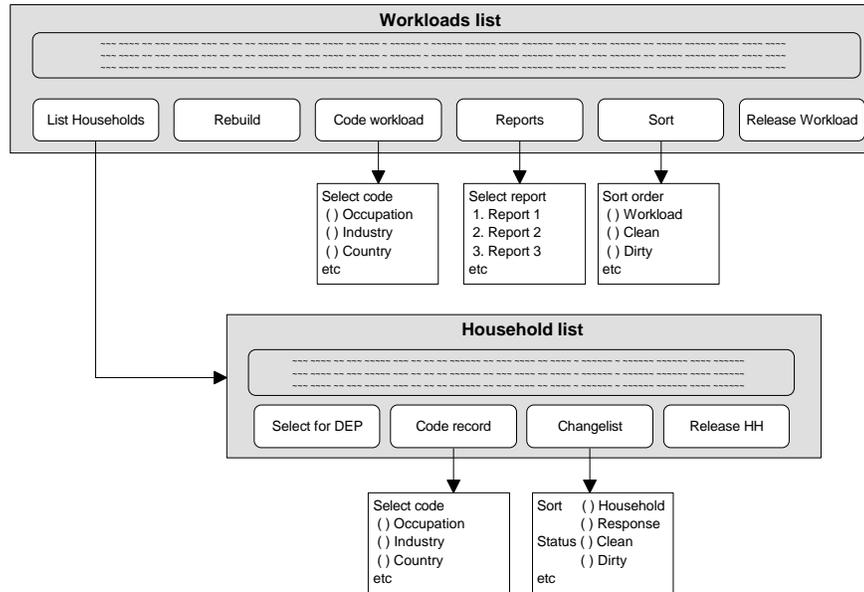


Figure 4. Menus for office processes

A key element of the office processes is knowing the completion status of a survey record. A completed record is one that has been cleared by the interviewer, has had coding carried out (if applicable), is considered clean by Blaise and has also been released by the office staff. Status information is stored either within the survey record or in workload summary control files. The initial status of a record is determined by the system but release of that record requires manual intervention through the Maniplus office interface, and can be done at the record or workload level. Rules are also applied within the system to ensure that there are no inconsistencies in the status elements.

The field system developed for CAI at the ABS provides a facility for interviewers to annotate any respondent record with comments to be followed up in the office. These comments, known as queries, are made available to office staff through the OMS who can enter the respondent record using the Blaise Data Entry Program (DEP) to clear up the identified problem.

Although coding can be carried out in the field using Blaise coding facilities there are various reasons why some coding may be done in the office: the complexity of some classifications (eg occupation) may make it impractical to carry out coding in the field; special systems may already exist in the office to provide for computer assisted coding; or it may be desirable to use existing office coding procedures to maintain comparability with previous surveys. All these reasons were applicable at the time that the OMS was being developed. In addition, coding facilities were not fully functional in the version of Blaise that was available at the time. The OMS, as it has been developed, supports office coding by displaying

relevant record details and allowing a code to be typed directly into the record. It was not possible to link the system to corporate computer assisted coding facilities at the ABS because they make use of SQL Windows. Such links may be possible once Blaise is moved to the Windows platform.

6. Some Technical Issues and Solutions

6.1 Reallocation of respondents from one interviewer to another

One of the more difficult issues to handle has been the reallocation of respondents to other interviewers or for further enumeration by someone in the office. This issue was particularly prevalent in the longitudinal Survey of Employment and Unemployment Patterns where many of the respondents have moved address in the time between the waves of the survey. Some of the situations which have arisen are :

- respondent moves to another interviewer's area,
- respondent moves to a location which is under the control of a different regional office,
- refusal or non response requiring the office or another interviewer to follow up,
- reallocation of a partly completed workload to another interviewer.

Early versions of the OMS dealt with these situations only partially. Although respondents were successfully reallocated, it would result in duplicate respondent records (one from each interviewer), because the record would then be in two workloads. Deletion of the record from the original workload was not practical because the original interviewer may have transmitted a partially completed workload to the office before the reallocation took place. The record would then need to be deleted from all copies of the original workload (both in the office and on the interviewer notebook) as well.

Where duplicate records exist between interviewers (or regions) it is not possible to simply accept the latest data received because the original interviewer may transmit again and after the record was re-allocated to another interviewer. Instead, it was considered sufficient to accept the record which was more complete based on a "response status" code applied by the interviewer. This approach was not fully effective because some response status codes do not indicate which one is more complete.

To overcome this duplicate data problem, the latest version of the OMS contains a Blaise based *control file*. Each time a respondent record is moved to or from any of the processes (ie. Prepare Processes, Notebook Processes, Receive Processes), 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 on the *control file* to indicate the change in control (See Figure 5).

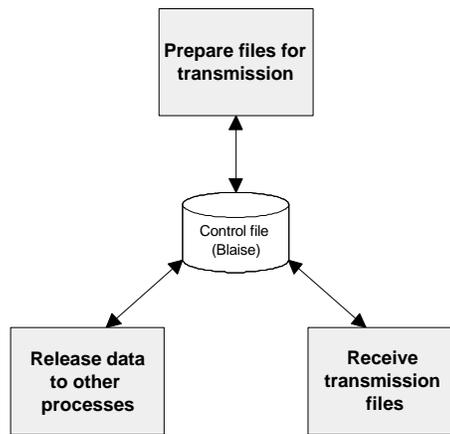


Figure 5. Controlling the movement of records

With this design, a record for any respondent may be located in more than one place but only one is considered to be the active record at any one time. This form of control was only needed in the processes which handle transmissions to and from the field because that is the place where multiple copies can arise. It was not necessary to have a similar arrangement for the pre or post-field processes since there would be only one copy at that time.

An example of the way in which the control file is used (for reallocation) is as follows :

- A record is prepared for a respondent and transmitted to an interviewer in Region 1. The Prepare Transmission process adds a record to the Control file to indicate that interviewer now has control for that record.
- The interviewer discovers when he/she attempts to make contact with the respondent, that he has moved to Region 2. The interviewer marks the response status on the record for that respondent accordingly and it subsequently is transmitted back to the office as part of the interviewer workload.
- The Receive Transmission system processes the transmission and loads the record to a Workload Blaise File. The Control file is updated to reflect that the Blaise OMS now has control of that record.
- An office person reviews the record transmitted from the interviewer and releases the record to be sent back to Oracle. The Blaise OMS Export process moves the record to Oracle and the Control File is updated to reflect that Oracle now has control (actually done by removing the control record for that respondent).
- Someone in the office for Region 2 uses an Oracle process to reallocate the record to an interviewer in their region. The Prepare Transmission process then prepares a transmission for the new interviewer. A new record is created in the Control file to reflect the new interviewer in region 2 now has control.
- Any further transmissions of that respondent record from the first interviewer will be ignored because the control file will indicate that it is no longer with the first interviewer.

While this control process has the potential to increase the time taken to move a record/workload from one interviewer to another, the benefit of knowing exactly where a record is outweighs this disadvantage.

6.2 Running batch jobs in Blaise

The Blaise processes for preparing and receiving transmission were found to be time consuming when run in interactive mode. There were also performance and access problems when other users were accessing the same data files. Arranging for these processes to be run overnight also had problems because :

- being interactive LAN processes, the computer needed to be left logged into the network overnight, causing a potential security problem,
- often these processes needed to start at a much later time (eg after midnight) when all users are out of the system or after interviewers have transmitted their work to the office.

To overcome these problems a WAIT function has been developed in our Maniplus interface for the processes concerned. The function calls a WAIT DOS utility (developed at the ABS) which "locks" the computer until a specified time. After the specified time is reached the computer is freed to run the selected process. At the end of the selected process, WAIT is invoked again to "lock" the computer until the same time the next day. Weekend processes can be activated to start on Sunday by allowing the entry of start date into the interface as well as a start time. The WAIT utility is then invoked in a loop until there is a match with the specified date.

The WAIT utility requires the user to enter a password at the time of submission. If at any time the WAIT process is interrupted the same password is required to allow processing to continue. The only way to interrupt the WAIT process (or the delayed processing) without the password is to reboot the computer (after which normal LAN security prevents unauthorised access).

A problem also arose with delayed processes that produce screen messages which require the user to press the <enter> key to continue but an option to turn off screen messages is now available in the latest releases of Blaise.

6.3 Performance and data volume issues

With the Blaise Office Management System running on the ABS network we have encountered a number of performance problems/issues. In the early version of the system, transmissions that were received from the field were loaded into a single Blaise data file for each region and placed on a Central office server to make them more accessible to survey management staff. With the longitudinal Survey of Employment and Unemployment Patterns this created data files of around 40 Mb each and a national file of approximately 112 Mb.

The problems we encountered with data files of this size were that extraction or updating of workloads during office processing was taking far too long (anywhere from 3 to 20 minutes); large files seem to be more susceptible to "corruption" caused by users aborting processes or from network dropout; some processes required exclusive access to the file (preventing others from using it); and rebuilding the files when they became "corrupt" could take hours to perform.

Two important improvements have now been incorporated in the OMS to deal with these :

- locating the processing files on a regional server;
- maintain a large number of small files during office processing.

Moving the data files to a local server made a significant difference to performance in the regional office because it reduced the delays caused by network traffic. While this improved regional performance there was a consequential reduction in the performance of Central office processes (such as prepare workloads and receive data) but these were not interactive and could be run overnight using the WAIT facility described above.

The most significant improvement to the OMS design has been to maintain a large number of smaller data files during the office processes, rather than a single regional file. By using a separate local file for each workload, it is no longer necessary to "extract" a workload from the main file for local processing, corruption of a workload file no longer affects other workloads and rebuilding such a file can be carried out quickly and does not interfere with other processes. In addition, the prepare and receive processes could now be run during the day without interrupting other processes. Keeping a larger number of smaller files does, however, necessitate a very organised and maintained directory structure on the network and is assisted by the existence of a workload summary file (such as that mentioned in section 5.7) to enable workloads to be located and report on the status of each.

6.4 Management of processes

An important requirement of an office management system is to be able to manage the separate processes, particularly where there is a work force of interviewers that operate remotely from the office. Each interviewer and the each of the processes which supply and receive work to them require separate control procedures. At any one time it must be possible to see which interviewers are expecting work, who has received work, who has returned work and whether the processes which carry out the various send and receive tasks have succeeded or failed.

The OMS contains a control screen in both the prepare and receive processes which provide this functionality and makes them easier to manage (see Figure 6).

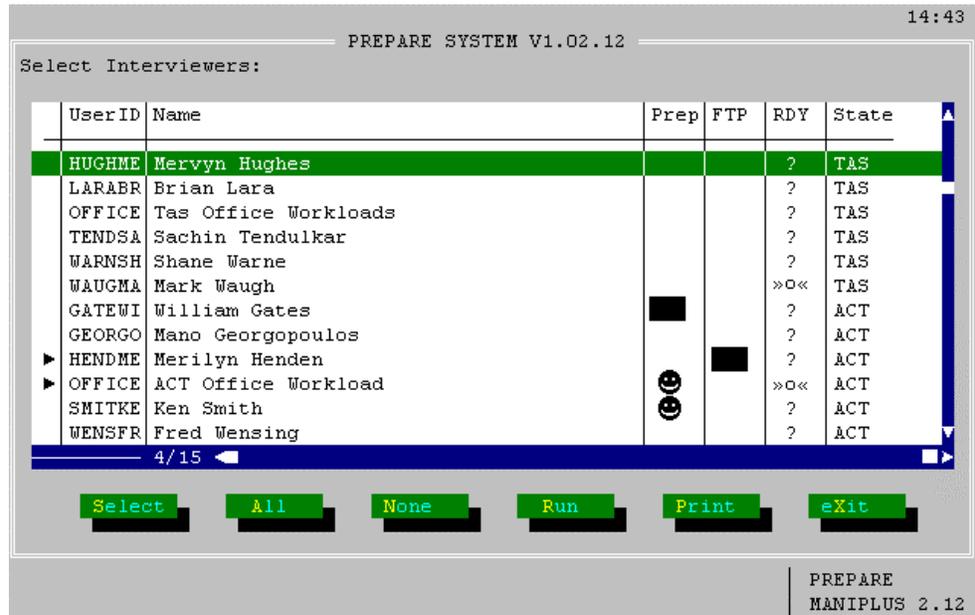


Figure 6. Typical control screen for prepare processes

The control screen presents a list of interviewers and a number of columns which report the status of each process. To make the status easier to interpret, the status report is presented using graphic symbols rather than words or numbers to indicate whether a process is ready to run, has run successfully or failed to run successfully during the current session.

The control screen includes indicators (arrows on the left) that identify when an interviewer is "selected" for a process to be applied, and buttons for selection of "all" or "none". Once a selection of one or more interviewers is made, the required process(es) can be selected using the "run" button which presents a list of processes that can be activated to run immediately or later (using the WAIT facility described in 6.2). The "print" button produces a printed report of the screen details.

The main feature is that all the management information and action options are presented in a single screen. A similar screen been developed for the receive processes with a different set of status columns and underlying processes.

6.5 Clean and dirty records

Another important requirement of an office management system is the ability to determine the "clean" status of individual records. While Blaise provides a useful system-generated status indicator on each record (with codes for Clean, Suspect, Dirty and Not checked), this status can change with even a small adjustment to a record or the questionnaire instrument.

In developing the OMS it became important to understand the way in which the record status indicator is adjusted by Blaise. In that way it can be used to assist (and not hinder) the clearing of records received from the field.

Some important lessons which were learned from the behaviour of the status indicator are :

- as far as possible ensure that the field instrument contains all the office management fields because moving the data between even slightly different datamodels in the office will cause all suppressed edits to become un-suppressed,
- make use of the "check rules" option while running Manipula processes on the data file otherwise all changed records will become "Not checked",
- a "dirty" record may indicate data has been added to fields that are not on the path, so it is important to carefully check the behaviour of all processes.

Since the development of the office management system described in this paper an additional option has been added to Blaise to enable the software to "force" the rules in a Manipula process and remove data that is not on a valid path through the instrument. This feature should assist office processes in the future.

7. Conclusion

The system described in this paper highlights some of the logistic, technical and data handling issues that have to be overcome in office management of computer assisted interviewing. Maniplus has provided the means to develop a series of integrated functional procedures to deal with these issues, which can handle the response data while it is a Blaise format, and present the office staff with an easy-to-use interface to carry out their functions.

8. Acknowledgements

We would like to acknowledge the contribution of other staff from the Household Surveys Facilities development team and the Population Survey Operations Section at the ABS who have assisted in the design, development and testing of the systems described in this paper. Special reference must be made to Julie-Anne McDonald and Sharon Baker who were responsible for developing the early versions.