Journal of Official Statistics
Vol. 9. No. 4, 1993, pp. 765–782

Management of Computer Assisted Interviews in the Netherlands

Lon P.M.B. Hofman and Wouter J. Keller[1]

Abstract: The Netherlands Central Bureau of Statistics (CBS) has developed and used Computer Assisted Interview systems for some time now. One of the results is the Blaise System. In order to be able to implement such systems successful good case management is essential. The CBS developed case management systems both for personal interviewing in the field and centralized telephone interviewing at the bureau. Both systems will be presented in this paper.

Key words: CAPI; CATI; case management.

1. Introduction

This paper describes the management of Computer Assisted Interview systems (CAI systems) at the Netherlands Central Bureau of Statistics (CBS). The CBS distinguishes three main types of CAI systems: Computer Assisted Personal Interviewing systems (CAPI systems), Computer Assisted Telephone Interviewing systems (CATI systems) and Computer Assisted Data Input systems (CADI systems). This paper will focus mainly on the management of CAPI systems used by interviewers in the field.

A number of national statistical agencies already use techniques like CAPI and CATI; some are planning to use them or are evaluating them seriously (see Statistics Sweden 1990; Braslins, Coutts, Jamieson, and Williams 1992; Nicholls and Matchett 1992). Agencies that do not use CAPI or CATI are in the minority. With the use of CAPI on laptops in the field or dozens of interviewers conducting concurrent CATI, good CAPI and CATI case management systems are essential. Examples of development or use of CAPI case management systems can be found in various places. An operative system can be found, for instance, at the Office of Population Censuses and Surveys (OPCS) in London. They have a CAPI management system for their continuing Labour Force Survey. This system uses diskettes to send address information to the interviewers and uses the telephone to obtain completed interviews. More information can be found in Manners (1991). Information about other systems can be found, for instance, in Kerepesi (1992) and Kuussela and Merisalo (1992).

In Section 2 we will give a short overview of the development of CAPI systems in the field and its management at the CBS. Section 3 then gives a description of the

[1] Netherlands Central Bureau of Statistics, Statistical Informatics Unit, Automation Department, P.O. Box 959, 2270 AZ Voorburg, The Netherlands.

recently developed CAPI case management system which has been in use at the CBS since 1991. To conclude the paper, Section 4 briefly describes the management of CATI surveys at the CBS.

2. Overview of CAPI Developments at the CBS

The Dutch experience with CAPI started in 1983. The first experiment was conducted using a Price Survey. In this survey, prices of commodities were observed in order to calculate the monthly consumer index numbers. The interviewers visited shops with the computer, recorded prices, reported changes in commodities, located and recorded new shops whenever a specified shop closed down, and answered special questions in supplemental questionnaires. Although there were some problems with the equipment, the experiment was a success for various reasons. First, it was possible to formulate the requirements a laptop computer should satisfy. Second, the experiment showed that respondents (shopkeepers) had no objection to computer assisted interviewing. And third, despite inconveniences experienced with the equipment, the interviewers developed a positive attitude towards this way of working. And training required only a couple of hours. After the success of the Price Survey experiment, another experiment was carried out to test the reactions of the general public when confronted with a laptop computer. This time a regular questionnaire was programmed with questions borrowed from the Consumer Sentiments Survey. A total of 173 interviews were conducted with the laptop computers and 167 interviews were carried out in the traditional way with paper and pencil. A comparison of the results showed that the use of laptop computers did not increase the non-response. There was no indication of feared psychological ("Big Brother") effects. The interviewers quickly learned to handle the equipment after only one hour of training. They reacted enthusiastically to the new technology.

Because of the success of these experiments, in 1987 the CBS started using CAPI in the continuing Labour Force Survey (LFS) and as a result was one of the first bureaus using CAPI on a large scale. An interview program called QUEST was used. About three hundred interviewers conducted the survey using a (currently outdated) Epson CP/M handheld computer. (MS-DOS handheld computers were still rare and expensive in 1986!) Each month the interviewers, equipped with the laptops, visited 12,000 addresses. After a day of interviewing, they returned home and connected their computers to the electrical outlet to recharge the batteries. The laptop computer is also connected to a telephone and modem. At night the collected data are automatically transmitted (with a speed of 300 Baud) to the statistical office. The next morning the interviewer finds a recharged machine with a clean work space, ready for new interviews.

LFS data processing at the CBS (for instance coding of occupation) is time critical. By contract, the data have to be delivered by the field work department to the statistical department within five working days after the end of the survey period. For this reason, there has to be a continuous flow of interview data from the field to the CBS. This type of transfer can be achieved thanks to data communication by telephone.

The LFS is the only survey the CBS conducted this way. The QUEST program was especially developed to function on a "small" CP/M handheld computer: a small screen, little work space, little internal mem-

ory. Because of this, the specification language is very user unfriendly and not very powerful, which makes the development of a questionnaire time-consuming. Loading of a questionnaire into the Epson laptop is not easy. It can only be done at the bureau, one at the time. Because of this, changes to the LFS questionnaire, which took place about once a year, had to be carefully planned. In this sense the whole system was not very flexible. Although the QUEST interview program could in principle also have been used for other surveys the limitations and deficiencies prevented this.

While the LFS was still under development, the CBS began work on the Blaise System (CBS 1987; Bethlehem and Keller 1991). In just a few years the Blaise System became an important tool for survey processing at the CBS and other agencies. In 1989 the CBS started the first Blaise CAPI household survey using a Toshiba T1000 MS-DOS laptop computer. All data exchange with the interviewer takes place with 3.5 inch diskettes. This was necessary because the old LFS telephone data communication system could not be used and the CBS did not want to invest in a new system at that time. Gaining experience with the use of diskettes was thought preferable. A final decision on using diskettes or telephone was postponed.

The procedure for conducting a survey with Blaise is simple: A diskette with the interview program is sent at the beginning of a survey period together with a list (on paper) of addresses to visit. During the survey the interview data are stored on diskette. At the end of the survey period the diskette is returned to the CBS. Although the procedure is simple, conducting a survey this way proved to be labour intensive: for instance, at the CBS many steps are necessary before all interview data of one period are assembled for further processing. One drawback was that the interview data were not backed-up (duplicated) in the field. This proved to be a precarious way of conducting a survey as any mishap could lead to all of the interview data for an interviewer for an entire survey period being lost.

At the beginning of an interview the interviewer enters the identification number for the interview and a number of purely administrative questions, like the interviewer identification and the municipality code. The answers to these questions can be found on the list of addresses. Because the identification number has to be entered by the interviewer, there is no guarantee that, when all the interviews have been gathered from all the interviewers, every questionnaire has a unique identification. This creates difficulties in the processing of the survey results. It would be advantageous if the address information were present in the laptop. Then, no errors would be made with the identification number and the administrative questions could be entered automatically.

At present a number of small surveys are conducted by the CBS with Blaise on the same laptop, mainly on a continuing basis. Each survey has its own diskette. It gets difficult for the interviewers to keep track of the various diskettes as the number of surveys increases. Also, the interviewer can easily lose sight of how extensive his/her workload actually is. The need grows for a management system on the laptop.

In the beginning of 1990 it was decided to replace the old LFS system by a new one. There were two main reasons for this decision:

- Spare parts for the Epson handheld computer were difficult to obtain. It would not have been possible to keep the old LFS system up and running using the Epson computers after January 1992.

- The questionnaire had to be redesigned. Blaise allows the construction of a questionnaire which is far better and easier to maintain. The LFS wanted to be able to enjoy the blessings of Blaise.

In April 1990 the preparations started for the conversion of the LFS system from QUEST/Epson to Blaise/Toshiba. A lot was learned from the use of the old LFS system and from Blaise CAPI surveys conducted with the Toshiba and diskettes. One such lesson was that it would be advantageous to have a system in which every address in the survey can be "traced." In such a system, it would be possible to answer questions like: Has the address been sent to the interviewer? Was it received back? Was it a non-response? and so on. It was thought possible to design and build a CAPI management system based on that "traceability" principle. For the new system, a solution had to be found for receiving the interview data in time from the interviewers on a regular basis since the processing of the LFS is time-critical. More about the design and development of the new CAPI management system can be found in Section 3.

3. The CAPI Management System

In April 1990 the design of the new CAPI management system was initiated with a preliminary study. There were several conditions which had to be satisfied:

- All addresses to be visited are stored in the automated Interview Administration System (IAS). This makes it possible to give each interviewer addresses to visit in machine readable form (either on diskette or by telephone). The CBS interviewers always conduct two forms of accounting for a household survey: sample accounting (e.g., how many households live at an address and how many were selected) and the visit accounting (e.g., when the address was visited and what the result was). This should be done using the computer. In this way, a lot of paperwork can be avoided and it is also possible to get a more "balanced" accounting of the addresses by means of the IAS (using the principle "what goes out must come in ..."). Every address is traceable at any time. Answers about the status of each address must be given by the IAS.
- The system must be capable of handling several surveys at the same time. To conduct a survey, a Blaise CAPI interview program is required. It should be possible for different surveys to use the same interview program. For example, in a rotating panel survey each panel wave is a different survey; different panel waves can be held at the same time, but they all use the same questionnaire.
- Data communication can be carried out with diskettes or by telephone. Data should be stored in a safe way (for instance, encrypted and compressed). Interview data should be sent on a regular basis to the bureau to ensure delivery to the statistical department in time.
- The system should not exceed the limited disk space available in the laptop.

The system we had in mind before the preliminary study started consisted of two modules. One is called the central management system and is located at the CBS. The central management system has to manage the data stream from the IAS at the bureau to the interviewers and vice versa. The

second module is called the decentralized management system and is located in the laptop of each interviewer. The decentralized system supports the interviewers in their work with laptop computers. Because of the hardware configuration, the decentralized management system is diskette-based.

3.1. The preliminary study

The system handles two different types of data: interviewer–independent data and interviewer–dependent data. An example of interviewer–independent data is the interview program. Every interviewer receives the same program. Most of the CBS CAPI surveys are continuing. The interviewing program is normally used for a longer period, for instance one year, and it suffices to send the program only once and send updates as they are required. Such data can be sent easily to an interviewer on a diskette. Interviewer–independent data stored on a diskette can be reproduced fast, very cheaply, and without errors in a disk copy machine. The amount of data is large, possibly up to 500 K for one interview program and must be sent in a short period of time (a high data peak). The telephone is not a good transfer mode for this kind of data. Even with a transmission speed of 2,400 Baud it takes a long time and the sheer number of interviewers (400) makes it a difficult job to execute by telephone. If the data were sent by telephone, an empty diskette would also have to be sent to the interviewer to store the data on. So why not write the data on the diskette from the very beginning? We concluded that diskettes are very useful if the amount of data to be sent is large, but the frequency of sending is low.

Dealing with interviewer–dependent data is a different story. The amount of address data for one interviewer is small (a few kilobytes) and it has to be sent frequently. The amount of accounting and interview data is also small when data are compressed. Several alternatives were studied for the communication of this type of data:

- All communication by means of diskettes: Both address information to the interviewer, accounting, and returned interview data are sent on diskette.
- A mixture of diskette and telephone: Address information is sent to the interviewer on diskette. The CBS receives the accounting data and the interview data over the telephone line.
- All data communication by means of telephone: Address information is sent to the interviewer via telephone lines. The CBS receives the accounting data and the interview data over the lines. This is called *electronic interviewer mail.*

Storing interviewer–dependent data at the CBS on a diskette and ensuring that the diskette is sent by mail to the correct interviewer often proves problematic. It is impossible to determine the contents of a diskette simply by looking at it. The easiest way to see a diskette's contents is to put it in a disk drive and display the contents on a screen. On the other hand, this method is not very handy when you mail several hundred diskettes each week. Barcode labels are another way of identifying diskette contents. With a Copydisk machine it is possible to read a barcoded label on the disk. When data are stored on the disk, a registration can be made using the link barcode-interviewer. The barcode can be read when the diskette is put in an envelope. This produces an address label which is then attached to the envelope. The mailing of diskettes in this way (or another) is labour intensive (and a boring job for the person who has to do it).

Receiving interview data on a regular basis (for instance for the LFS at least once a week during the monthly survey period) is difficult when diskettes are used. Special provisions have to be taken for backing up the data in case a diskette goes astray or becomes unreadable. What we call a *receive confirmation scheme* was worked out to ensure the receipt of survey data. When a diskette with survey data comes in, *a release back-up instruction* is written on the diskette that is sent back to the interviewer. If the interviewer (the laptop) does not receive this instruction the next time he/she mails a diskette to the CBS, the back-up data are automatically included on the diskette. The receipt of interview data in this way is difficult to manage.

From the management point of view, it was concluded that the best way to send and receive interviewer-dependent data is via telephone. Electronic interviewer mail is easier to manage than physical interviewer mail!

The following question remained: If we chose electronic mail, how should it be done? Use a commercial electronic mailing facility or build our own? The CBS has strict security regulations to guarantee optimal privacy of respondents. We concluded that the interview data could be better protected by using an electronic mailing facility developed in-house. Ignoring the development cost of such a facility, this would also be cheaper than using a commercial facility.

The old LFS data communication system consists of nine personal computers, each with its own modem. Each PC serves a fixed group of interviewers. Each interviewer has his/her own time slice during which communication is possible every night. During this time slice the laptop can call its PC at the bureau and transfer data if necessary. Because the internal clocks of the laptops all have a slightly different time and because the communication takes place at 300 Baud, the time slices were made rather wide (ten minutes). Due to strict timing of the calls for each interviewer (call it inflexibility) the system has an overwhelming overcapacity. The effective use is only 3%.

A communication speed of 300 Baud required only four modems and could do the job if the idea of time slices and fixed groups of interviewers for each PC was abandoned. An in-house technical expert concluded that he could build a new system based on only one PC running under MS-DOS which serves up to eight modems simultaneously in full duplex mode. This new system will be able to handle 300, 1,200 and 2,400 Baud modems at the same time.

From the management point of view (and the technical point of view) electronic mail was preferred. But what about the costs? Also from this point of view the telephone proved the best alternative. Although the investments in hardware for the interviewers and the CBS are high it takes about three years before the break even point with physical mail is reached. If a depreciation period longer than three years is used it becomes even cheaper to use electronic mail. The variable costs of electronic mail are less than the variable costs of physical mail.

All interviewers who participate in the old LFS are equipped with a 300 Baud external model. Although a speed of 2,400 Baud is preferred, the old 300 Baud modems can also be used in the new system without causing capacity problems. This way the initial investments in the new data communication system can be lower. Only the interviewers who do not participate in the LFS have to be equipped with a new modem. Nevertheless gradually all

the old modems will be replaced by new ones.

Based on the findings of the preliminary study, a new system can be designed and built in which all interviewer–dependent data will be exchanged using the telephone and all interviewer–independent data will be sent on diskette.

3.2. A description of the new system

In May 1990 the design and building of the two management systems and their various interfaces to other systems was started. The decentralized system received the name LIPS. LIPS denotes Laptop Information system for Personal Surveys. LIPS is designed to support the interviewers in their work with laptop computers. It takes care of various aspects of the field work. The central management system, the counterpart of the decentral system, is called SPIL, which denotes System to Perform data Interchange with Laptops (or Lips). SPIL is the central system which takes care of managing the data flow between the interviewers (LIPS), the Interview Administration System (IAS), and the statistical departments.

Figure 1 depicts the various systems and their data flows.

All samples of addresses for the various CAPI surveys are stored in the IAS. The system is implemented as an Oracle data base on a mini-computer. This mini-computer acts as an Oracle server and is accessible by Oracle Clients on the Local Area Network of the CBS. The IAS allocates the addresses to the different interviewers taking into account their maximum workload for all surveys in a given period. The workload for a certain survey depends on the average interview time and the number of allocated addresses. The workload is calculated in hours and is matched with the total number of hours the interviewer is willing to work for the CBS in a given period. The automated allocation process also uses the postal area code of the addresses to minimize travelling costs. The resulting allocation is then corrected manually. On request, the IAS sends these allocated addresses to SPIL. SPIL ensures that the addresses will be sent to the right interviewers. The entire procedure is fully automated.

SPIL splits up the incoming information (from the interviewers to the CBS) into two streams. One stream leads back to

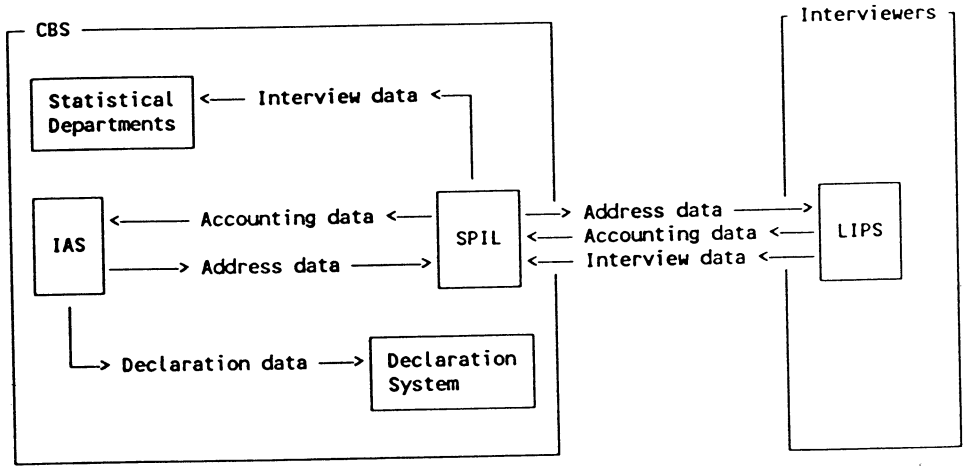

Fig. 1. The CAPI management system

IAS and consists of sample and accounting information, and addresses that have not been visited yet (*return addresses*). In the future IAS will send the accounting information to an automated payroll system, which will calculate the interviewers' wages. The unfinished addresses may be sent back to SPIL for redistribution. The other stream consists of the survey data that have been collected. These data will be sent to a coding system if necessary. From the coding system the data will be sent to the statistical departments for further processing and evaluation.

The IAS performs a key role in delivering management information regarding the different surveys. For instance, each day for each survey a very important indicator is computed: the percentage of the survey that has been completed. Because the IAS stores all sample and visit account data collected with LIPS, it can also answer many other questions about a survey, for instance, number of refusals. It is a potentially rich source of information for all kinds of studies regarding the field work.

3.2.1. Conducting a survey with LIPS

LIPS is completely menu-driven. By using the cursor keys, the interviewer can scroll through the menus and lists. By pressing the enter key, the interviewer selects the item presently indicated. At various places the interviewer can perform an action by just pressing one of the keys indicated on the bottom line of the screen.

LIPS can handle data for up to 15 surveys simultaneously. A survey on the laptop consists of a survey definition and a list of addresses to be visited. In the survey definition several characteristics of the survey are stored, for instance, a short description, the name of the Blaise questionnaire to be used for interviewing, the period in which the addresses of the survey have to be dealt with. All the available surveys on the laptop are presented to the interviewer in a menu. In Figure 2 an example of the survey selection menu is found. Each survey in the menu is identified by the short description and the survey period, for instance *Labour Force Survey, 01-11-1992-31-11-1992.* While performing actions on the address level the survey is always identified in that manner on the screen. The addresses of a survey are presented to the interviewer as a scrollable list. An example of such a list is found in Figure 3. The list can be sorted according to a number of parameters if required, e.g., on postal code, alphabetically on street name, or on status value. Each address in the list has an attribute that indicates the current status of the address in the survey. This status attribute is very important. LIPS uses the status value to determine the current point in the cycle of the address in the survey. The status value is used by LIPS in many places, for instance to determine whether a certain action can or should be performed with the address. Examples of status values are "*approached*", "*completed*" and "*sent to CBS.*" The status value "*approached*" indicates that the interviewer provided information to LIPS about the address. If the status value is "*completed,*" LIPS will send all the available information provided by the interviewer to the CBS when data communication is performed. If the status value is "sent to CBS" the interviewer cannot select the address anymore (although the address still appears in the list).

LIPS can present a list of all the addresses for all surveys available on the laptop. This can be handy for determining the total workload and an efficient route to visit the addresses. The interviewer is presented this overview when he/she selects the last option in the survey selection menu, "*addresses of all available surveys.*"

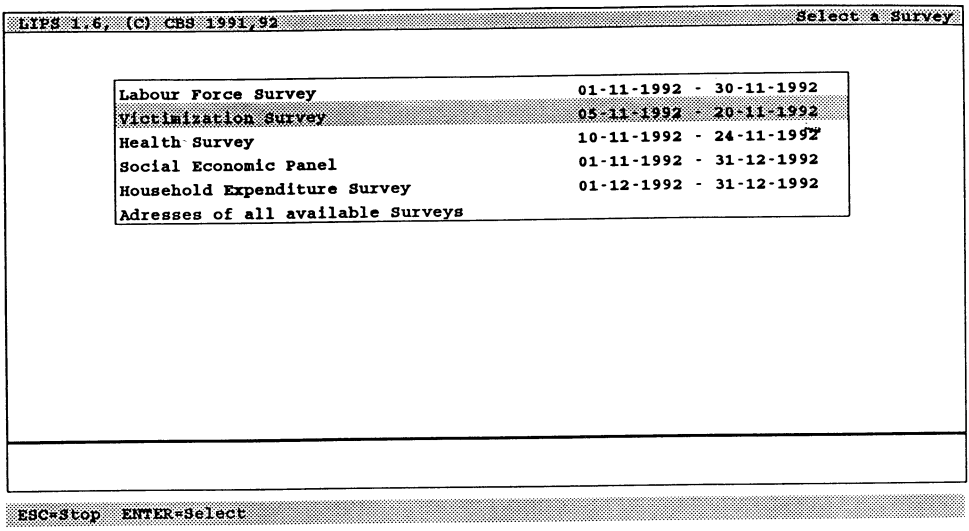

Fig. 2. The survey menu

To carry out an interview the interviewer follows a number of steps:

1. Selects the appropriate survey from the survey selection menu (see Figure 2). Next a list of addresses for that survey is presented to the interviewer (see Figure 3).
2. Selects the address to be visited from the list presented.
3. Fills in some information about the address, like the number of households living there (this information is called the *sample account*). For each address, up to nine households can be interviewed.
4. Selects a household and starts the interview program. First, LIPS checks whether the designated CAPI–program is available on the laptop. If so,

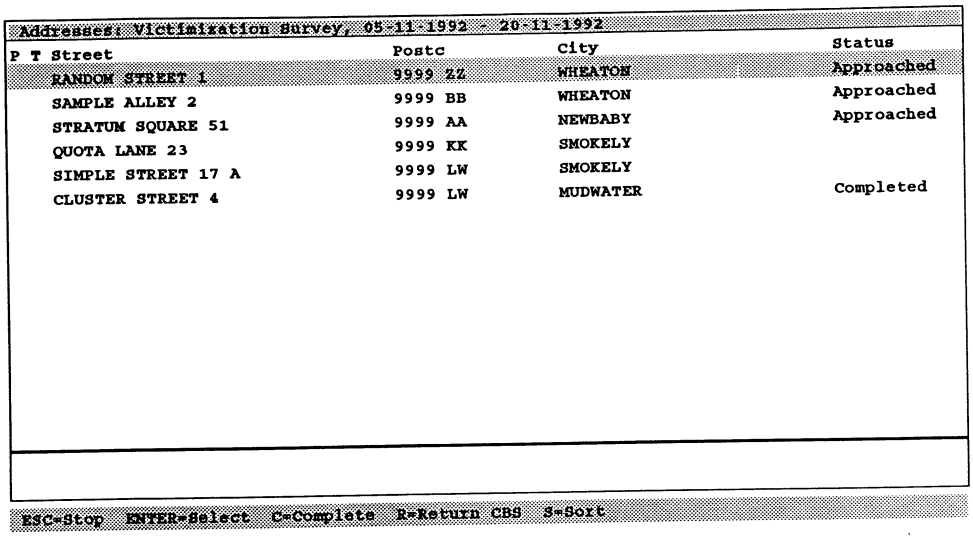

Fig. 3. List of addresses for a survey

the program is started with the correct parameters (for instance, the key value). If the CAPI–program is not available, LIPS will ask for the disk with the CAPI-program (and copy the program from diskette to the laptop if the laptop has enough disk space to store the CAPI–programs). Finally, LIPS starts the CAPI–program. After the interview has been concluded LIPS stores the interview and the main menu of LIPS reappears on the screen.

The interviewer fills in visit account information for each household, like the time of the visit, the result of the visit, etc. See Figure 4 for an example of the visit account screen. One of the principles during the design of LIPS was that the program should be, whenever possible, very flexible: It should be an aid to the interviewer, not a strait-jacket. For instance, the filling in of the accounting information, until now done with pencil on very flexible and tolerant paper forms, should not hinder the interviewer in executing his/her primary task with LIPS: interviewing. In this way, LIPS gives the interviewer the freedom to fill in the accounting information whenever it suits him/her best. He/she can do it directly after the visit or in the evening at home. LIPS does not offer an agenda function or schedule.

The interviewer indicates that an address is ready to return to the CBS by pressing a function key in the address list screen. LIPS then checks whether all accounting information has been entered correctly (the same principle as in Blaise CADI: check on request). If all the accounting information is correct the status value is changed to "*completed.*" If the accounting information is not correct, an error message is issued. This error message can be overruled by the interviewer, but if he/she does, the information is marked *dirty* and receives special attention at the CBS. Such an address receives from LIPS the status value "*return, not completed.*" Furthermore, the interviewer can return an address to the CBS, for instance, if the workload is too large. In this case LIPS changes the status value to "*return to CBS.*" All addresses

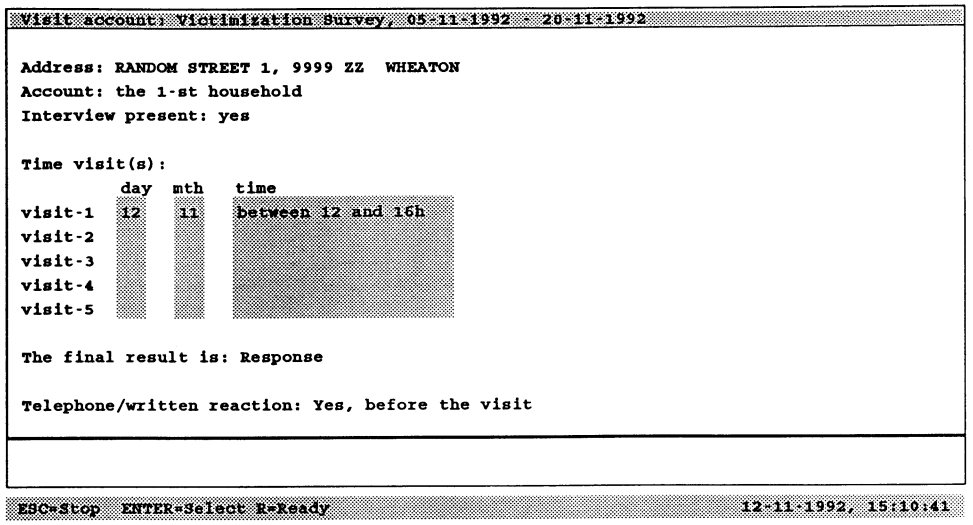

Fig. 4. The visit account screen

with one of the mentioned status values are ready to be sent back to the CBS. LIPS sees to it that they will be sent to the CBS the next time the interviewer performs data communication.

On request LIPS shows information on the surveys the interviewer has currently available on the laptop, like the name of the survey diskette, the survey period and the number of addresses in the survey.

The LIPS system can also be used for panel surveys. Panel data are stored in the IAS. They can be sent together with the allocated addresses to SPIL, which directs the panel data to the correct interviewer. LIPS ensures that the panel data are available during the interview as external data for Blaise. These data can be used for routing or for performing desired edits.

3.2.2. Data communication

All interviewer-dependent data communication between the CBS and the interviewer takes place via modem. All interviewers are supplied with an external modem. An external modem is not only cheaper than an internal modem, but it has the advantage of not contributing to the weight the interviewer has to carry around.

When interviewers want to send data to the CBS or receive data from the CBS they select the option *Communication* in the management menu in LIPS. First, the interviewer enters his/her personal password which is stored at the CBS. Second, he/she has to indicate whether he/she wants to perform data communication at once or during the night. If night communication is preferred (in The Netherlands it is cheaper to phone during the night) LIPS selects a random time when the CBS is to be called. LIPS always checks whether the modem is plugged in correctly.

In the main the following steps are performed by LIPS:

1. LIPS prepares an archive file with all the data that are ready to be sent to the CBS. The data to be sent back to the CBS consist of the interviews, accounting information, and return addresses. LIPS also makes a back-up of this file, which remains available for about two months.
2. The program then dials the telephone number of the CBS. In the event the initial call does not connect (something goes wrong) the program makes three attempts at calling CBS. During the night this process is repeated twice if necessary, each time after 20 minutes.
3. When the link has been successfully established, LIPS checks whether the interviewer is known to the system and whether the interviewer password stored at the CBS is the same as the interviewer password entered in LIPS. If so LIPS sends the data to the CBS or receives address data from the CBS, or both. There is also the alternative that LIPS receives a message for the interviewer, for instance: "please telephone your coordinator." During data communication the time of the laptop is synchronized with the time at the CBS.
4. LIPS takes care of the data received, if any, and presents an overview of the results of the session, like the number of new addresses or the reason why the communication action was not successful, for instance: "password not valid."

After all data from a survey have been sent to the CBS, LIPS removes all remnants of the completed survey and prepares the laptop to receive a new survey. A survey is completed when all addresses of a survey have the status "*sent.*" LIPS removes these surveys automatically when the survey

period expires. The next time the interviewer uses LIPS "sent" surveys no longer appear on the survey menu.

The interviewers still receive some printed information from the bureau. For instance, interviewers send letters soliciting cooperation from potential survey participants. Interviewers receive a paper list of addresses. When an interviewer receives this list, other survey materials are available from the bureau via data communication. In this sense the arrival of the list triggers the other data collection activities.

At present only 2,400 Baud modems are used for data communication with the CBS. The data are sent over the line using an in-house developed protocol, so no one has to worry about data integrity. At the CBS a dedicated PC with four external modems (which are handled simultaneously in full duplex mode under MS-DOS) receives the data from and sends data to the interviewers. Each laptop uses the same phone number for contacting the data communication facility at the CBS. The computerized telephone exchange directs incoming calls to a free modem. The interviewers can send data to the CBS-PC for the greater part of the day (typically 18 to 20 hours per day). At the CBS the data received are stored on a hard disk. When the communication lines are open, the PC is not connected to the CBS local area network. Once every day the PC is connected to the network (the telephone data communication lines are then disabled) and all data that have been received during the previous day are then transferred from the hard disk to the file server(s) on the network to be processed by SPIL.

3.2.3. LIPS file management

For each survey, several files are stored on the laptop: a file containing the addresses, a file containing accounting information, and a file with the interview data. A back-up of this information is stored on a diskette. The CAPI programs are stored on separate diskettes and, if enough disk space is available, on the laptop as well. All data are stored by LIPS in scrambled format: either in password-protected archives (which gives a very scrambled result if you do not know the password) or by using our own scrambling if no archives are used. If ever a diskette with data goes astray, it is very hard to make sense of the files stored on it.

In order to keep LIPS as general as possible it was designed so that no knowledge of the Blaise structures of questionnaires or data is required. Only the key for the questionnaire and the menu option to start Blaise is used by LIPS. For LIPS a special Blaise data format has been developed: Each questionnaire is stored in a separate, compressed file. All files for a survey are stored together in one archive file with a password. Just before data communication LIPS removes the questionnaire files from the archive file and stores them in the file that will be sent to the CBS with the accounting information. This scheme of archiving and de-archiving special Blaise format files has several advantages. The two most important are:

1. The data are stored in compressed (and password-protected) archive files. Thus the total amount of required disk space is significantly reduced and illicit access to the data is very difficult.
2. When questionnaires are sent back to the CBS, the related information is deleted from the laptop. As a result of this, the size of these special Blaise archive files remaining on the laptop decreases. The size of normal Blaise data files does not decrease if an interview is removed.

The LIPS/Blaise-data files are merged and converted to the normal Blaise format at the CBS by the form manager, a tool in the Blaise system.

3.2.4. Training and support

Training and support are two important subjects. They are essential for the good functioning of the complete system.

It is obvious that interviewers have to be trained in using a computer. Good training is crucial to the success of the computer in the field. On the other hand, such training does not have to be too extensive if a system is user friendly and takes care of the nitty-gritty details by itself. LIPS is such a system.

The interviewers get a basic five-day training course when they start working for the bureau. They are trained in the traditional paper-and-pencil interviewing and in accounting for their work. For accounting a special *sample* and *visit account form* (SVA form) is used, which is the same for all household surveys at the CBS. During the basic training computers are not used.

Only interviewers having received this basic training may participate in the training for LIPS. This training takes an additional three days. The following topics are discussed:

- A short introduction to operating the laptop.
- A description of LIPS in the survey processing (something like Figure 1).
- The principles of internal and external memory of the laptop. The different sorts of diskettes that play a role in LIPS are explained, back-up procedures (automatic and manual), etc.
- Rules for data communication and recharging the batteries.
- The different menus of LIPS, the meaning of menu options.
- Accounting with LIPS: During the design of LIPS much consideration was devoted to analyzing the way the SVA form is used by the interviewers. Accounting with LIPS bears a strong resemblance to accounting with the SVA form. This makes training and operating less complicated.
- Training facilities in LIPS: For training purposes, LIPS offers a practice survey that is in all ways identical to a regular survey except that the survey is never deleted from the system when all addresses receive the "*sent*" status. It contains a simple Blaise questionnaire and can be used over and over again without any special attention from the CBS.

During a separate two-day training session the interviewers learn to use Blaise. This is always done in combination with a training for a survey. A topic dealt with in this training is the treatment of *edits* in the Blaise questionnaire. In general, much effort is devoted to the design of the Blaise CAPI questionnaire and to proper messages for the different edits. It is important that each edit results in a clear, comprehensible message as the system is designed to keep the number of edits to a minimum. Training for a survey without the Blaise component lasts one day.

LIPS allows an interviewer to train on an arbitrary Blaise questionnaire without having to select an address. This is especially valuable when a new survey is introduced and the interviewer wants to get familiar with the Blaise questionnaire. The interviewer just inserts a valid "survey" diskette in the disk drive and selects the option *Training* in LIPS's main menu. LIPS takes care of the rest.

Beside the extensive training the interviewers receive, there are always ample

staff from the interview department to provide support. For instance, two staff give technical support. They deal with hardware failures, are responsible for maintenance, etc. One person deals with all data communication problems. Five LIPS trainers help interviewers who have experienced problems operating LIPS. These persons are also responsible for monitoring the return of addresses and the incomplete account information. An interviewer must have a reason for returning addresses or sending incomplete accounts to the bureau and preferably the interviewer should ask permission to do so (LIPS does not enforce this).

3.3. *The introduction of the new system*

In the second half of 1990, two small field tests were carried out with the new system. The first one in September 1990 was to find out whether LIPS was adequately interviewer-friendly and to evaluate the LIPS training program. A group of 10 interviewers was trained; they used LIPS for two of the continuing CAPI surveys during one month. Data communication was not used. The results of this first test were very promising. After some alterations to the system a second test was carried out in November. About 30 interviewers participated. In this test the complete system was used. This test was also fairly successful.

We planned to introduce the new system on a large scale in April 1991. Before this date, 300 interviewers had to be trained on the new system. A maximum of 12 interviewers were admitted to each three-day course. So at least 25 courses had to be given. The LIPS training courses started in mid-January 1991, and by the end of March all 300 interviewers had been trained.

In late March we started with two complex surveys. One survey was a wave of the Socio-Economic Panel (SEP). Data from the previous SEP wave were sent to the interviewers for use in the new wave. The other survey was a recruitment for the SEP. Both surveys used the same Blaise questionnaire. SEP used to be a Blaise CADI application: All data collected with paper and pencil were entered at the bureau using Blaise. The CADI questionnaire was used as a basis for the CAPI questionnaire. Blaise makes this step easy.

During the remainder of 1991 the system was fine-tuned and gradually the number of interviewers was increased to approximately 500 by the end of 1991. A lot of extra features were added during the adjustment phase to make LIPS a more secure system and easier for the CBS staff to support. For instance:

- At every start-up, LIPS checks whether the system is completely installed and whether no unidentified files are present in the laptop. LIPS also checks the data integrity of all surveys in the laptop. If something is wrong the interviewer is instructed to contact the CBS.
- The option of sending an exact copy of all survey data stored in the laptop to the CBS was added. This proved to be important in case a disk drive failed.
- A number of statistics regarding the use of LIPS are sent to the CBS each time data communication is performed.

During 1992 all 300 Baud modems were replaced by 2,400 Baud modems. This was done because the old modems lacked full compatibility and sometimes caused problems. Also during 1992 all laptop computers were upgraded with 2 MB extra non-volatile RAM. This made it possible to run the Blaise questionnaire from RAM instead of diskette. The start time of a questionnaire was thus reduced from approximately one and half minutes to a few

seconds. Also, by adding extra RAM the disk-drive is used less which saves battery power.

The following monthly figures illustrate the use of the data communication system. The figures are the average of the first nine months of 1992.

Addresses sent:	14,000 for seven different questionnaires.
Sent to field:	870 KB in compressed format in 500 sessions (an average of 1.8 Kb per session).
Received at CBS:	6,300 KB also in compressed format in 1,600 sessions (an average of 4 Kb per session).

4. CATI Management

The CBS has been conducting CATI surveys for about 10 years. The original ran on a mini-computer. CATI surveys are the responsibility of the field work department, the same department for which the CAPI management system described in Section 2 was created. The old system had two main disadvantages:

- The mini-computer did not fit into the standardization of hardware in use at the CBS (see Bethlehem, Kellenbach and Keller 1991). The policy of the bureau is that primary data entry should, if possible, be done on PCs and not a mini-computer.
- The options for defining a questionnaire were inferior to those of the Blaise system. It was believed better to have only one computer language for defining a questionnaire. The use of Blaise was widespread in the field work department, while knowledge of the old CATI system was restricted to only two or three people. If Blaise could be used for conducting CATI surveys the knowledge of the new system would also be spread without a great deal of training.

But Blaise lacked proper management for a CATI survey. For this reason the CBS tried to build a CATI management system into Blaise as part of the general replacement of the old CATI system by Blaise. This project was started in April 1989.

There were two main starting points:

- The new system had to be fully integrated in the existing Blaise system.
- There had to be numerous options for setting an appointment. The convenience of the respondent had to be accommodated for whenever possible.

The project resulted in the first operational version of the Blaise CATI system in September 1989. This CATI system has been used since that date and has replaced the old system completely since January 1990. The remainder of this section gives a short description of the Blaise CATI system.

4.1. Blaise CATI

The functioning of the questionnaire in CATI is the same as in CAPI. Displaying the questions on the screen, entering the answers, and reporting and correcting errors are performed in exactly the same way. CAPI and CATI differ only in the management facilities of the survey. Blaise CATI has an elaborate management system. The main task of the management system is to ensure that the telephone numbers for a survey are presented to the interviewers, while constantly maintaining appointments at the stated times.

To perform its task, the CATI management system needs information which is stored in the *survey definition*. The survey definition includes, for example, the survey

period, a specification of the days on which interviews will be held, and the number of interviewers, i.e., shifts. The survey definition also indicates which treatment must be given to telephone numbers in several situations. For example, how many times a number should be called back if the line is engaged or no answer. This definition can, if required, be changed while interviewing is in progress.

Before interviewing can start, a list of telephone numbers has to be entered into the system. This can be done with the Blaise conversion module. For a large CATI survey the management system would have to manipulate a large number of telephone numbers, which could deteriorate the performance of the system. To avoid this problem Blaise CATI works with *day batches*. A day batch is a list of telephone numbers to be contacted on a certain day in the survey period. A day batch is created by the supervisor of the system using a module that is part of the management system. The day batch is valid only on the day for which it has been created. Whenever an interviewer asks for a new telephone number, it is retrieved from the day batch. The management system chooses the next number by taking into account the priorities of the numbers in the day batch.

When all these things (survey definition, telephone numbers and day batch) have been taken care of, the interviewing itself can start. The interviewer starts the CATI program and asks for a telephone number. The treatment of a telephone number is called a *dial*. A dial can produce one of several results. Two groups of dial results are distinguished:

- *No contact*: Characteristic of this group of results is that the questionnaire need not be started. This group distinguishes the following result types: *no answer*, *busy*, *answering service*, *disconnected*, and *other*.
- *Contact*: A result of this group can only be obtained if the questionnaire has been started. This group distinguishes the following results: *response*, *non-response*, and *appointment*.

The dial result determines the further treatment of a telephone number by the management system in the day batch: a number with the result *no answer* or *answering service* will be given the no answer treatment, numbers with the result *busy* will be given the busy treatment, and a number with the result *appointment* will be given the appointment treatment. The other results (*disconnected*, *other*, *response* and *non-response*) are considered to be concluded.

The idea behind the busy treatment is that a number with dial result *busy* must be quickly dialed again as there is a fair chance of getting someone on the line. It is, however, also possible that the receiver has not been replaced. If a number gives the dial result *busy* several times in a row, it is better to stop the busy treatment. The survey definition specifies how many quick re-dials should be tried for the busy treatment. It also specifies the time between subsequent dials.

With the *no-answer treatment* a number of dials are performed during the rest of the day. Should this not lead to an end result, then the procedure is repeated for a number of days.

For both the busy treatment and the no answer treatment we speak of *dials* and *calls*. For a given telephone number all dials on a given day are called a call. The survey definition specifies the maximum number of dials per call. It also specifies the maximum number of calls to be performed for a given telephone number.

With the *appointment treatment* the interviewer makes an appointment with the respondent for a new contact. Four appointment types are distinguished:

1. *No preference*. The number will be redialed during the survey period, but not on the same day.
2. *Appointment for a specific date and time*. For this appointment type a date and a time have to be specified. The appointment must be met at the specified time (if, of course, the CATI system is operational at that moment).
3. *Appointment with preference for a certain period*: This appointment type specifies two dates and two times. The interviewer does not have to fill in the dates or the times him/herself. In that case the entire survey period or the entire day will automatically be selected. The number will be called back in the specified period and time slice.
4. *Appointment with preference for a day in the week*. An appointment of this type consists of one date and possible two times. If the respondent wants to be called back on a Thursday, then CATI looks for the date of the first Thursday in the survey period after the day on which the appointment was made.

CATI assigns a priority to telephone numbers based on the type of appointment that has been made and the number of dials that have been performed. For an appointment for a specific date and time (type 2) the priority is always *hard*. For an appointment with preference for a period (type 3) a number of days are usually available to contact the respondent. In that case the priority is *soft*. On the last day in the specified period the priority is *medium*. If no appointment or an appointment without preference (type 1) has been made, or if previous dials have led to the result *busy* or *no answer*, then the priority is *default*. Finally, appointments that have been made by the supervisor have the priority *super*.

The management system retrieves the telephone numbers to be presented to the interviewers from the day batch. Of course, this is not done randomly. The choice depends on the *priorities* of the available numbers.

The management system works in five-minute intervals. The priorities are updated every five minutes. In the meantime treatments for numbers may have been concluded, new numbers may have become available, and new appointments may have been made. Although the day batch contains the same numbers all day, the set of active numbers varies continuously. The priority assigned to a number may also change many times.

In the CATI management system a number of interactive tools are available for the supervisor to help him/her with this task. There is a tool which offers a dynamic view of all the cases present in the day batch. Another tool generates a report of the complete survey. It is also easy to obtain production statistics for individual interviewers.

4.2. *Evaluation of the new system*

The new system now has been used for several different surveys. It took the interview department some time to get used to some of the new features in the system.

For instance, the numerous opportunities for setting appointments caused planning problems. In the old system there were only limited options for making an appointment. Based on experience the number of interviewers could be planned one month in advance. It turned out that the old experience was of limited value in the new system.

From an information analysis point of view, it is good practice to store data for running a process separately from data that are part of the process. With this in mind we should have stored all data for the call management separately from the interview data. We did not do this, because CATI had to be fully integrated in the existing Blaise system. For this reason we made the call management data (and the appointment data) part of the questionnaire and this turned out to be an excellent choice. All the tools in the Blaise system for manipulating data can also be used for CATI. For this reason no separate tools had to be developed for loading telephone numbers into the CATI system.

The CBS has a group of free-lance interviewers who do only CATI interviews. This group received training in Blaise, with extra attention for the CATI options. The step from the old system to Blaise caused no significant problems for the interviewers. Nevertheless, in comparison with the old system, the numerous operational possibilities of Blaise did cause some minor adjustment problems.

5. References

Bethlehem, J.G. and Keller, W.J. (1991). The Blaise System for Integrated Survey Processing. Survey Methodology Journal, 17, 43–56.

Bethlehem, J.G., Kellenbach, F.M., and Keller, W.J. (1991). Computer Assisted Statistical Information Processing at the Netherlands Central Bureau of Statistics. CBS-report, Netherlands Central Bureau of Statistics, Voorburg.

Braslins, A., Coutts, M., Jamieson, R., and Williams, B. (1992). The Building of an Integrated Collection Operation in Statistics Canada's Regional Offices: Time for a Change. Proceedings of the Annual Research Conference, U.S. Census Bureau, Washington, D.C.

CBS (1987). Automation of Survey Processing Select 4. Netherlands Central Bureau of Statistics, Voorburg.

Manners, A.J. (1991). Development and Implementation of Computer Assisted Interviewing for the British Labour Force Surveys. Paper presented at the ASA 1991 Joint Statistical Meetings, Atlanta, GA.

Nicholls, W.L. and Matchett, S.D. (1992). Computer Assisted Survey Information Collection Issues at the Census Bureau. Proceedings of the Annual Research Conference, U.S. Census Bureau, Washington, D.C.

Kerepesi, K. (1992). Labour Force Survey at the Hungarian Central Statistical Office. First International Blaise Users Meeting, Voorburg, The Netherlands.

Kuussela, V. and Merisalo, A. (1992). A Prototype System of Data Exchange for Statistics Finland's CAPI system. First International Blaise Users Meeting, Voorburg, The Netherlands.

Statistics Sweden (1992). Computer Assisted Data Collection Production Tests in the Labour Force Surveys, August 1989. Stockholm, Sweden.

Received December 1991
Revised May 1993