

Features of Case Management in CAI Systems

*Vesa Kuusela, Social Survey Unit, Statistics Finland and
CMS working group set by the Blaise Corporate License Users' Board (BCLUB)*

1. Introduction

A Case Management System (CMS) is a central part in all data collection systems which are based on Computer Assisted Interviewing (CAI). In general, case management here means procedures and arrangements for handling sample points¹. This definition presupposes that a sample with adequate contact information is drawn prior to the actual data collection. This excludes for example RDD sampling methods and quota sampling. This definition also excludes other tasks in survey undertaking, such as the installation of surveys (questionnaires and supporting files) and monitoring of data collection. However, these functions are needed in a CAPI system. Especially, the up-to-date monitoring of fieldwork is an essential task in supervision. Figure 1 shows schematically other relevant parts of a CAPI system and their linking.

Technically data collection modes differ considerably from each other, but even more they differ in how data collection is organised in practice. The fundamental division is between the modes of administration: Face-to-face interviewing (CAPI) and telephone interviewing (CATI) are interviewer-administered modes while web interviewing (CAWI) is a self-administered mode. Especially, in self-administered mode the case management is inherently different from that in interviewer-administered modes. Occasionally, it has been desired to establish a CMS that would handle all data collection modes. This paper analyses the typical requirements for a CMS in order to define a universal kernel.

Data collection with CAPI requires a field interviewer organisation which covers sufficiently the region where selected households (sample points) may reside. Interviewers work with laptop computers (in stand alone mode) and they have an arrangement to communicate electronically with the office. The size of a field organisation is determined by the size of country (or the area from where data are collected), time table of data collection, the number of languages, and the actual field organisation. A field organisation may be centralised in the sense that supervising and management is done in one place only, or the organisation may be partly or completely decentralised. A CMS has to comply with the field organisation and with the requirements given for data collection.

In CATI mode, data collection is carried out by a centralised system. Interviewers are working in one or more CATI centres and their work-stations are connected to a network and a file server. The central part of a CATI system is an application called Call Scheduler (CS). The CS feeds cases from the sample file to available interviewers (actually to work-stations) and maintains appointments and management information.

In CAWI, no interviewers are needed. Respondents have to be invited to open the questionnaire and answer the questions. The central component in CAWI survey is the computer system accessible via the Internet, in which the sample is loaded prior to data collection. Case management tasks are inherently different from CATI or CAPI. Within the technical system there may be some tasks resembling case management but they are excluded here.

Applicability of different data collection modes depends essentially on the data provided by the sample file. In many cases, a sample is drawn from an address register or from a database from a previous census.

¹ Case management is called sample management in some occasions.

In this case, the sample point is an address but it is not known whether it is inhabited or who is living in it. In some countries, the address may be to a block of flats and one flat may include several households. In this case, the final stage of sampling is carried out by interviewers according to given rules.

Some countries have well updated registers of population and households. Samples from such registers pin point exactly the person or the household to be interviewed. In addition, in some countries there are registers for telephone numbers which can be used for sampling or it is possible to find telephone numbers for persons sampled from the population register. Obviously, email addresses are not available in public databases so that they could not be included in the sample data.

In multi wave surveys, the information about a sample point available before the first wave and later in the subsequent waves may differ considerably. Therefore, also the case managements may be organised differently.

2. Case management in different data collection modes

Because of their technical and organisational dissimilarities, the different modes of CAI set rather different requirements on how CMS should be arranged. Also the possibilities to design the CMS vary considerably between modes.

2.1. CMS in CAPI

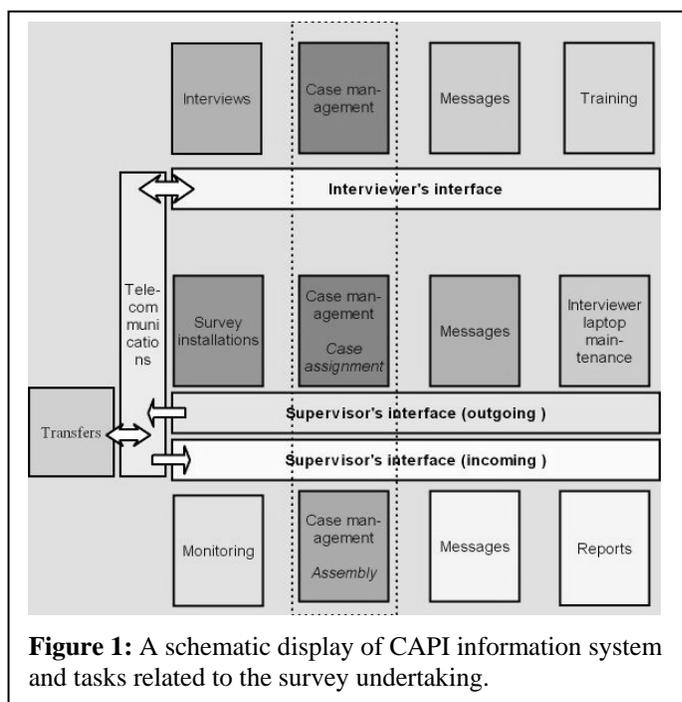
The field organisation for CAPI has to be geographically organised because interviewers have to visit the selected households. The average distance between interviewers' homes and potential respondents has to be reasonable in order to keep the travelling costs reasonable. Because of the disperse nature of the organisation, the case management system for CAPI is far more complex than in other modes. In addition, the field organisation partly determines the production process in CAPI surveys.

Characteristic features of the data collection are that all interviewers work with stand-alone laptop computers; they have dedicated samples, i.e. the total sample is exhaustively divided between interviewers and the sample file for each interviewer includes only his or her sample points; interviewers organise their work and make appointments on their own; and they return cases, both completed and non-responses, within the given time limits. This work process introduces three critical tasks for the CMS: case assignment, delivery of sample files to appropriate interviewers, and receipt of completed interviews from interviewers. A special problem in the CAPI case management is how reassignments or case transfers are handled. The reassignment of a case is needed when the interviewer to whom the case was assigned first is not able to carry out the interview. In this case, the case information has to be moved from one interviewer's laptop to the new one.

The interviewer workforce for CAPI can be organised in several different ways and this organisation determines how the case management can be organised. Obviously, the personnel organisation is a more decisive factor than any other part of the production process and therefore, the case management system has to be adapted to it - not vice-versa. Changes in personnel organisation are difficult and slow and they are avoided if other solutions are possible.

From the perspective of CMS, the important question is how supervising is arranged. A major difference between field organisations is whether the supervisors work in one place (centralized organisation) or at regional offices (decentralized organisation). In some cases, the regional organisations are extensive and very strong. The regional offices may even have more influence on survey undertaking than the central office, which may have only a co-ordinating role. On the other hand, in some organisations all management and supervision is done in one place. Often there are firm historical reasons behind the existing organisation.

Despite the fact that the organisation of the interviewer workforce varies, the technical part of the CAPI system includes always some essential parts as show in the Figure 1. The schematic display refers to an automatic system, but the same tasks appear also in a completely manual system.



An automated CAPI information system includes three different user interfaces: two for supervisors and one for interviewers. All user interfaces include case management, but its roles differ because the tasks in different phase of a survey are different.

In the interviewers' interface, the case management application passes the sample information to the CAI application (such as Blaise DEP), and transports the interview data and changes in sample information to a place from where it will be transmitted during a telecommunications session (or by other means of transmission). The architecture of the CMS is an important factor in how case management is performed on interviewers' laptops. The solutions can be divided into two major classes: database based systems and object based systems. In a database based system the completed interviews are stored in a database which is transferred to the office system as one file; in object based system each interview is stored in a separate file and the separate files, "objects", are transferred from interviewers' laptops to the office system. The object based system requires a specific application on both ends to handle the objects adequately, but this architecture makes the case management technically fairly straight forward and sturdy.

In the supervisors' interface, the tasks serving outgoing and incoming activities differ essentially also in what comes to case management. In the outgoing part, the most important task is case assignments for interviewers. If it is automated, it requires some sort of a database application which connects respondents to interviewers using geographical information. Therefore, there has to exist a database which includes information on interviewers' whereabouts.

In the interface for handling of the incoming data, the most important part of the case management is the process that assembles data sent by different interviewers to a single data file. Another important part is the application that processes changed contact (sample) data in multi wave surveys, such as the LFS.

A necessary part of a CAPI information system is the communications system which is often a telecommunications system. Typically, CAPI interviewers' communication system is not active all the time. Instead, interviewers contact the central facility at a given frequency. The frequency of communication may vary considerably between organisations. Probably, the most common arrangement is that interviewers open telecommunications system once a day, but the frequency may range from several times a day to

once a week or even once in a month. In some countries interviewers do not have communications facilities at their homes. Instead, they have to visit the regional offices to send and receive data.

A communication system is a central part of the CMS and its features partly determine the structure of the CMS. For telecommunications, a CAPI information system requires a telecommunications server that handles the traffic between interviewers and the office. There are several technically different systems in use and obviously they are all built in-house. The most difficult task, however, is to design the communications protocol: how each interviewer receives his or her data once and only once and how interviewers' data are received safely. This protocol is intimately connected to the CMS.

In CAPI surveys, interviewers seek out the responding households using the information they have in the sample file and using their knowledge of the local circumstances. Occasionally the knowledge of local circumstances is the most helpful asset. It is necessary that to some level the case assignment is based on geographical information to keep the distances between interviewers' residences and respondents' homes reasonable.

In some cases, the need for a CAPI arises from the fact that only the whereabouts of the sampling unit (e.g. address of the sampled household) is known, but not who the respondents are or whether the dwelling is inhabited at all. CATI and CAWI are out of the question in such cases. This situation occurs frequently when a sample is drawn from an address register or in connection with area sampling. This situation may also arise if the sampling frame is inadequately updated.

CAPI sets some stringent requirements for the sample information: The locations of sampling units or households' (respondents') addresses are necessary. Names of household members and telephone numbers are not necessary for CMS, but they facilitate interviewers work in setting appointments.

2.1.1. A common CMS kernel for CAPI systems?

The CAI software packages do not include a ready-made CMS. Considering the facilities that have been incorporated for CATI management, it should not be unreasonable to expect some survey or case management facilities to be provided for CAPI. Especially, it would be important for smaller organisations that don't have existing facilities for survey operations. Wensing (2009) presented an example of a CMS for small organisations.

Is it possible to find a common kernel for an automatic CMS? A common kernel requires a definition of a standard CAPI system. The CMS is inherently related to installation of the technical environment for CAPI. Obviously, it has to have separate solutions for the central parts: case assignment, CMS within interviewer's laptop and assembly of a single data file from the received interviewers' files. In addition, an automated CMS cannot be completely detached from the communications systems.

Obviously, the installations of the CMS within interviewer's laptop do not vary too much. For that part a simple solution serving the basic needs is fairly straightforward. It does not solve the total needs but it would help small organisations in the beginning.

Case assignment is at the core of the CMS and it depends essentially on the information available about sample points. In some countries case assignment is partly manual because available information on sample points is not sufficient for an automatic system. For automatic case assignment, it is necessary that there is a sample file that contains respondent information, and a database containing interviewer information. Both databases have to include a field (or fields) that are used to connect a respondent to an interviewer. This field can be called an Enumeration Area Code (EAC). In CAPI it has to be based on geographical information, such as postal area code. The structure of the EAC can be hierarchical and consequently the case assignment can be done in two or more phases. In interviewers database the EA codes have to be unique for each interviewer and the codes have to cover completely and exhaustively the geographical area to be surveyed to end up with proper case assignment.

If no Enumeration Area Code is available, automatic case assignment is not possible and it must be done manually or semi-manually. The Appendix contains a description on how case assignment is done in some organisations.

The end results of case assignment are the sample files dedicated to each interviewer. The files will be delivered to the proper interviewer. Delivery of the files is closely related to the communications arrangements.

2.2. Case management in CATI surveys

A CATI system is typically a centralised facility running in one place (or several places) with interviewers using (desktop) work stations that are connected to a computer network and a file server(s). The sample file is handled by a file server and a program and the cases are not dedicated to specific interviewers. An application called Call Scheduler manages the sample. In contrast to a CAPI system, in CATI there is no case assignment in the same sense and there is no need for sample file delivery. Completed interviews are stored automatically in the database and case status is updated automatically. Because there are no dedicated samples there are no case transfers in the same sense as in CAPI. Similar situations may arise if cases should be transferred from one group of interviewers to another group (e.g. between language groups) but usually CATI systems include tools for supervisors to perform the transfer easily.

A major difference to CAPI is that interviewers working in a CATI centre do not actively organise their own work. Cases are automatically assigned by the Call Scheduler and interviewers are not able to - and supervisors do not need to - influence the process. The Call Scheduler manages the sample by feeding new cases to interviewers according to predefined rules, and maintains the statuses of cases. The Call Scheduler also manages appointments and interviewers have only limited possibilities to influence the process. If manual operations are needed in case management, it is a task solely for supervisors.

In a CATI survey, it is necessary to know respondents' telephone numbers and preferably also their names. Postal addresses (or email addresses) are needed if advance letters will be sent but it is not necessary for CMS.

2.3. Case management in CAWI surveys

Data collection by CAWI does not involve interviewers and consequently there is no case management in the same sense as in CATI and CAPI. Especially, there is no need for case assignment, which is the most demanding task in CAPI.

In a CAWI survey, case management includes installation of the survey on a web server including the arrangements for the authentication of respondents; publication of the survey, i.e. the announcement of the address of the server and login information for respondents, accompanied with the presentation of the survey; follow-up of accumulation of responses; and sending reminders to non-respondents.

In a CAWI survey it is difficult to know the status of cases during the survey because respondents answer at different rates. Even though the answer has not been stored in the server by the given date it does not mean that the respondent is not going to answer later. It is a characteristic feature of a self-administered survey that the collection of data has to be stopped at some stage although answers are still arriving. In interviewer-administered surveys the schedule of data collection is set before the survey begins and the status of each sample point is known.

In a CAWI survey it is necessary to know respondents' names and postal addresses or their email addresses.

3. General case management system for mixed mode surveys

In a mixed mode survey, at least two different data collection modes are used simultaneously or in succession for data collection. For example, data collection starts as a web survey (CAWI) and in the next phase respondents who have not answered are called (CATI) or interviewers visit those respondents whose tele-

phone number is not known (CAPI). Another example may be a CAPI survey in which a refusing respondent is offered a possibility to answer via the web. In multi-wave surveys, different modes are frequently used in different waves, but strictly speaking, they are not mixed mode surveys.

The possibility to carry out a mixed-mode survey depends on the information that is available on the sampling units. More information is needed if all modes should be possible because all modes require different contact information. In cross-sectional surveys this may pose a problem. In a multi-wave survey it is possible to add contact information after the first contact and maybe also to settle the mode of the next interview.

Installation of automated case management for a mixed mode survey is not straight forward because the tasks of CMS are inherently different in CAPI, CATI and CAWI. It is difficult to find common factors in them. Even the interviewer-administered modes, CAPI and CATI, differ essentially from each other in this respect. In addition, the information available on the sample points determines what is possible. For example, if a telephone number is not available, CATI is out of the question; or if there is no adequate address available for respondents, CAWI is not possible.

Obviously the automated case management for mixed mode surveys would require a management server in which a daemon process handles contact information and passes cases from one mode to another. Each mode has a separate CMS incorporating its typical features. A case can be active only in one CMS at a time. Each separate CMS has to know the status of each case. Status includes also information about in which system the case is active. When the status of a case changes, the new status should be made known in each CMS. A technical difficulty is posed by the fact that unlike in CATI and CAWI, in CAPI the connection to interviewers' applications is not continuous. Moving cases from or to CAPI interviewers' workstations is delayed and the length of the delay is not known. In addition, moving cases in or out of interviewer's laptop requires consent from the interviewer.

4. Conclusions

Each data collection mode has its characteristics which set different requirements for case management. In addition, case management may be an important factor behind the efficiency of the data collection system. It is a danger that if case management systems of different data collection modes are forced in a single system the efficiency of data collection in all modes will be compromised.

In a CATI system, the well functioning call scheduler is an important factor behind the efficient data collection. Technically, it would be possible to install a CAPI system with direct access to case management at the office using new wireless telecommunications facilities - even a system similar to a typical call scheduler. However, it is difficult to see any benefits in a system like that. Cases to be interviewed face-to-face have to be directed to interviewers who are living close to the respondents. This, in turn, leads to dedicated samples and the call scheduler does not add efficiency. Probably it would make data collection less efficient.

The design of a common CMS for mixed mode surveys is not straightforward. It is technically challenging, but not too difficult. It is another question whether it is possible to find a common kernel, which all organisations could apply. Probably all organisation who have collected data by CATI and CAPI for many years have a working case management system of their own, which they presumably will use and develop in the future. A common kernel for a CAPI CMS would be most useful for new (and probably small) organisations that are starting computer assisted data collection.

References

Wensing, F. (2009). Development of Survey and Case Management facilities for organisations with minimal survey infrastructure. In *Essays on Blaise 2009*. Proceedings of the 12th Users' conference, Riga, Latvia. (Also at <http://ibuc2009.blaiseusers.org/papers/2c.pdf>)

Appendix:

CAPI case assignment in some countries

INSEE (France)

In France, case management is done in two phases:

- Phase one: an automatic pre-assignment is done centrally, based on a database containing the area for each interviewer that he/she covers, and the survey that he/she is interested in (this database is updated once a year). This is done by the method unit.
- Phase two: regional offices check the case assignments and can change what they want. The tool for manual re-assignment is unfit for large assignment. This work is done by the interviewers' supervisors.

ONS (United Kingdom)

At ONS, the sample is drawn using the sampling system and stored on a pre-defined table, which contains all sorts of geographical information about the location of each case.

Cases are aggregated into quotas of work for interviewers. This is usually done as part of the sampling so either we select clustered samples from postcode sectors (with a fixed number of addresses selected per postcode sector), or, for unclustered designs, we pre-define a geography and see how selected cases fall within it (producing a variable number of addresses per quota of work).

There is a separate allocation system for assigning quotas to field interviewers. This is de-centralised, the allocation to interviewers is performed manually by field managers, based on the information they have about availability.

Once the allocation is finalised, IT packages the individual cases into case objects, and scatters them to interviewers along with the relevant questionnaire object. We have systems for creating the case and questionnaire objects, and a separate system which places the objects into the correct mailbox, based on the allocations. When the interviewer connects to the Office, the objects are automatically downloaded onto the laptops ready for manual installation by the case management software.

Interviewers work off-line, and need to connect to the Office to transmit completed cases. Casebook zips up the completed case objects and places them in a specific location on the interviewer's laptops. When the interviewer connects, all objects in that location are picked up by the Gather systems.

Statistics Canada

SC has an in-house interviewer messaging system that allows messages to be passed up and down the reporting hierarchy. Interviewers can only send/receive these messages when they dial in. Interviewers work off-line and transmissions are usually once per day.

Case assignment is manual. For ongoing surveys, where the sample units come back periodically, the previous cycle's assignment information is used as the starting point, and is manually reviewed and revised.

For new samples, the data collection managers in the regions look at the geography information and manually assign the sample to the appropriate interviewer. Generally, the basic information about where the sample is being selected is provided in advance so that hiring decisions can be made

The actual assignment planning activity (assigning cases to interviewers) does not occur until the sample is loaded into the case management database and made available to the managers in the regions.

Statistics Netherlands (SN)

Interviewers work off-line. They have a communications-system on the laptop for connection with the office. When an interviewer connects to SN, the cases (addresses) are automatically downloaded onto his/her laptop. Interviewers connect to SN on a daily basis. At the end of a survey, interviewers return interviews (questionnaires) to SN together with logistic information.

The sample is drawn by a sampling system. The sample includes information about some (background) aspects of the respondents. The sample is then loaded into the system. If respondents have an actual phone number, they will be interviewed by phone (CATI), otherwise (for respondents without actual phone numbers) a face-to-face interview (CAPI) will be held.

Respondents are first allocated to field interviewers automatically by the system. Region Data Collection Managers (all 13 regions have own data collection manager) can make some changes manually in the allocation. They can assign manually the sample to the appropriate interviewer. When the allocation is finalised, addresses are dedicated centrally to each interviewer.

Statistics Norway

The interviewers work off-line from their homes but at least once a day they connect to the office.

A sample is drawn by a separate sampling system and imported into the CMS. The sample is drawn according to our geographic cluster plan (interviewers are located in pre-sampled clusters). The case assignment is done automatically according to the geographic clusters and postcodes. If some cluster is without interviewer (due to illness or vacancy), the field managers do the re-assignment partly manually.

After the case assignment process, cases are prepared for transfer by zipping each in a file. When an interviewer connects to central office he/she reaches his/her own homepage which shows workload and progress, distributed on the different surveys. From the homepage the interviewer starts data transfer. During this process questionnaires and cases are copied out to the interviewer's laptops and completed cases are sent back to the office. The zipped files are unzipped automatically and interviewers find the new cases in the case list. By Statistic Norway the assignments the manual reassignment work, and re-assignments (e.g. in case of illness, slow progress etc.) are time consuming.

Statistics Finland

Case assignment is done centrally and automatically. Each CAPI interviewer collects data on a specified geographical area. The areas are defined using zip codes (postal codes). Technically, the case assignment is done simply by assigning the interviewer code on each record in sample file by a merge procedure based on the zip code. Supervisors do some refinement on the assignments based on interviewer's workload or temporary leaves.

In the earlier version of the CMS, the sample file was divided in parts so that at the end there was a dedicated sample file for each interviewer. In the current system the whole sample file is sent to each interviewer and the interviewers' interfaces produces a view into the sample file with only the specific interviewers' cases.

Case assignment is still based on postal area codes but now the cases for an interviewer are given in a small file that includes only the interviewer code and sampling unit ID (separately for each survey). This file is sent to each interviewer with the total sample file and it is used in the CSM in interviewer's laptop to create the view into the sample. Cases are transferred from one interviewer to another simply by changing the information in this case assignment file. All installations in interviewer laptop take place automatically in the laptop.

Completed cases are returned to the central office automatically when interviewer opens the telecommunications system which takes place usually once a day.

Some other countries

In previous Eastern Block countries the arrangement in survey work is close to the organisation in France. Most of the practical work is done in the regional offices. Automation of survey data collection is still under way but in most cases the plan is to use regional offices. In few countries (e.g. Latvia), the system is centralized. In those countries where CAPI is in use, case assignment is semi-automatic.