Around Blaise-surveys at Statistics Netherlands

A new ship for a voyage to the Mixed Mode Archipelago

Marien Lina
Blaise support
Statistics Netherlands

Introduction

In 1998 and 1999, all Blaise surveys at Statistics Netherlands were converted to Blaise for Windows. Around the questionnaires and Cati management, a variety of software applications were active for survey management and administration. These applications are in part inconsistent and old-fashioned. The Data input division for Person and Household Surveys of Statisics Netherlands searches for a new system to replace the available variety of tools. All surveys should be able to supply consistent information to this system. Increasing demands on flexibility (for example, combining mixed mode with multiple waves) ask for one integrated system. This paper aims to outline the shape of this integrated information system for managing the survey process at Statistics Netherlands.

This project of developing a new management system for the survey process is still in preparation. Needless to say that the final project may differ from the ideas below, taken from the draft version of the project plan.

Blaise Island in the past century

In the past century Statistics Netherlands has been developing Blaise. The developments went fast. In the eighties there were merely tools for data entry and cleaning forms. In the early nineties Blaise was in use on laptops for the Dutch labour force survey. At Statistics Netherlands, this was the starting point of computer assisted interviewing. At that time the main focus was on computer assisted interviewing itself. At Statistics Netherlands software development for survey research concentrated on the Blaise data entry program: the part of the software that enabled to show question texts on screen and to arrange the order of the questions in a flexible routing mechanism.

In 1989 the Cati-management system has been added. An increasing demand for information asked for larger samples, new surveys and more frequent panel interviews. Apart from CATI-management, other features at that time (Abacus, Manipula, Bascula, Set-ups, Conversions, Coding, Form manager) are not really integrated in Blaise but they were a bunch of independent "tools" placed in a sub-menu of Blaise 2.5. Tools were added to Blaise incidently if there was time to develop a new functional tool for somebody who asked for it. The most important tool around Blaise at that time was Manipula. In the past 10 years Blaise III and 4 have introduced Maniplus and a more consistent language definition for Blaise and Manipula.

CATI island in the past century

On Cati Island there are about 100 interviewers. They work in shifts. The CATI-calling centre at Statistic Netherlands has 54 chairs.

The survey procedures at the CATI call centre are well structured when it comes to handling appointments, non-response and the status of a call. Many credits here for the Cati-call management system. The history files supply standard information, enabling a structural approach of dial results. For example, for monitoring the performance of the interviewers are retrieved from the history files. Management information (time consumption for calling and time in between) from the history files is used for planning interviewers capacities. This kind of small investigations were not performed on a structural base, mostly using Manipula (or home made Pascal routines). Apart from the data in the Cati management administration there was no big administration. The supervisor has a simple (Blaise III like) menu to create overviews of the Cati-surveys (per interviewer, per day or survey period). Non-Response reports are edited manually. Data from the Cati-management system Tools for coding, converting and finishing data were developed outside Cati-island. The advantage of the cati-management system was that the information about Cati in the interview data is uniform.

CAPI island in the past century

For CAPI there are about 550 interviewers in 80 districts around the country. They all have a laptop with a modem at home. They contact field work managers at Statistics Netherlands directly.

There is no Blaise standard CAPI-management system (like the CATI-management system). Outside Blaise a interview administration produces standard tables. It includes 'reasons for non-response' tables from the viewpoint of the data collection department. The figures may differ from the non-response definition for the survey from a statistical viewpoint. The non-response classification in EA is merely arranged to give account of proper handling of visiting addresses by the interviewers.

Additional tools have been developed outside Blaise. They cover CAPI-management (like CATI-management) but also tools for data communication, the status of an interview, processing visit records and for keeping the salary-records. These CAPI tools have been developed completely separated from the tools for CATI.

Despite an increasing CATI share, CAPI remains an important input source for statistical analysis. The old tools work together as a solid machine, at least, if they are handled correctly. However, many of the tools for CAPI management ask for manual interventions, are survey dependent, old-fashioned and therefor they ask for replacement by a more consistent Capi-management system.

The Survey Bay

At Statistics Netherlands, organising and executing the field work for the surveys is the task of GEP or "Gegevensverzameling; Enquetering Personen" (in English: "Data collection; Person Interviews"). GEP consists of 70 staff-members persons. They supervise and execute the survey process. Each survey has its own process manager. They guide and monitor the whole process: maintaining the questionnaires, training interviewers, collecting data, coding, quality control, planning and data delivery. The interviews are performed with Blaise.

The organisation of the GEP-staff roughly is devided into CAPI and CATI-surveys. A closer look shows that survey procedures are different for every survey. This asked for survey-dependent protocols. In this sense each survey has its own boat to get the Blaise data from Cati- and Capi-Island. For example, the nature and the amount of finishing jobs (coding, quality control and data delivery) depends on the survey. In the past special help-tools and procedures have been developed for specific surveys. These individual tools are based on a diversity of DOS-applications (Blaise 2.5, Paradox, Pascal, Cobol). The applications needed frequent maintenance.

The rowing boats to Cati Island

For each Cati-survey there is a "rowing boat". The boat delivers the sample data, and if necessary a new questionnaire, external files, or even a new Blaise version to Cati-Island. Afterwards the rowing boat picks up the interview data (including Cati-management data), updates statuses of interviews, creates some non-response figures, determines the destination of a record and does some administration on board. The procedures on board imply many manual interventions. If someone disturbs the cadence there is a big problem to transport the data from wave to wave. The boat usually manages to get the data safely to the Survey Bay.

The old vessels to Capi Island

A lot of CAPI-management is still done manually. An increasing amount of data collection asked for a new system around Blaise, reducing error caused by manual actions. In the nineties there were four sub-systems around Blaise that managed a part of the required functionality around a survey. Together they form the system called "SPEER". The sub-systems are:

SPIL (System to Perform dataInterchange with Laptops). This system has been designed to bring sample data to the laptops of the CAPI interviewers at home.

LIPS (Laptop Informationsystem for Personal Surveys).
This system is on the interviewers' laptops to start datacom, load interiews, make visit reports, remarks and send completed interviews back to the office.

COBS (Computer assisted data manipulation)
This system has been designed for automated data manipulations after the interview data returned at Statistics Netherlands (for example for coding).

EAS (Interview Administration System)
This system keeps track of the status of an interview task. The survey managers use the system to check the response figures. It is also used for the pay-roll of the interviewers. This system has been changed and adapted to meet the needs of different surveys. The result was merely a number of different management systems than one standardised system.

The old system has many loose boats and ships, drifting somewhere in the ocean. Nevertheless also in the current system, automated sub-procedures (such as keeping the pay-roll) and fixed protocols ease management and supervisors. To call this

system a fleet of old vessels is not meant to be a disqualification of its design or the expertise and/or the many investments of its builders. Until today these boats are still on duty and they still succeed to deliver the load they have to.

The exploratory drilling rig called POLS

Until now there is not a consistent system to direct the boats of the fleet in the right direction. As already mentioned each survey has its own management and there was no drive to a uniform approach. In fact every Cati / Capi and Cadi (sub-)survey was an island. This has been changed when the Life Situation Survey (POLS) came alive. POLS was the start of a new approach in which all surveys were regarded as one entity. POLS combined three important surveys (about medical care, life conditions and legal protection) in one system.

For those surveys uniformity in the interview organisation and administration has been reached. However, the three surveys have in common that the observation unit is a person and that for each address only one person has been selected in the sample. The whole system around POLS has been designed for the situation where only one person of a household was questioned.

Unfortunately, other surveys (for example the EBB (labour force survey) and the PAP (a survey about private car use) do not fit on the POLS platform. They have other sample units (households and companies) and may imply multiple observation units per sample unit (persons and license plates). So to say, the drilling isle was too much restricted to one situation, not flexible enough and there was a need for a new ship.

Currently the system meets the point where innovation is required. Changes are needed to improve the handling of mixed mode surveys, stratified sample units (persons in households), panel management and consistency in administration and organisation.

Searching for a new vessel.

To gain the profits of new technology in Blaise 4, one new ship should replace the old fleet. The aim is a new parameter- and menu-controlled system that can handle all relations between CAPI and CATI, suitable to load and carry all survey designs and minimising maintenance. The aimed system includes:

- handling complex survey designs
    - combining different interview modes (Cati, Capi, Casi, Cadi)
    - combining different sample units (households, persons, addresses, license plates)
    - handling data flow from wave to wave in panel surveys

- standardisation
    - consistent procedures for survey supervisors and interviewers.
    - standardised procedures to deliver data to the statistical departments
    - uniform administration for all surveys.
    - compatible data and standardised blocks for non-response treatment

- Version control (data models, questionnaires, external files, Blaise versions) and documentation.

- Windows orientation, optimising use of available ADO components.

A model for a new ship: The body.

This model shows the main flow from sample to data delivery. This model focuses on the essential process in general terms: to keep the survey going. Many sub-processes are reduced to one here.

The main purpose of system is to combine:

1. controlling the information flow between panels and modes and the statistical database
2. keeping the records of the complete person and household survey administration of Statistics Netherlands.

The system keeps track of the survey process between sampling and data delivery and manages the subsequent steps to be taken.

Figure 1. Simple process model of the survey organisation.
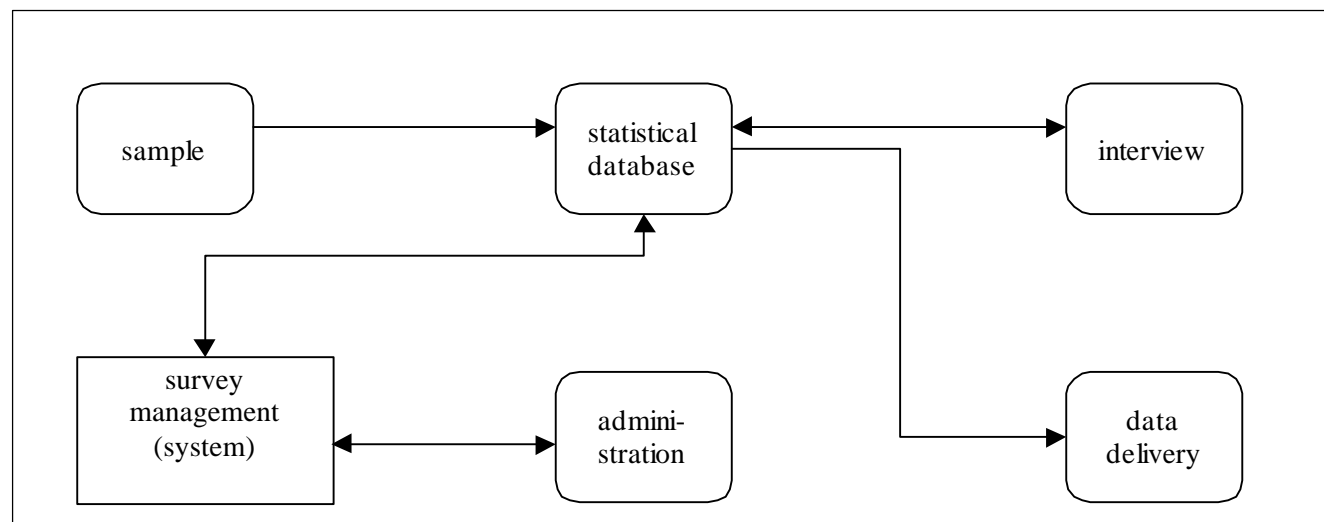
The *sample* is the input of this process model. The output of the of the model is the *data delivery* process.

Data from the sample are directed to a *statistical database*. The database keeps information from the sample and the interviews. The statistical data include keys to link to the administration. Each sample unit that enters the database gets a status. This status is kept and updated in the *administration*.

If records are sent to the *interview* process, the interview will be carried out and the results (completed forms, reports) are updated in the statistical database (data).

The co-ordination of the whole process is done by the *survey management (system)*. The brackets indicate that a part of it may be automated. The survey management system is the wheelbox of the process. For each survey, there is a blueprint of the steps to take for the sample unit before a record it is ready for data delivery. The flow to specific surveys and waves is controlled by these blueprints and the status in the administration.

The survey management system systematically:
1. Checks if there are changes in the statistical database to update the status of records in the administration.
2. Activate sample units for interview or data delivery based on the updated status in the administration.

Based on information from the administration and the statistical database the management determines the next step for each record. This may be the *interview* process or *data delivery*.


The interview process
The process model above is the general process model for the main stream of interviewing data. The decision to activate the interview process is carried out by the survey management system. The organisation of moving between the statistical database and the interview process is more complicated.

Depending on the survey, codes are supplied with the interview record, controlling the route (i.e., to which questionnaire, wave, interview mode, interview period, version control, check on external data, Blaise version). The system will have to allow for different sample units (person, household or others). Until now much of the decisions are done "manually". The new approach is to automate the decision rules as much as possible.

No matter how complicated the decision rules are, they are made in advance. The decisions are taken outside the interview process. The interview process is reduced to IMPORT the records in a standard way, to direct them to the proper interview mode (Cati, Capi or Cadi), do the interview and EXPORT them back to the statistical administration.

Figure 2. Process model of the INTERVIEW process.
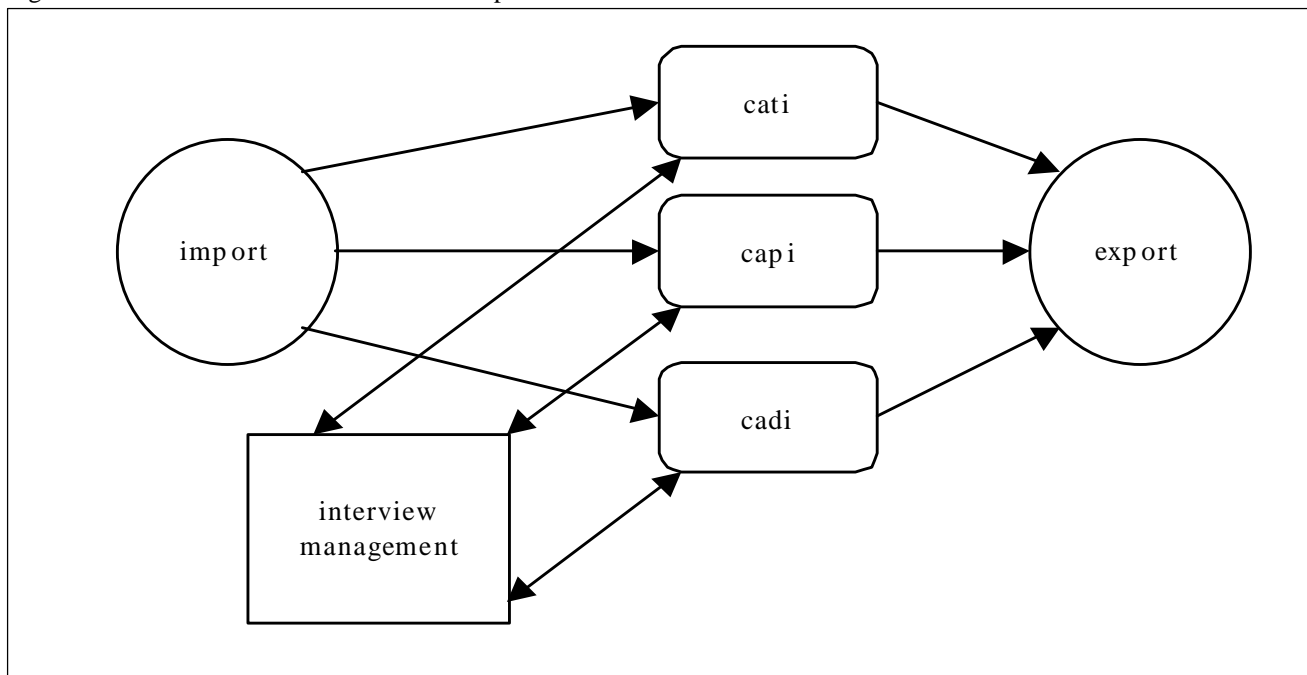
All information exchange with the statistical database passes the IMPORT or the EXPORT module. This ensures uniform data handling. The interview management in the person of a CAPI- or CATI supervisor keeps having an active roll. The supervisor continuously checks if the "interview machine" is working properly. They manage the scheduler for the interviewers. If an interviewer is absent the interview manager takes care for replacement. If there is a problem with the system they report it to the survey management.

Data delivery

The data delivery process in the first model implies two processes: *finishing* and *data delivery*.

Figure 3. Process model of the data delivery process.
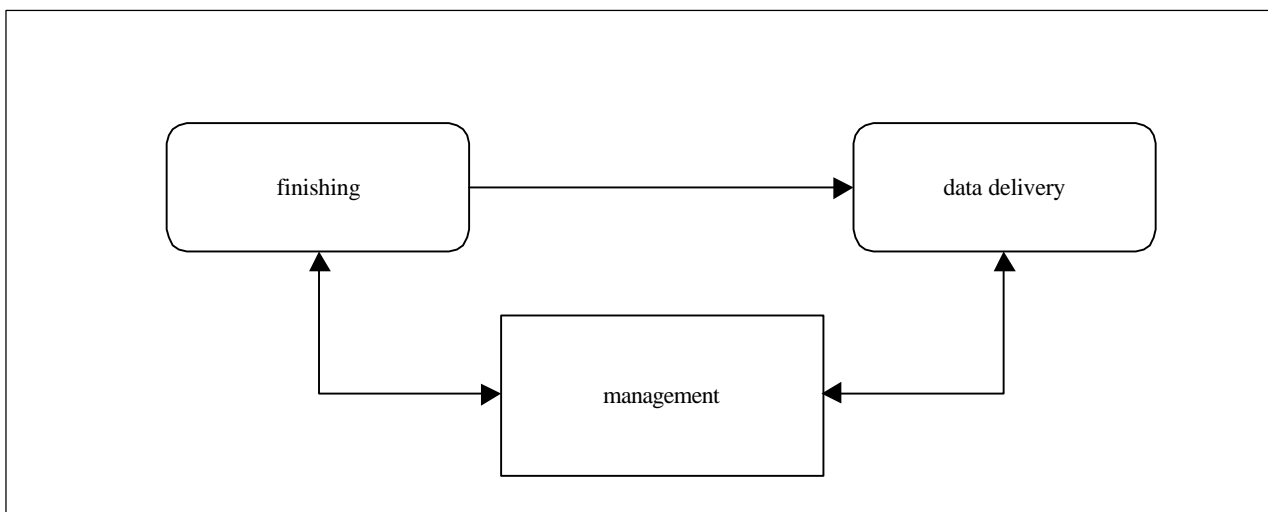
Before data are directed from the statistical database to the client, there is a *finishing process*. The number of finishing jobs and their nature depends on agreement with the client (the statistical departments). This finishing process may include coding jobs (for example education and kind of job), cleaning data, quality control of the data and computation of derivations.

The data delivery does not necessarily have to be reduced to the interview records. This may also imply other information (such as the visit reports, remark files, history files, survey reports).
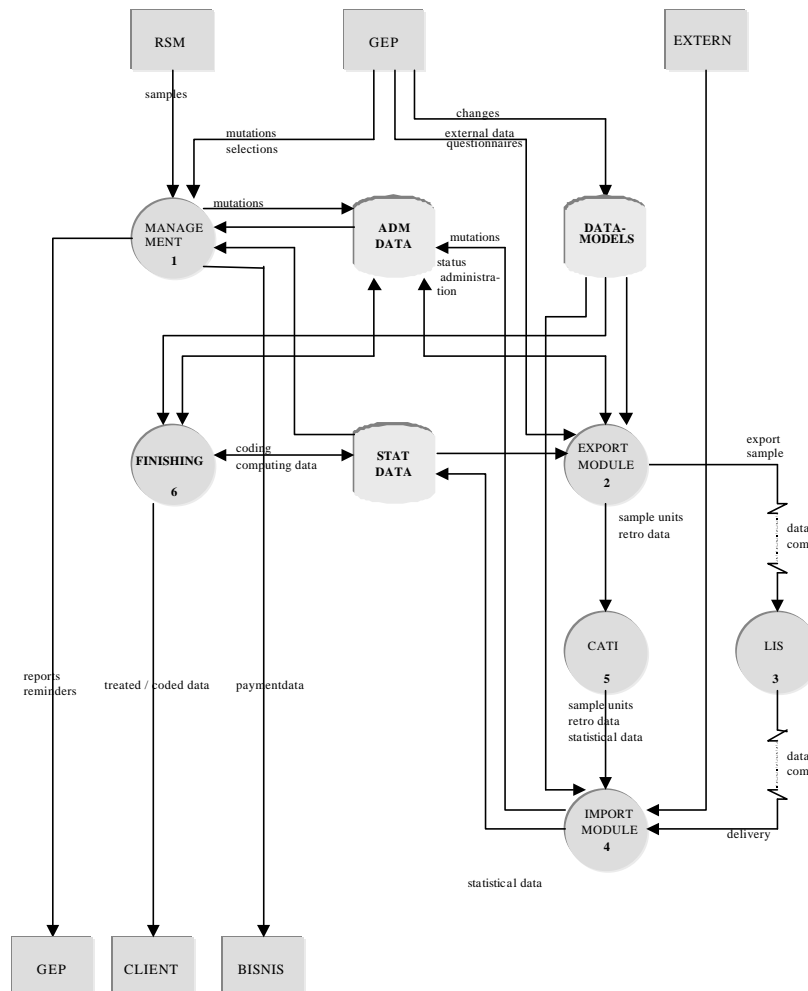
Automated administration

In the previous figures, there is only one line to the administration. The Survey management updates all information in the administration. In fact, the plan is to automate the flow of information to the administration. To reduce the amount of arrows in the figures all of them have been left out.

The Blueprint of the ship

The models above have been put into one complete blueprint. Each survey is supposed to fit in this model.

Figure 4. Draft version of the process model

RSM — GEP — EXTERN

samples

mutations
selections

changes
external data
questionnaires

mutations

MANAGEMENT 1 — ADM DATA — DATA-MODELS

mutations
status
administra-tion

FINISHING 6 — coding computing data — STAT DATA — EXPORT MODULE 2 — export sample

sample units
retro data

CATI 5 — LIS 3

data com

sample units
retro data
statistical data

reports reminders — treated / coded data — paymentdata

IMPORT MODULE 4 — delivery

data com

statistical data

GEP — CLIENT — BISNIS

This process model implies more than the simple models mentioned earlier. There is more detail in this model, nevertheless, this is a draft version.

RSM is the department delivering the input of the system: the sample. This is the only relation of RSM to the system. GEP is the department in charge for survey management. Added in this model is the GEP input in the process, for example, datamodels, mutations on them and mutations on administration data. At the centre of the system the *administration* (SQL-based) and the *statistical database* (Blaise based) can be recognised.

Standardising Import and Export to Capi and Cati Island

On the right side of the figure the IMPORT and EXPORT module link records from the statistical database to the interview process (CATI and CAPI). LIS is the system for CAPI.

Data from interviewers are imported solely by the import module. Similarly, data to transmit to the interviewers must always pass the export module. The EXPORT module takes care for exporting sample data in empty interviews to CAPI and CATI and the IMPORT module takes care for importing completed interview forms into the statistical database and the administration. Using only the modules for import and export ensures that it happens on a standardised way. If for a certain survey the interviewing job is put out to an external enterprise, also the data to and from these EXTERNAL interviewers should pass the import and export modules.

Overall survey management

GEP manages the whole survey process. Part of the system will always ask for manual interventions. If there is a change in any corner of the system, GEP is involved. If the system is out of order GEP organises a repair. For example, when a data communication problem occurs between interviewers and Statistics Netherlands, necessary interventions are controlled by GEP.

The checklist for the ship

A lot of details often are forgotten. This is why a checklist has been made. In fact it concerns too many details to make it an amusing story. Here is just a small part of it.

Survey organisation
- Creating and distributing interview materials (letters, answer cards, panel cards, removal cards)
- Re-distributing returned untreated sample units.
- Reminding interviewers and possible respondents
- Automatic computing analyse fields from values in basic interview fields.

Preparing datacom and Cati
- Preparing and installing the required questionnaires (including Cati system and external files)

Software on the Laptop: LIS, the Laptop Information System enables:
- Importing sample data (unpacking, error handling, log-file sent back to Statistics Netherlands)
- Checking completed interviews for datacom

Planning

Until here is where the project got so far. The plan is still to be finalised. The project aims to start in May 2000 and is planned to be finished at the end of 2001. The current planning of this draft project plan speak about 11 FTE years (not engaged yet).

Credits

Credits go to the project managers and advisers Theo Huijs, Antoon Hogervorst and Frans Kerssemakers for their information about this project.