

Blaise at Statistics Netherlands

Marien Lina, Statistics Netherlands

1. Introduction

Statistics Netherlands developed Blaise as a system for data entry and survey design. Being the producer of the system, Statistics Netherlands is also one of the bigger Blaise users. This contribution focuses on Blaise related applications at Statistics Netherlands that use the Blaise system.

The starting point is Blaise data entry and interviewing. At Statistics Netherlands Blaise has been the standard environment for developing and executing CAPI and CATI interviews for 15 years now. Surveys, carried out by Statistics Netherlands differ in size and complexity of the questionnaire, willingness to respond, accessibility of the sample, the need for personal interaction, respondent burden, response sensitivity and subjectivity of question content. Different surveys ask for a different solution, different data entry types, different ways to address respondents and different techniques, such as applying external data files for typing and coding, or running external procedures while entering data in a Blaise data entry application.

The “traditional” Blaise data entry modes, used at Statistics Netherlands are data entry of paper forms (CADI), Face-to-face interviewing (CAPI) and Telephone interviewing (CATI). They keep playing a significant role in the data entry process at Statistics Netherlands. The first part of this paper will review surveys and sample sizes for various data entry modes.

Recent developments in Blaise offer new technical solutions developed for specific demands. New technology enables new functionality (for example, Internet interviewing) and simplified solutions for complex functionality (for example, combining information from different surveys or interview modes). New Blaise developments were both a challenge and a necessity for carrying out various projects successfully. A number of innovative projects will be reviewed here. The selected themes are related to: *Web interviewing, automated data imputations, mixed mode surveys* and the administration of the survey process of these mixed mode person and household surveys.

This contribution will review:

- “Traditional” data entry and data manipulation: An overview of Blaise interviewing, data entry and Manipula applications at Statistics Netherlands for several person, household and establishment surveys.
- Technological Blaise innovations
- Innovative projects at Statistics Netherlands
- The SSA System for Survey Administration for person and household surveys
- The IMPECT project
- The Electronic Data Reporter for establishment statistics
- Web interviewing in a survey to measure establishment turnover.

2. “Traditional” data entry and data manipulation

At Statistics Netherlands, Blaise has been the standard environment for developing and executing CAPI and CATI interviews for 15 years. In 2003, Statistics Netherlands will perform approximately 175,000 field and 175,000 telephone interviews in Blaise.

2.1 CATI interviewing for person and household surveys

Most of the CATI interviews at Statistics Netherlands are carried out with Blaise. The Labour Force Survey is the largest household survey.

CATI-interviews with Blaise for person and household surveys:

Survey	CATI interviews *)
Labour force survey	122,000
CCO (Consumers prosperity)	18,000
EDISENT reminders	5,000
Price observation petrol	< 1,000
Non-Response analysis survey	8,000
School-leavers panels	8,000
Woonkosten A (house rent survey)	< 1,000
Experiments Budget survey	1,000
Budget Survey first contact	11,000
Total Blaise CATI interviews	175,000
OVG – traffic research	43,000
Environmental costs survey	<1,000
Experimental	6.000
Total Non Blaise CATI interviews	50,000
Total	225,000

*) approximate planned numbers + / - 20 %

Most of the person and household CATI interviews at Statistics Netherlands are conducted with Blaise. An exception is the OVG, using paper booklets to enter moves and means of transportation. The special Neu Kontiv Design that has been used for the OVG survey implied specific experimental methods that asked for a specific approach. Nearly 80% of the person and household CATI surveys use Blaise for data entry.

2.2 CADI interviewing for person and household surveys

CADI interviews with Blaise for persons and households planned for 2003:

Survey	CADI interviews *)
CPI – price observations	117,000

*) approximate planned numbers + / - 20 %

In the price-observation survey data are gathered on paper and data entry is done afterwards in Blaise. The questionnaire is very small and it concerns just the prices of a number of goods in shops. In fact the survey population are goods here and not persons or households. Data entry is carried out by the same organisational group as for person and household surveys.

2.3 Blaise CAPI interviews for persons and households

Face to face Blaise interviews - persons and households. Laptop / 2003:

Survey	CAPI interviews *)
EBB (labour force survey)	96,000
Gezinsvorming (Survey Family Settling)	16,000
POLS – health survey	17,800
POLS – living conditions and behaviour	8,900
POLS - legal protection and security	8,900
POLS - the young	7,100
ECP - Eurostat corrections	1,000
Budget Survey – first contact	5,000
Budget Survey 2003	2,400
Budget Survey 2004	14,200
Experiments	400
Total Blaise surveys	175,700

*) approximate planned numbers + / - 20 %

Blaise data entry is used for every field survey among persons and households.

2.4 Blaise interviewing and data entry in establishment statistics

Compared to person and household questionnaires, establishment statistics use more secondary data and the surveys usually are much smaller. Some of the establishment statistics do not depend entirely on data entry, but they merely are in need of additional figures to adjust the data readily available, when the bulk of the information is retrieved through other channels.

The following overview gives numbers of data entry forms, entered yearly in Blaise. Some data collection activities may not be listed as there is no central administration of these surveys.

Data entry records per year – establishment statistics.

Statistical topic	Forms entered per year *)
Impect Ps (long list)	80,000
Impect Ks (short list - additional questions)	400,000
International trade **)	600,000
Financial enterprise statistics	40,000
Bus transportation	50,000
Road transport (freight)	300,000
Canal shipping trade	60,000
Sea shipping trade	10,000
Fire Departments statistics	500

*) approximate numbers + / - 20 %

***) the 600,000 Blaise records are additional to 74,000,000 from secondary data.

The table shows that the number of records entered with Blaise is about 1,5 million each year, about 5 times more than the 350,000 person and household interviews.

One should keep in mind that data collection for establishments is usually data entry for a limited number of data fields, and the time needed to fill in a questionnaire or data form is limited.

There are more establishment statistics and some of them don't use Blaise data entry. For example, data entry of about 280,000 forms for a survey on job vacancies and absence through illness is done with other data entry software. There may be more surveys of this kind that have very small data records. They may be retrieved from and/or combined with specific database formats, not related to Blaise.

Apart from the mentioned figures, there are many records entered in Blaise by establishments outside Statistics Netherlands. In the Iris project, this concerns about 21 million each year.

The tables are not complete (especially the one with the data concerning establishment surveys) but the most extensive surveys are listed here.

2.5 Data manipulation at Statistics Netherlands

Apart from data entry, another "traditional" Blaise functionality is data manipulation. The relevant software parts in Blaise are Manipula, Maniplus and Cameleon. The three tools are applied to execute procedures for cleaning data, typing and coding, file management and preparing data for analysis in other software applications.

The number of calls per year of the Manipula program at different parts of Statistics Netherlands is more than a million. More than 40% of the Manipula calls are relatively "old fashioned" and designed to process ASCII files. More and more of the old systems are being replaced by new technology using new Blaise 4 functionality or Blaise Component Pack applications. COM technology ensures communication options between Blaise and other applications.

The Blaise Component Pack may carry out Manipula-like procedures without activating Manipula. The million calls to Manipula exclude the calls to VB (including Blaise Component Pack) for similar purposes. Some of these (in fact the applications in the SSA project) replace previous applications that used Maniplus and Manipula setups.

2.6 General conclusions on traditional data entry and data manipulation

The practical usage of Blaise for data entry at Statistics Netherlands, derived from the mentioned figures above illustrates the strong quality of the Blaise system for data entry and manipulation.

The choice for Blaise as a data entry machine is obvious when questionnaires and data structures are larger and more complex. For very complex data entry machines like the Labour Force Survey and POLS (life situation survey) using Blaise has become the de facto standard at Statistics Netherlands.

Smaller data records are found at establishment statistics. Statistics Netherlands has a standardisation policy for data collection methods. This policy encourages the use and development of Blaise data entry machines. The aimed practical merit of such standards is increasing exchangeability of data in the future.

3. Technological Blaise innovations

At Statistics Netherlands, Blaise plays and keeps playing a significant role in the survey process, including applications for data entry, typing, coding, manipulating data and preparing files for analysis. The "traditional" functionality is very useful to perform these tasks.

3.1 Blaise developments “the old way”

Over the years, new functionality has been added to the Blaise system in numerous tools and utilities. For example, at the beginning of Blaise developments in the late 80s, the CATI system has been added. Over time it has been extended and developed as a stable toolbox with “out-of-the-box” standard procedures. The CATI modules allow for specifying many parameters in standard dialogues and parameter files. Yet it remains a standard tool you may use “as-is”. As a result, the CATI management system works fine for all standard CATI interviewing. One can imagine, special requirements ask for tailor-made adaptations on-site. Specific non-standard functionality may be added with the programming tools Manipula, Maniplus and Cameleon. Functionality has been added to Manipula to increase control on the CATI management system (for example the `daybatchadd` keyword). This illustrates how Manipula, Maniplus and Cameleon enable creating tailor-made applications around Blaise. The building of applications in these languages requires specific programming skills. With the programming facilities of Manipula, Maniplus and Cameleon, meta and data manipulations are handled separately (meta information via Cameleon and data via Manipula or Maniplus).

Different institutions and different survey types using Blaise asked for specific features not implemented in the standard tools. As there are many users, this called for more flexible tools, enabling Blaise users to create “home-made” applications to meet their own special needs.

3.2 Blaise Component Pack

Blaise Component Pack (BCP) 2.0 was necessary to carry out the projects that will be reviewed below. The projects in turn have been a driving force for developing and implementing specific required functionality in BCP 2.0.

New Blaise Component Pack technology opens a flexible way to create Blaise and Blaise related applications. BCP did not only increase flexibility for user-made additions to the system, it has also been used for developing new standard tools by the Blaise development team. Among others, the BCP technology has been used to develop the Blaise Internet Services application (Blaise IS).

Some elements of this new approach with BCP are:

- Creation of VB applications that can be activated from within Blaise data entry applications or Manipula jobs (alien procedures and alien blocks).
- Using the OLEDB approach and data exchange via BOI file enables easy communication between the Blaise data format and other data formats.
- BCP enables the creation of Visual Basic applications that read and write Blaise data. VB applications that run outside the Blaise environment can be developed to include and combine DEP, Manipula, Maniplus and Cameleon functionality.

Both developments are continuing:

- adding standard tools to the existing Blaise system
- developing new technical options in Blaise Component Pack 2.0.

4. Innovative projects at Statistics Netherlands

New Blaise technology has been used in a number of projects at Statistics Netherlands. Some use BCP technology, others use new Blaise 4.6 functionality.

For the person and household surveys, the need for controlling and administrating the survey process (on a record level) has resulted in the SSA system. This System

for Survey Administration implements the control and organisation of subsequent steps for processed records in person and household surveys. It combines many surveys in one database and controls data flow. The system is built with VB tools that use Blaise Component Pack technology, especially the Manipula functionality.

For establishment statistics the IMPECT project at Statistics Netherlands combines Blaise with VB and SQL-Server applications. The global outline of IMPECT will be reviewed below. The IMPECT project has a number of sub-projects, each with their own functionality and relations with Blaise, some of which will be treated here, such as LogiQuest, controlling the organisation of data collection.

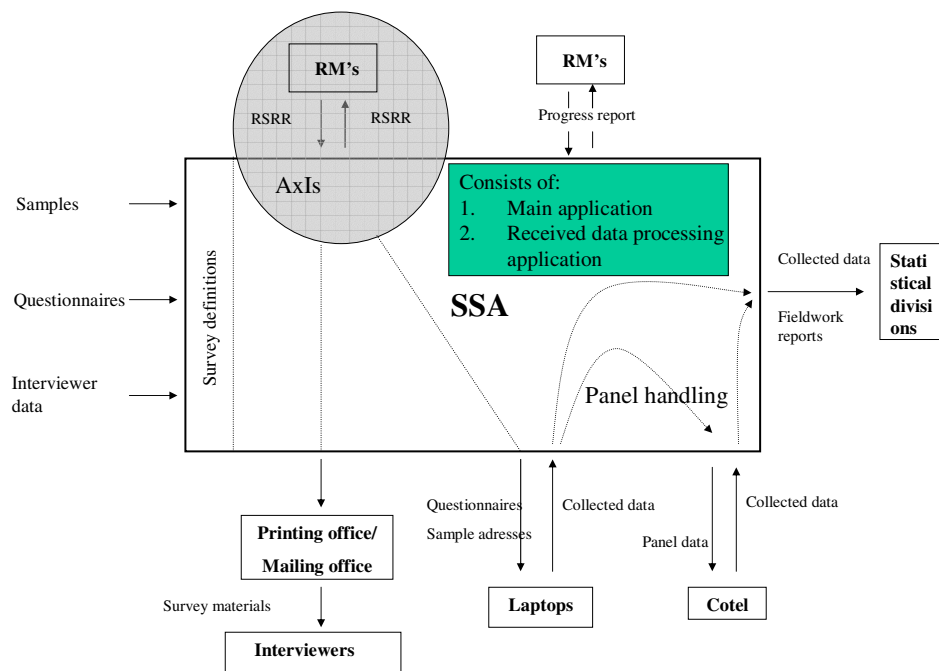
Data may be collected in two new ways: by e-mail or by Web-interviewing. The Electronic Data Reporter organises data collection via e-mail and will be reviewed briefly.

One statistic within the IMPECT project measures yearly turnover figures. Data collection for this statistic is executed with Web interviewing.

5. The SSA administration for person and household surveys

In the past, administrations have been a mix of Oracle, Paradox, Blaise and other applications. The putting out of interviews to laptops of field interviewers and the CATI server at Statistics Netherlands has been a process that needed a lot of manual actions.

The SSA (System for Survey Administration) project puts an end to the multitude of administrations and manual actions. The two biggest surveys: POLS (life situation survey) and EBB (labour force survey) are CAPI surveys that have a follow up with CATI. These and other surveys have been accommodated in SSA.

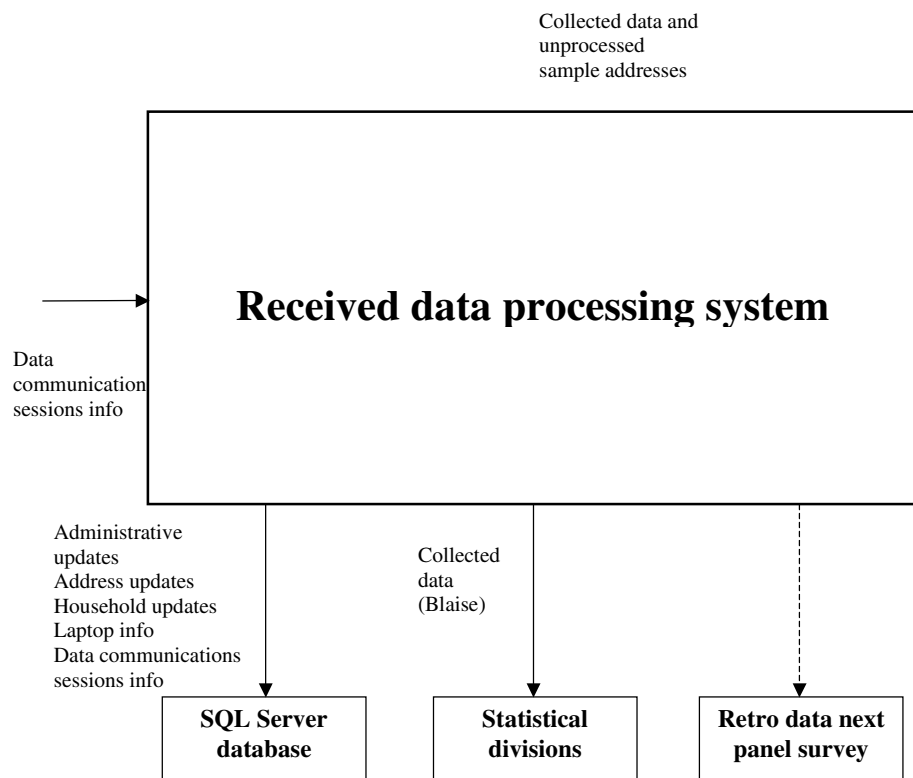


The previous blueprint of SSA shows how it is organised. The survey design of every survey is included in SSA.

The system input consists of the sample (elements), the questionnaire and the interviewer data. Based on the survey design, the system determines when a record is to be activated. It does not only cover the organisation of activating a record, it also executes the sending of interview materials to the interviewers. It loads data on the laptop, collects and processes information that comes back from laptops in the field, creates a regional overview about carrying out interviews, generates fieldwork progress reports and organises that interview forms are guided from CAPI to CATI. After finishing data entry, records are guided to classification procedures (if needed).

The SSA functionality to put out sample elements to interviewers (CAPI and CATI) is written in VB, using Blaise Component Pack (BCP version 1). Completed interview records are guided to and stored in the central SQL-Server database.

When data are received from an interviewer there is a complex process that generates parallel actions on various levels, updating information in the SQL-Server database, producing data for statistical analysis and passing on records to additional panel waves or coding machines.



At this moment, most person and households surveys are part of the SSA system. Statistics Netherlands is still working on accommodating all person and household interviews in SSA.

The system controls the flow of CAPI and CATI interviews from one wave to another. Redirecting cases to systems for coding and typing is included in this system.

6. New developments for establishment statistics

The traditional way of collecting establishment statistics used to be by sending out paper forms, receiving the filled forms from the establishments and then entering the data in a Blaise or relational database, depending on the complexity of the data model. Blaise offers new additional methods for entering data by the establishments.

Software has been developed to use Blaise data entry on-site at the establishments. Establishments use EDR to enter the data 'at home' in electronic questionnaires and send them in by e-mail to Statistics Netherlands. The recently developed **EDR** (Electronic Data Reporter) replaces *EDISENT*. It uses some of the new Blaise 4.6 technology and is operational since the beginning of 2003.

Another data entry mode that is expected to be helpful is Blaise IS. Blaise Internet Services is in use now for collecting data on behalf of turnover statistics of establishments.

There are a lot of different establishment statistics. The IMPECT project is the integral project that aims to integrate these projects into one framework. To achieve the goals of IMPECT, Web-interviewing and using the Electronic Data Reporter for establishment statistics is on the wish list. Before focusing on EDR and Web interviewing, a short introduction of IMPECT might be helpful.

6.1 The IMPECT project

The full name of the IMPECT project is: "IMPlimentation of EConomical Transformation process". It combines a large number of establishment statistics and the tasks that are performed. IMPECT is a program for a number of projects. Some of the main goals are to speed up the whole process of making statistics and lowering the respondent burden.

6.1.1 LogiQuest

The sub-project LogiQuest is the central data collection system. It generates paper questionnaires and electronic contact items. It generates data entry tools and organises data-entry and statement selection.

LogiQuest connects to a *Central Observation Database*. The *Contact Registration* of the LogiQuest system keeps track of all contacts with establishments. Whenever there is a telephone contact with an establishment, the caller can have a view on all relevant previous contacts.

LogiQuest generates establishment questionnaires in various modes. Information requests can be sent out either on paper, as an electronic Blaise questionnaire, or as a Blaise Web interview (for the time being, an HTML-form only).

LogiQuest manages the questionnaire server, sending the information request to the establishments and handling the returned questionnaires. These actions use new Blaise technology to generate questionnaires electronically.

6.1.2 E-mail and Web interviewing

A number of projects deal with data entry modes for IMPECT: The Electronic Data Reporter replaces *EDISENT*. It enables suppliers of information to send in data with e-mail facilities. Another sub-project uses Blaise HTML forms for data entry. The current off-line HTML version is realised with Blaise IS and is in production now. It enables suppliers of information to fill out a form "at home". Once the form

is completed, it can be submitted using an internet connection. An on-line version of a questionnaire is not yet operational in production.

6.1.3 UniEdit

A second IMPECT project called UniEdit has been realised to automate data checking and cleaning. It manages error detection in data, automated data corrections, automated editing and (if other methods failed) manual editing. Suspect data may be imputed on estimations or may send a signal to replace the data at the source. Automated rules checking of data, cleaning up data and imputing missing or inconsistent data save time, replacing manual procedures.

In this process of data checks and imputations, Blaise comes in on many stages: for statement selection, error detection and correction, automated editing and interactive editing, analysis, and not in the last place for generating new Blaise source for sub-data models, based on selective fields of existing data models. The *SLICE* tool checks on irrational data combinations and corrects them automatically (if possible). It is written in C++ and uses Blaise Component Pack to link to the Blaise Data and Blaise rules. If the automatic corrections are not successful, manual corrections are carried out with Blaise data editing machines.

6.1.4 The analysis project of IMPECT

Analysing establishment data is the last step before publishing statistics. However, it is an illusion to think that analysing is a straightforward separate process *after* completing data collection, data checks and corrections. The analysis is an ongoing process that starts as soon as the first results of data collection are available. The analysis results show stepwise advancements, based on intermediate results of the last response and the last corrections that have been made in the data. Statistical corrections of the data involve improved methods for applying imputations (automatically *and* manually), weighing, detecting outliers and checking them with external data such as value added tax data from the tax office. A part of the checks and corrections are done with Blaise related tools. *Bascula* is the Blaise related tool for weighing. The acceptance uses SPSS procedures. After correcting, data are transformed into statistical variables.

The required speed for accurate production of statistics asks for this approach in which intermediate results are available in “stand-by” mode. Finally, if the data are accepted, the “final touch” is put and the data are ready for statistical publication.

6.2 EDR: the Electronic Data Reporter

More and more, data collection makes a shift from paper questionnaires to the Electronic Data Reporter and Internet interviewing. With new Blaise technology these methods have been applied successfully. Statistics Netherlands is also experimenting with touch-tone data entry for establishments.

The Electronic Data Reporter (EDR) is an advanced data collection module, and is mainly used for establishment statistics. The module uses new Blaise 4.6 technology. Except for the launcher (a small C++ program), the EDR module has been developed completely in Blaise. It is operational at Statistics Netherlands since the beginning of 2003.

The Electronic Data Reporter has been developed to replace the EDISENT module. EDISENT had a number of restrictions, it was build for a 16-bits environment and therefore, it is difficult to maintain and difficult to extend. EDR has been set up in a more flexible way. EDR applies the Blaise language for questionnaire definitions. This increases the flexibility of questionnaire design. The Electronic Data Reporter must be installed on-site at the establishment (or other supplier of information).

The installation can be delivered on a CD-ROM and includes the Blaise run-time engine for data entry.

The software can be extended and updated by e-mail. Different statistics may use the same Electronic Data Reporter. The questionnaires can use different code lists e.g. for looking up answer values.

As code lists can be shared, they will be stored only once, also when they are used in more surveys. This reduces the claim on the disk space of the supplier of information when a code list is large. Tests with code lists containing over 35.000 elements did not affect the performance in any way.

Every questionnaire may now contain its own hierarchy of keys ten levels deep. All kinds of key types (predefined using a lookup or just open) are supported. Predefined values for keys can be maintained by the data collector when sending questionnaires. Routing and checking is available for open keys. The Electronic Data Reporter offers complete statement management including duplicates.

6.3 An establishment statistic featuring Blaise IS

One of the survey themes for economical establishment statistics is about the yearly turnover of establishment code groups. The survey used to be carried out with paper and pencil forms sent to the establishments. The establishments included in the sample used to be addressed by ground mail, if needed with reminders by telephone.

New Blaise technology made it possible to experiment with Web interviewing. This is done in the IMPECT project with Blaise IS 1.1. Establishments are asked to supply information for statistics on volume of trade. Compared to most of the person and household studies, these questionnaires are very small, about 5 or 6 questions, and in some cases the questionnaire may exist of one question only, in simplified terms: "what was your turnover in 2003".

If an establishment does not reply on the first electronic request to fill in a form, 2 reminders may follow to win them over to participate in the survey and become a member of the responding population. If this is not successful, then the specific cases are handed over to the survey management and they might use "old-fashioned" methods to address the establishment in another way to get the data, or they may decide to ignore the non-respondent and impute values based on estimations. This depends on the amount of non-response for establishments in the same code-group and the weight of the expected turnover of the specific establishment.

Blaise IS is not to be regarded as being the replacing methodology, but merely an addition to data collection by mail, telephone and e-mail. It takes away a lot of reminding and paper work. With the Blaise IS method, the response comes in faster than when using ground mail and telephone only. After one week, 30% of the establishments responded and response at the end became 70 to 75%.

6.4 Conclusion on Blaise IS and EDR for establishment statistics

It is hoped that these new methods decrease the total amount of time needed to produce statistical figures. For specific establishments, the paper and pencil method remains a popular means of passing on information. Reminders are still necessary to get the required amount of response and only time can tell what response rates the new Web and e-mail way of data entry will have in the future.

So far, the traditional CAPI, CATI and CADI data entry of paper forms are being used as long as a certain proportion of response is not coming in via EDR or Blaise IS applications.

Nevertheless, the basic conclusion on EDR and Blaise IS is that more and more establishments are moving away from the old paper and pencil method and that they welcome EDR and Blaise IS for data entry. The merit for Statistics Netherlands is that this decreases the amount of data entry, paper work and correspondence and it may help to have response in a smaller amount of time.

7. Summary and Conclusion

At Statistics Netherlands, Blaise is the standard tool for data entry. Yearly, over a million and a half data records are entered with Blaise CAPI, CADI and CATI machines.

New Blaise technology has been used successfully to integrate surveys for person and households into one system. The Blaise Component Pack made it possible to construct the System for Survey Administration in VB, connecting to Blaise meta and data. This system substantially reduced the amount of manual work in the organisation of the person and household surveys.

Data manipulations to prepare data for analysis are performed a lot with Blaise tools like Manipula and Maniplus (there are about 1 million calls to Manipula each year). These Manipula procedures may be replaced with BCP applications in the future.

New technology and new functionality of Blaise 4.6 has been used successfully in the IMPECT project and other projects to achieve a more efficient production of establishment statistics.

Relevant data entry applications here are the Electronic Data Reporter as well as Internet applications created with Blaise IS. Both methods (e-mail and Web interviewing) are used to collect data for short term establishment statistics.

Data corrections, checking plausibility, weighing, checking for outliers, and many other things are arranged in Blaise applications, Bascula and the SLICE application that uses BCP technology in a C++ application.

The options to combine Blaise applications with others, like C++, SQL-Server and SPSS are increased with Blaise Component Pack. The Blaise software with its specific characteristics for statistical surveys, its data model properties, data entry machines and dynamic routing and checking mechanism keeps playing an important role at Statistics Netherlands.

