

# The Co-existence of Blaise 4 and Blaise 5 in CYSTAT

*Charalambos Charalambous, Charoula Charalambous, Niki Chrysostomou, Costas Diamantides and Konstantinos Mina, Statistical Service of Cyprus*

## 1. Abstract

CYSTAT has a long history in the use of Blaise. The first electronic questionnaires were developed in 1995 and since then, Blaise has become the main tool for data capture. During the first years of the use of Blaise in CYSTAT, several employees from different production units participated in the trainings offered by Statistics Netherlands. In this way, there were no centralized procedures for development and consequently the sharing of knowledge was lagging. In 2011 it was decided to establish the Blaise team of CYSTAT with the task for the development of all the applications with Blaise. Nowadays, the Blaise team of CYSTAT is part of the Methodology Section and comprises of one developer.

Despite the limited human resources several applications are being developed and supported by the Blaise team. Blaise 5 was recently introduced in CYSTAT and it is now used in two surveys, the ICT Usage in Enterprises and e-commerce (ICT-ENT) and in Employment and Job Vacancies. The data collection in both surveys is multi-mode and the data collection tools are developed with Blaise 4 and Blaise 5. Another example of a major survey incorporating multi-mode data collection is the Income and Living Conditions (EU-SILC). In this case, however, all the data collection tools are developed in Blaise 4 and considering all its advantages the aim is to convert to Blaise 5. The EU-SILC is the only survey that is not being developed by the Blaise team.

The aim of the paper is to present the experiences of CYSTAT in multi-mode data collection by using both versions 4 and 5 and to share the future plans in using Blaise 5.

## 2. Background

This paper examines the use of Blaise in three surveys, the ICT-ENT, the Employment and Job Vacancies and the EU-SILC. This section provides all the background information for all three surveys.

### 2.1 ICT Usage in Enterprises and e-commerce (ICT-ENT)

The ICT-ENT is being carried out annually since 2004 by applying the method of Computer Assisted Personal Interviewing (CAPI) implemented in Blaise. The aim of the survey is to collect data about the Information and Communication Technologies (ICT) used by the Enterprises. The major topics covered are: use of computers, access and use of the Internet, employment of ICT specialists and their skills, ICT security and E-commerce. The data collected through the survey are necessary for the implementation of policy programmes at both the European and National level.

The responsibility for carrying out all the steps of the survey – from the design to the dissemination of the results - lies on the ICT Section of CYSTAT with the support of the Blaise team. The methodological manual, the model questionnaire as well as the validation and transmission tools are all provided by Eurostat.

The data collection takes place during January - May and covers approximately 2.000 enterprises with 10 or more employees in almost all categories of economic activities. The method used for the sample selection is stratified random sampling. Two variables are used for the stratification, the economic activity

(16 groups) and size (3 size groups: Small enterprises (10-49 employees), Medium enterprises (50-249 employees) and Large enterprises (250+ employees)). The survey covers all the government controlled areas of the Republic of Cyprus and the reference period of the survey is the current year unless otherwise stated.

Since 2017, the data collection is being carried out either through a web questionnaire (CAWI - Blaise 5) or personal interviewing with the IT manager of the enterprise (CAPI - Blaise 4.8). The web questionnaire is available during the first phase of the survey period (first 4 - 5 weeks). If an enterprise fails to fill in the web questionnaire in the first phase then it is obliged to provide the data through the personal interview in the second phase. The duration of the second phase is up to 2 months. It is important to note that during the second phase the web questionnaire remains to be available online.

As soon as the data collection is completed, all the datasets (from both CAWI and CAPI) are joined by using Manipula and exported in ASCII format. Then the complete dataset is imported in SPSS for editing. As soon as the file with the clean data is ready all the necessary output is produced.

## **2.2 Employment and Job Vacancies**

The Employment and Job Vacancies survey is carried out on a quarterly basis. All the enterprises with 20 and more employees are covered and for those with 1-19 employees a representative sample is used. The survey covers the vast majority of economic activity categories.

The sampling frame used for the sample selection is drawn from the Business Register which contains all enterprises as well as their local units that carry out any economic activity irrespective of their size. A stratified sampling technique is used as sampling method. The strata are defined by the cross-classification of economic activity groups with size classes (enterprise with no more than 1 employee, enterprises with 1,5 - 19,5 employees and enterprises with more than 19,5 employees). All enterprises employing more than 19,5 employees are selected for the survey whereas for the other size categories, the sample within each stratum is selected using simple random sample. The sample size in categories with no more than 19,5 employees is based on the number of enterprises in each category in the population. For the case of public sector and the publicly owned enterprises, those are fully covered. The same sample is used every quarter and the sample size is about 3300 enterprises.

The survey collects the monthly employment levels for each cell of a matrix determined by: 1) full-time/part-time employment and 2) status of employment (working proprietors, unpaid family workers, employees). It is noted that working proprietors are included also in cases where they are self-employed and the only labour force of their enterprise. Part-time employment is defined as those persons working less than 30 hours in the reference week. In addition, the questionnaire collects the number of hours worked by part-time workers as well as an indication of the normal hours worked by a person in full-time employment. It also collects information on Job Vacancies like the number and the description of the position.

The data collection lasts for 3 weeks and it is carried out either by Computer Assisted Telephone Interviewing (CATI Module – Blaise 4) or through a web questionnaire (CAWI – Blaise 5). Both options run concurrently and there is a mechanism put in place which ensures the unique collection of data from each enterprise. As soon as the data collection is completed, all the datasets (from both CATI and CAWI) are joined by using Manipula script and exported in ASCII format. Then the complete dataset is imported in SPSS for editing. As soon as the file with the clean data is ready all the necessary output is produced.

## **2.3 Income and Living Conditions (EU-SILC)**

The Income and Living Conditions (EU-SILC) is an annual cross-sectional and longitudinal household sample survey, based on European Commission Regulation and is the main source for data and indicators on income, poverty, social exclusion and living conditions in the European Union. Cross-sectional data pertain to fixed time periods, with variables on income, poverty, social exclusion and living conditions, while longitudinal data pertain to changes over time, usually observed over four years. The survey was first launched by CYPSTAT in 2005.

The questionnaire of the survey is designed every year according to Eurostat's guidelines. The guidelines concern the methodology including the definitions of all the variables (both primary and secondary).

The sample is based on a rotational design of 4 replications with a rotation of one replication per year. This design is appropriate to serve the cross-sectional and longitudinal aspects of the survey. Each year the sub-sample that is in the survey for 4 years is dropped and a new sub-sample is selected by one-stage stratified simple random sampling. The sample selection is conducted by the Methodological Unit of CYPSTAT. The total sample size (all 4 sub-samples) is around 5.000 households and complies with the effective sample size requirements for the cross-sectional and longitudinal components set by the Commission regulation. The measurement units are the households and their members.

The main mode of data collection is CAPI using BLAISE 4. From 2013 onwards, CYPSTAT combines both CAPI and telephone interviewing. Telephone interviewing is used only for a sub-sample of the survey, that is Nicosia's households which are in the survey for 3 or 4 years (near the 20% of total sample). The data capture during the telephone interviewing is carried out on the same electronic questionnaires as in CAPI and thus, the CATI module is not being used.

An addition to the data collection process, from 2013 onwards, was the access to administrative income data from various Government departments such as the Grants and Benefits Service, the Department of Social Insurance Services, as well as the Treasury of the Republic. The use of administrative records has proven to be very effective. It serves two purposes: accuracy and reduction of the burden on the respondent. By using administrative records, it was possible to check the quality of the data collected by the interviewers and also to record income data in the cases where this was not provided by the respondents.

As soon as the data collection is completed, all the datasets (from both CAPI and telephone interviewing) are joined by using Manipula script and exported in ASCII format. Then, the complete dataset is imported in SAS for editing. As soon as the file with the clean data is ready, all the necessary output is produced.

## **3. Use of Blaise in CYPSTAT**

This section contains descriptions of the processes that are applied in the use of Blaise in the three surveys with the aim first to present how Blaise 4 and 5 coexist for the purposes of the same survey and second to identify the problems encountered.

### **3.1 CAPI - CAWI**

As described in 2.1 the ICT-ENT survey has been carried out annually by CYPSTAT since 2004. The data collection had always been carried out by applying the method of CAPI implemented in Blaise. In 2015, a web questionnaire was developed and pilot tested for the first time in Blaise 5.

Although a member of our staff participated in a Blaise 5 training course organised by the Blaise Team of Statistics Netherlands the knowledge gained in Blaise 5 was insufficient for the implementation of a web questionnaire. In the framework of the long standing cooperation of CYPSTAT and Statistics Netherlands the Blaise Team provided assistance to implement the first web questionnaire according to CYPSTAT’s requirements. Moreover the Blaise Team provided guidelines on how to make all the installations on the web server.

As a result from the transfer of knowledge, from 2016 onwards the web questionnaires are developed by CYPSTAT. In 2016, a second but in a larger scale pilot was carried out. Since 2017, the option for completing the web questionnaire is being offered to all enterprises which are included in the sample. The web questionnaire option is available to all enterprises during the first phase of the data collection survey. At the second stage, for those enterprises that do not fill in the online questionnaire a personal interview is carried out and the questionnaires developed in Blaise are filled in on netbooks/ laptops. When the second phase of data collection begins the respondents still have the option to fill in the web questionnaire. So at this phase both CAWI (Blaise 5) and CAPI (Blaise 4.8) are running.

In the 2019 survey, the web questionnaire was developed in Blaise 5.6.1. No problems were encountered in the transition from previous versions. There were some difficulties regarding the compatibility of the web layout after the upgrade that have been resolved. The CAPI version was developed in Blaise 4.8.

Until the full implementation of the survey in Blaise 5 there are procedures put in place for the co-existence of Blaise 4 and 5. In particular, the responsible Statistics Officer of the survey has the daily task to export all the questionnaires that were submitted (or started) through the web. Then these data are converted by running a Manipula script and imported in the Blaise database built in Blaise 4.8 (the same is used in CAPI). In order to avoid double counting the Statistics Officer removes the questionnaires completed online from the workload of the enumerators. The editing is carried out in all the questionnaires included in the database which is built on Blaise 4.8. The editing stage includes the import of data for those surveys that were also included in the samples from previous years. Figure 1: Processes in CAWI displays all the major processes concerning the web questionnaire and Figure 2 the processes to combine CAWI and CAPI.

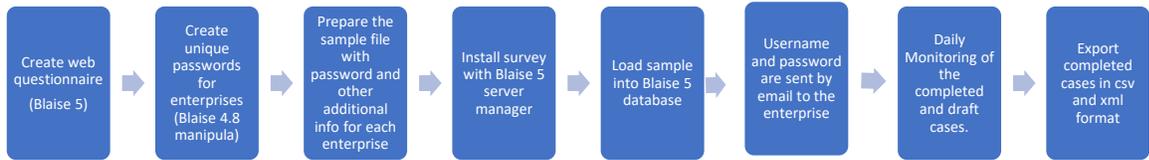


Figure 1: Processes in CAWI

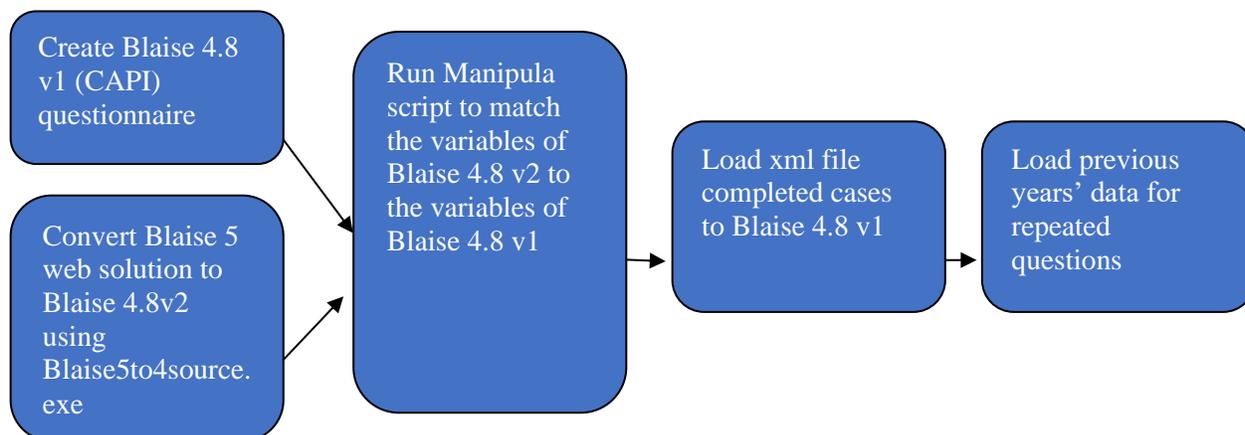


Figure 2: Processes to combine CAWI and CAPI

The process described above for converting from Blaise 5 to 4 is time consuming and prone to errors. This period is considered, however, as transitional to the full implementation of Blaise 5 and the planned steps are described in Chapter 4.

### 3.2 CATI - CAWI

Employment and Job Vacancies is a quarterly survey in which, traditionally, the data has been collected through CATI. Since the third quarter of 2016 the Blaise 4.8 CATI module has been utilised for data collection and management of the survey.

In the framework of the implementation of CYSTAT’s strategy to offer respondents alternative means to provide the required information, the CAWI option built on Blaise 5.2.5 was offered for the first time in the second quarter of 2018. This web questionnaire was developed by CYSTAT based on the knowledge gained from the ICT\_ENT implementation.

The Blaise CATI module offers the opportunity to the operator to arrange an appointment with the enterprise in order to fill in the questionnaire at a later stage, to complete the questionnaire for any other reason (e.g. closed enterprise, merged with other enterprise etc.). The questionnaires that are completed and there is no need for any further treatment are considered as “Completed” and are not active in the day batch. Touched questionnaires that are not completed are considered as “Pending” and appear in the day batch file as they are routed. All completed cases collected through web are imported every morning before the creation of day batch file in CATI Blaise program and are routed back at the end of the day to the interviewers to check the questionnaire and proceed with coding.

Figure 3 describe the processes that are related to the CATI implementation. It is noted that the processes to join CATI and CAWI are the same as described in Figure 2 with the exception that instead of CAPI in this case we have CATI. Furthermore, there is no need for the last step (load previous years’ data for repeated questions) in CATI since this step is already done at the phase of the creation of the Blaise 4.8 CATI questionnaire.



Figure 3: Processes in CAWI

Similar to the case of CAPI and CAWI it is envisaged to have the CATI module in Blaise 5 in order to be able to work in a homogeneous environment and thus, avoiding having the time consuming conversion and updating procedures.

### 3.3 CAPI

The main mode of data collection in the EU-SILC survey is CAPI using BLAISE 4. From 2013 onwards, CYPSTAT combines both CAPI and telephone interviews. The telephone interviews are not implemented with the Blaise CATI module. Instead the operator type in the responses during the telephone interview to the questionnaire developed in Blaise 4. Telephone interview are used only for a sub-sample of the survey, that is Nicosia's households which are in the survey for 3 or 4 years (near the 20% of total sample).

The 75% of the sample are HH that are in the survey for the 2nd, 3rd or 4th time; hence a lot of information from the previous year's survey is preloaded. Some of this information is visible during the interview and some is accessible only during the coding and editing by the supervisors. For that reason, 2 ASCII files are exported from the previous year's survey and they are modified and merged with the new sub-sample location information in order to import them to the new year's electronic questionnaire.

Implementing BLAISE provides the researcher with a powerful tool for applying validation and consistency checks on the field and it also helps enormously in the correct routing of the long and complicated questionnaire of the survey. Additionally, by collecting data with BLAISE, the process of after the data collection data entry is abandoned, making the procedures more cost effective.

Every week during fieldwork, at supervisor – interviewer meeting, the completed questionnaires are transferred to the supervisor for coding and editing. During this phase, the supervisors are calling back to the households for clarifications and any inconsistencies found. Moreover, some reports are generated using Blaise to monitor the data collection, i.e. the number of completed questionnaires, the number of not located households and the response rate. The IDs needed for extracting the information from the registers are also exported.

At the end of the editing, all the supervisors' databases are merged and form the complete database for the year. Then, 11 ASCII files are created and imported to MS-Excel for cleaning by the Statistics Officers. Finally, SAS is used for the final data checks, the calculations of the weights and the creation of the 4 data files that are transmitted to EUROSTAT.

The EU-SILC processes in CYPSTAT are summarized in Figures 4a and 4b.

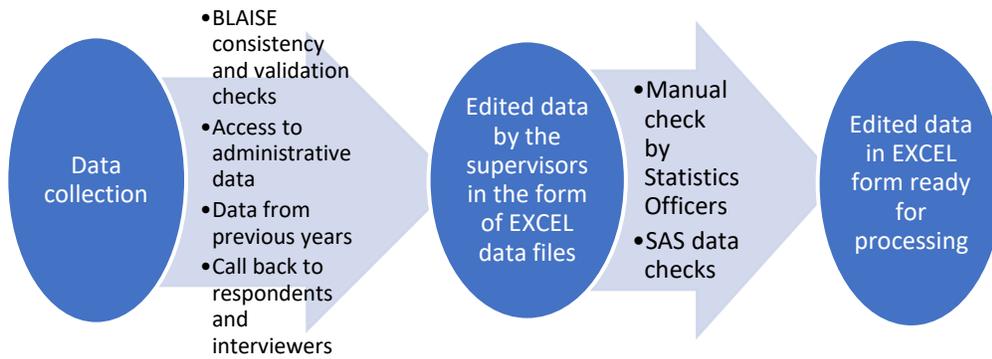


Figure 4a: Processes in EU-SILC



Figure 4b: Processes in EU-SILC

## 4. Future Plans

As already mentioned above, CYSTAT is a small office with limited human resources and in particular IT resources. The progress achieved during the years in the use of Blaise is mainly due to the significant support provided by the Blaise Team of Statistics Netherlands. It is envisaged that this excellent cooperation will continue and in the forthcoming years. Having as asset this excellent cooperation, CYSTAT plans to gradually transit to Blaise 5. Due to limited human resources, however, there is no specific timeframe for this transition. It is imperative that the size of CYSTAT's Blaise Team needs to increase in order to tackle all the challenges that are included in the transition.

One of CYSTAT's goals is to implement the ICT-ENT survey as multimode in CAPI and CAWI with Blaise 5. In this respect, in the 2020 survey one tablet has been given to an enumerator and he is given the task to carry out the CAPI by filling in the questionnaire developed in Blaise 5.6.4 Android App. The completed questionnaires are uploaded through the internet to the server. This process at the moment of writing this paper is still running and therefore, no feedback can be provided at this stage. However, from the experience gained so far there were some problems with the design of the layout, but those were tackled with the assistance of the Blaise Team of Statistics Netherlands.

From the relative short term experience in CAWI it can be concluded that most of the respondents prefer to complete a web interview using mobile device. However, this cannot be done due to the small size of text and buttons which cannot be resized. Question text, navigation buttons, and code frames are sometimes clipped if there is too much information on the screen. In such situations navigation is impossible because the buttons are unavailable. Moreover, the scrolls up and down buttons do not work. In this respect, the web screens should be optimized for the mobile too.

As regards the survey on Employment and Job Vacancies, the next step is to successfully convert Blaise 4 CATI instrumentation into Blaise 5. The Blaise 5 instrument should be executed in CATI, CAWI modes. In this respect, as soon as this conversion is achieved CYSTAT envisage to use Blaise 5 CATI in other surveys as well, for instance the EU-SILC. The long term goal is to set up a telephone centre in CYSTAT based on Blaise 5 for all the surveys that are carried out by telephone.

As far as the EU-SILC survey is concerned, in the 2021 the new IESS regulation will be implemented. The most difficult aspect of this regulation, for most of the countries, is the timeliness of the data transmission. The data must be transmitted at least 9 months earlier than the existing regulation. In order to comply with the IESS, a lot of the procedures must be reviewed and modernized. First, for the data collection it is necessary to upgrade from Blaise 4 to Blaise 5, especially as it supports multimode data collection. Additionally the field work needs to be shortened. It is essential to examine how to improve and further develop existing procedures in order to use administrative data more effectively. For example, questions regarding benefits, the salaries of the government employees and pensions from the Social Insurance Services and the Government should not be asked during interviewing when the Civil ID of the person is provided. This information should be uploaded automatically from the respective registers. Moreover, completed questionnaires from the interviewers should be transferred to a server and not locally, as it now, for better monitoring of both the data collection and the editing from the supervisors. This would help to decrease the length of the editing phase. In addition to that, the reviewing and modernization of the microdata editing and processing in Excel and SAS is necessary.

In 2021, the census of population will take place. Following the success of the previous census in 2011 where the data collection was carried out with netbooks and the information system was built on Blaise 4.8, it has been decided to carry out the census with tablets and Blaise 5. We expect that the experience to be gained in the whole process will be very beneficial in the implementation of the strategy to use Blaise in all surveys carried out by CYSTAT.

According to CYSTAT's strategy Blaise is the standard for data capture, CAPI, CATI, CAWI and CADI. Moreover, the data collection in several surveys is multi-mode. Considering that Blaise 5 offers the infrastructure to build the necessary tools for multi-mode data collection and coupling this with CYSTAT strategic aim, the full transition to Blaise 5 has no return.

## **5. Conclusions**

The transition from Blaise 4 to Blaise 5 is a long process which is further hampered by the lack of human resources. Considering, however, the benefits of Blaise 5 that mainly concern the support for multi-mode data collection, and relying on the support of the Blaise Team of Statistics Netherlands there is optimism that the strategic aim for the full conversion to Blaise 5 will be achieved.

The use of CAWI is expected to offer significant savings in the cost for carrying out the surveys. Indicatively, from the recent experience in applying CAWI in the ICT-ENT survey, 50% of the questionnaires were filled in online. The direct data entry by enterprises minimizes interviewers' contribution and the working hours required by the responsible officer.

Another advantage is automation and real-time access. Respondents import their own data and these are automatically stored electronically on the server. At any time it is easy to have access to the data. Daily reports are also available.

Finally, with the use of Blaise 5 it is envisaged improving the layout of the survey due to the design flexibility and respondents can provide the information at their convenience.