Data Transfer in Blaise Surveys with CAPI Collection Method

Lilya Luria, Central Bureau of Statistics, Israel

1 Introduction

Many surveys conducted by the Central Bureau of Statistics (CBS) in Israel are developed in Blaise. The most common data collection method in these surveys is the CAPI method, and therefore, there is a constant need to transfer data from the CBS network to the field and back.

The data from the field is transmitted through the 'Generic Broadcast System' - a user-transparent remote component which provides service for data transfer between remote environments and the internal network of the CBS. The data transfer is performed in both directions.
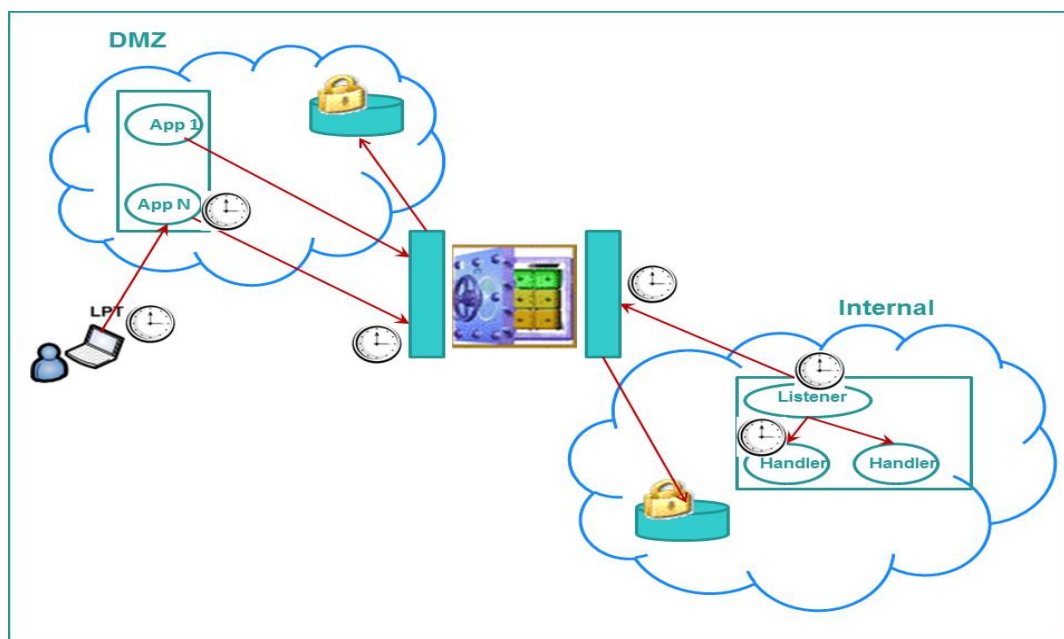
In order to connect the 'Generic Broadcast System' component, we have developed a set of generic applications which are used in all the surveys that are using Blaise for saving data on the one hand and using CAPI collection method on the other. These applications have been developed in Manipula and .Net using the BCP utility.

In this paper we will explain in details the process of sending Blaise files from the interviewer's laptop to the internal network and back and ways to troubleshoot.

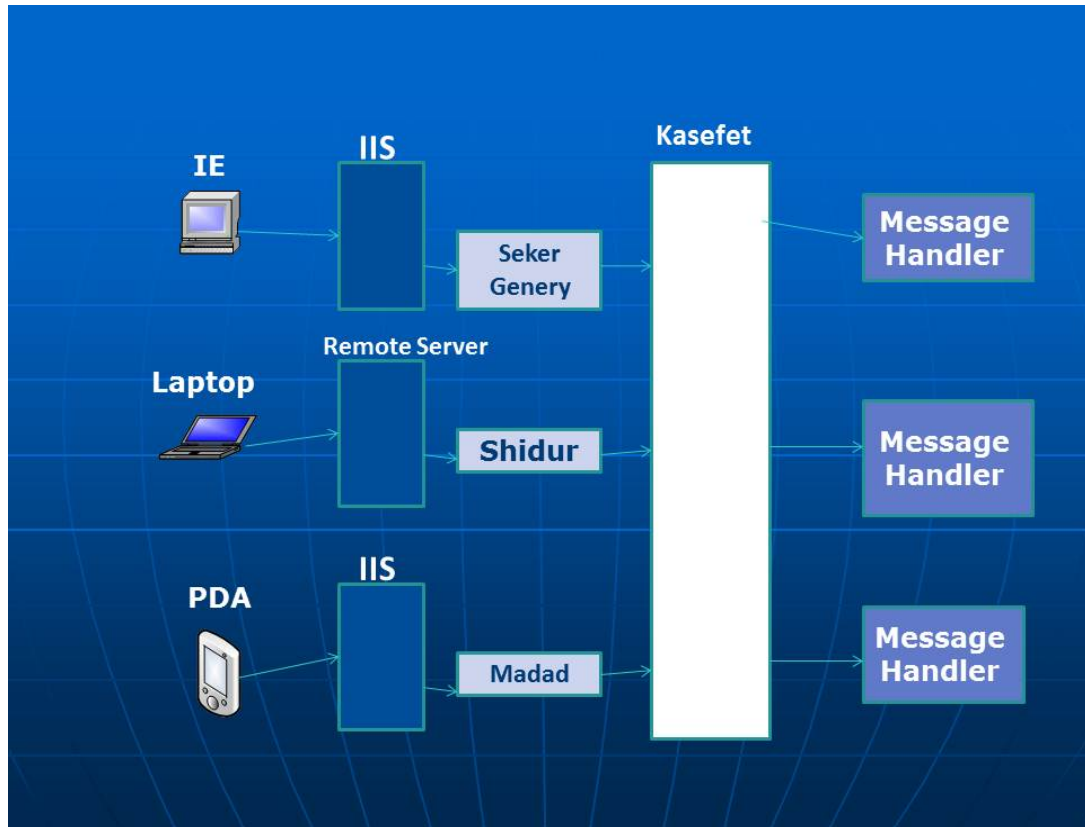## 2 The 'Generic Broadcast System' component

The 'Generic Broadcast System' is a component (.Net DLL) which has been developed by the team of generic applications at the CBS, and aims to meet the need to transfer data between different networks. This component is used today in all the surveys (Blaise surveys and others) which are using CAPI collection method.

The data is transmitted through the 'Kasefet' ("Safe") – the hardware & software component which main purpose is to store data and transfer it between networks.

In addition, the 'Generic Broadcast System' allows managing processes priorities, users' management and monitoring.

Data transferred via 'Generic broadcast' comes from different work environments: Web systems, laptop computers (CAPI), PDAs, etc., as shown in the following scheme:

In order to use the 'Generic Broadcast System', in each survey we need to implement a certain architecture which includes the following requirements:

Server Side:

- A dedicated collection server with folders for all interviewers: each interviewer has his own folder with a unique identification number (employee number).

- Within each folder, there are three sub-folders: Download, Upload and Backup.

- The Download folder contains files that are ready to be sent to the interviewers.

- The Upload folder contains files that have arrived from the interviewers.

- The Backup folder is used for a backup: at the end of each action all files are copied into the Backup folder. The Backup files are saved for a year - year and a half.

Client Side:

- In the laptop computers of all the interviewers we'll find the same three folders: Download, Upload and Backup.

- These folders are not per survey, they are used in all the surveys the interviewer works with.

- Each survey has an 'identification prefix' of 3 letters, for example, **Sch** for **S**aka**Ch**odshi (Monthly Labour Force Survey), **Pns** for **Pens**ion (Social Survey with an annual theme "Pension"), **Smr** for **SMAr**T (Long-term Households Survey), etc. As a customization for each survey, the specific prefix is indicated in the configuration file allowing the 'Generic Broadcast System' to differentiate between the surveys applications/services.

3   Data transfer via 'Generic Broadcast System' in Blaise surveys

The Blaise development team has developed a set of generic applications for transferring data via the 'Generic broadcast'. These applications are used in all Blaise surveys at the CBS.

Two of these applications are used on the server side (CBS statistic network), and perform the following functions:

- Preparing and 'packing' Blaise files that are ready to be sent to the interviewer's laptop and placing those files within a dedicated collection server.

- Updating the Central Blaise DB at the CBS network with the data received from the field via the 'Generic Broadcast System'. This action is executed through different management applications and could be performed by the initiative of survey coordinators or automatically.

Two other applications are used on the client side (interviewer's laptop), and they perform the following functions:

- Preparing and 'packing' the data collected on the laptop to be transmitted via the 'Generic Broadcast System' to CBS statistic network.

- Updating the Blaise DB on the laptop with the data received from the CBS network, as well as software version updates.

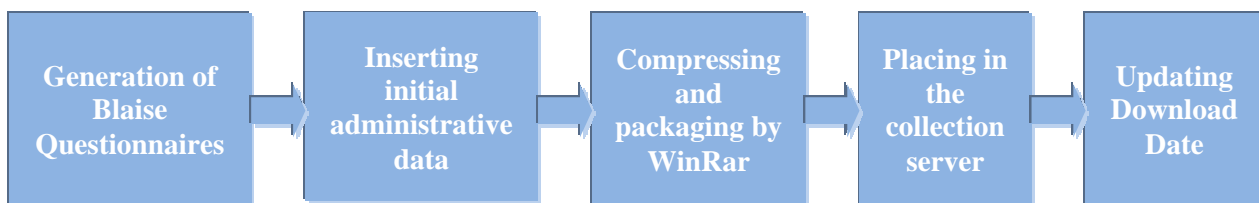# 4   Server Side Applications (Statistic Network)

4.1   Delivery Preparation Application

This application is activated by region coordinators. The region coordinator sets collection of questionnaires to each interviewer, and then performs 'Delivery Preparation'. This delivery includes all the questionnaires that are set to the interviewers and not been sent yet.

The main functions carried out by this application are:

- Filtering all sampled that were set to a specific interviewer and have not been sent yet.

- Generating Blaise questionnaires for those sampled.

- Inserting initial administrative data into the Blaise questionnaire, such as data taken from the Population Register, property tax files and other sources.

- If this is not the "first cycle" of the questionnaire (the questionnaire has been sent to certain interviewer and then returned to the coordinator, and now the coordinator set this questionnaire to another or the same interviewer), in this case the information produced by the previous interviewer inserted into the questionnaire, such as Actions Record Table – a table documenting all actions taken in this questionnaire by all the interviewers who had handled it.

- Generating separate Blaise DB for each interviewer who supposed to receive the package, with only her/his questionnaires included.

- Compressing and packaging of all Blaise files which supposed to be sent to the interviewers (.bdb, .bdm, .bmi, .bjk, .bfi, .bpk) by WinRar software.

- The generated RAR file name is composed of the following parts:
  - First 3 letters are representing the name of the survey
  - The fourth letter is **M** which means the file is sent from the CBS center (**M**erkaz)
  - Partial date composed of the month, day and time.
    For example, the file named **SchM**01231715 belongs to the Monthly Labour Force Survey (**S**aka **Ch**odshi) sent from the CBS Center (**M**erkaz) to the interviewer on 23/01 at 17:15.

- Placing the created RAR file in the 'Download' sub-folder of the interviewer's folder (The interviewer's "cell") in a dedicated collection server.

- Updating Download Date at the Central DB

Generation of Blaise Questionnaires	Inserting initial administrative data	Compressing and packaging by WinRar	Placing in the collection server	Updating Download Date

Although we do not know when the interviewer will draw the files sent to him from the collection server, once the questionnaires were sent they are considered as being under the interviewer's handling and the survey coordinator cannot perform certain operations on these questionnaires.
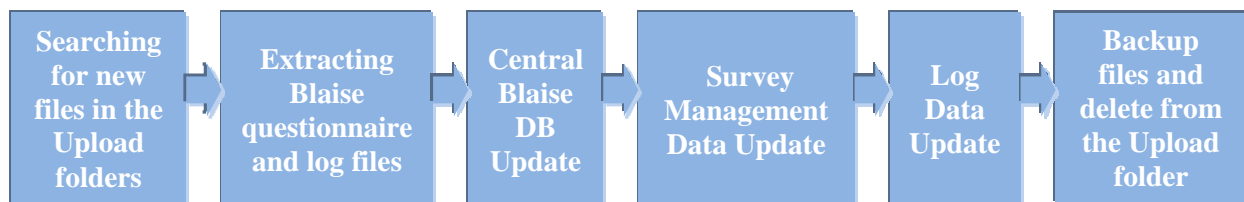
4.2 Receiving Questionnaires from the Interviewers Application

The purpose of this application is to withdraw from the collection server all the material sent by the interviewers and update the data in the Central Data Bases. In some surveys this is done automatically, on Windows Tasks Scheduler, usually at night, and in other surveys the material withdrawal is triggered manually by region coordinators through the survey management application.

The main functions carried out by this application are:

- Retrieving from the Survey Coordinator's management system a list of all interviewers working under the coordinator who activated the application.

- Check for each interviewer's number in the list: whether there is a material (files) in the 'Upload' sub-folder of the interviewer's "cell" in the collection server - those are the last files sent from the interviewer's laptop and have not been collected by the survey coordinator yet. For each survey, the application searches only for the files belonging to the same survey, and this identification is, as mentioned above, by the three first letters of the file's name, representing the name of the survey.

- The files arrived from the interviewers are compressed files with a '. RAR' extension and can be one of two types:
  - A file containing the questionnaires handled by the interviewer;
  - A file containing the log data (daily work reports each interviewer fills in the laptop).

- The received file name consists of the following parts:
  - First 3 letters representing the name of the survey;
  - The fourth letter is **S** (**S**oker - Interviewer) in a file containing the questionnaires, or **Y** (**Y**oman - Log) in a file containing the data from the daily reports;
  - Partial date composed of the month, day and time.

  For example, two files found in the 'Upload' sub-folder of the interviewer X:
  Sch**S**01261825.RAR
  Sch**Y**01261825.RAR

- The Blaise files (.bdb, .bdm, .bmi, .bjk, .bxi) are extracted from the RAR file at the collection server.

- Analysis of management information in each questionnaire that arrived, especially from the "Actions Record Table": the questionnaire status (completed / non-response), the reason of non-response, the follow up status, etc.

- The questionnaire management information is checked to determine if the questionnaire meets the criteria for updating, if it is - the arrived questionnaire is being saved at Central Blaise DB.

- Updating management data of the survey in accordance with management data in the questionnaires, such as Upload Date, Questionnaire Status, the reason of non-response, etc.

- Updating the log data in the survey management system.

- Backup files in the 'Backup' sub-folder of the interviewer's folder.

- Deleting files from the 'Upload' sub-folder of the interviewer's folder.

Searching for new files in the Upload folders	Extracting Blaise questionnaire and log files	Central Blaise DB Update	Survey Management Data Update	Log Data Update	Backup files and delete from the Upload folder

5 Client Side Applications (Interviewers' Laptops)

Each interviewer who works in one of the Blaise surveys, has another icon on her/his laptop's desktop in addition to the survey's application icon – the adjusted generic application by which 'Generic Broadcast' is activated.

Clicking on the icon activates the application consisting of 4 main parts:

- Preparation of Blaise questionnaires 'package' ready to be sent to the Central DB.

- Call the 'Generis Broadcast System' DLL which performs data transfer in both directions.

- Updating Blaise DB on the interviewer's laptop with the questionnaires arrived from the Central DB.

- Software Version Update (if relevant).

5.1   Preparation of Blaise questionnaires delivery

The application checks for questionnaires that must be delivered to the Central DB. In fact, the selection logic for sending questionnaires from the laptop is fairly simple - every questionnaire that has at least one action recorded in Action Record Table – is taken to the delivery. This application does not perform a thorough analysis of the questionnaire: whether it must be updated in the Central DB, what status it should be given, etc. This operation is performed on the server side.
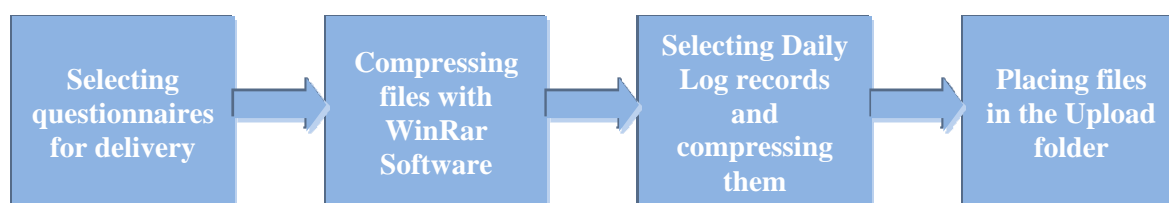
All questionnaires that meet the delivery criteria are copied to a separate Blaise DB intended for the delivery.

Then these Blaise files are compressed and 'packaged' by the WinRar software with a '.RAR' extension (similar to Delivery Preparation on the server side).

In addition, there is another Blaise application installed in all interviewers' laptops called "Yoman" ("Daily Log") - an application through the interviewer fills her/his daily work reports, and it is the same log for all the surveys.

While preparing the delivery, if the interviewer has filled new records in the log (daily report), they are packed for delivery as well, though in separate RAR file.

These RAR files are copied to the 'Upload' folder on the laptop, as a preparation for data transmission.

Selecting questionnaires for delivery	→	Compressing files with WinRar Software	→	Selecting Daily Log records and compressing them	→	Placing files in the Upload folder

5.2   Data Transmission

After preparing the delivery, our application triggers the 'Generic Broadcast' DLL.

For this purpose the interviewer must be equipped with landline or mobile phone (for transferring data between networks) + special identification token (for information security reasons).
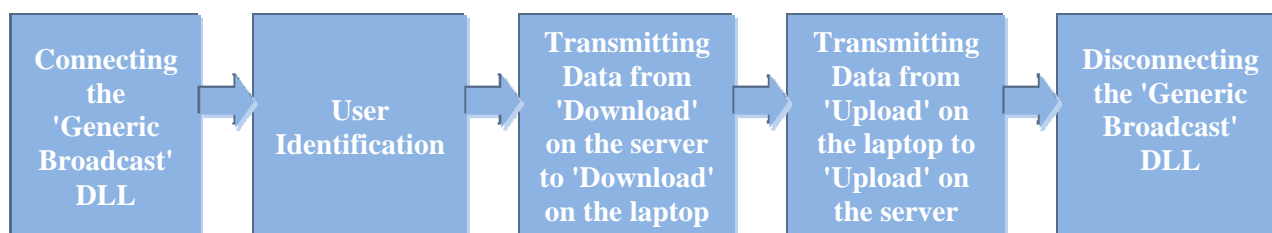
The interviewer receives the 'Generic broadcast' log-in window. The interviewer is required to enter predefined username and password, where the second part of the password is generated by clicking on the token, and expires after 30 seconds (for information security reasons).

Once connection is established, user authentication and password are being verified in front of the Central Broadcast DB.

Then the DLL turns to the collection server and searches in the 'Download' folder of the given interviewer whether there are files waiting to be sent from the collection server to the interviewer's laptop. If there are, the 'Generic broadcast' transmits these files from the ' Download' folder on the collection server to the 'Download' folder on the laptop.

Next step, the Broadcast component performs the inverse operation: searches in the 'Upload' folder on the laptop whether there are files waiting to be sent from the laptop to the collection server. If any, 'Generic Broadcast' transmits these files from the 'Upload' folder on the laptop to the 'Upload' folder on the collection server.

At the end of these operations our application disconnects from 'Generic Broadcast' DLL and performs data update on the laptop.

Connecting the 'Generic Broadcast' DLL	→	User Identification	→	Transmitting Data from 'Download' on the server to 'Download' on the laptop	→	Transmitting Data from 'Upload' on the laptop to 'Upload' on the server	→	Disconnecting the 'Generic Broadcast' DLL

5.3   Blaise DB Update on the laptop

If there are files in the 'Download' folder on the laptop, our application will extract these files (an UnRar action). The arrived files are Blaise files containing questionnaires sent by the coordinator. These can be new questionnaires or questionnaires that have been sent in the past, returned to the coordinator and now sent back to the interviewer (same interviewer or another one).

The application updates Blaise DB in the laptop with the arrived questionnaires and with the management data that is also found inside the Blaise questionnaires.

Rarely happens that the file arrived from the collection server, has the same name as already existing file in the 'Backup' folder, so the application checks the RAR file name before extracting files from it. If for some reason file with the same name already exists (including the date and time), the new file is saved in the 'Backup' folder with another extension (in order to not overwrite the existing file), for further clarification.

5.4   Software Version Update

Once our application has finished the questionnaires update process in the laptop, it checks whether there are also software version update files in the Download folder.

Occasionally it is necessary to update the software version (Blaise questionnaire or interviewer's management system) on the laptops. For this purpose, in addition to the usual file containing collection of questionnaires for the interviewer, software version update files are also sent from the central management to the laptop.

If the application detects there are version update files, the interviewer gets a notification about that.

For version update, 2 files usually sent to the interviewer's laptop:

- RAR file containing all the Blaise and Manipula files that must be updated on the laptop. If the new version of the questionnaire has different structure from previous version, the RAR archive also contains an '.msu' file that copies data from the Blaise DB on the laptop to the new structure.

- .BAT file containing all the commands to be performed on the laptop.

The application runs the BAT file, which usually performs the following commands:

- Extracting files from the RAR Archive.

- Running MSU file that transfers data from the old structure DB to the new structure DB (if required).

- Copying files from the 'Download' folder to the questionnaire DB folder in the laptop.

At the end of the process a backup of all files is done in the 'Backup' folder and all files from the 'Upload' and 'Download' folders are being deleted.

6 Issues / Problems & Troubleshooting

There are several problems / issues which arise from time to time during the process of data transfer.

The main issues are:

- **System Folders Update**
  The hard disk in the laptops is divided into drives C and D.
  For security reasons there is a very clear distinction between types of files installed on both drives.

  All the System files are installed on drive C, including Blaise System files.
  Drive D is designated to applications' installations of various surveys.

  When we send a version update via the 'Generic Broadcast', in terms of the interviewer's permissions on the laptop we can update only the files located on drive D. In most cases this mode is completely satisfying, since the survey files that need to be updated are all located in drive D.

  However, once in a while we need to update files in the Blaise System folder - a folder located on drive C.

  For example, in laptops using the BCP component it is necessary to update Blaise license every year - quite problematic when the interviewers with their computers are distributed across the country. Theoretically, we could send a file with the license key through a version update for the interviewers, but since this file must be placed in the Blaise folder on drive C, for security reasons the interviewer cannot receive it through the 'Generic Broadcast'.

  Therefore, as for today - just before the old license expired, all the interviewers arrive physically to the CBS regional offices and got their Blaise license updated manually by the technical team under Administrator permissions.
  There is no doubt that this situation is not very convenient, especially in large surveys with a large number of interviewers.

  Apparent solution in the near future:

  One of the CBS development teams is working these days on developing a generic application for software versions updates on all laptops. The application will be activated while transmitting data via the 'Generic Broadcast'.
  One of this application advantages is that the update will be performed under Administrator permissions, so it will be possible to update the files that are installed on drive C.

- **The need for a collection server as an intermediate stage**
  One of the questions that arise from time to time, especially in new surveys, is why when the interviewer initiates the data transmission operation in the laptop, it is not fetching the material directly from the Blaise DB and survey management system on the internal network, or directly updating this systems with files sent from her/his laptop.

The fact that the data transmitted in both directions "stops" in an intermediate stage which is collecting server, and "waits" to be picked up, forces the survey coordinators to activate two actions: sending material to the collection server and picking up material from the server to update the system.

The first reason for this is data security. To reach the maximum security level, the material passed between the networks stops at the collection server and does not go directly to the survey servers while the communication channel is open.
Only after the interviewer had disconnected from the broadcast and the communication channel is closed, the survey coordinator can fetch the arrived material from the collection server to the survey servers.

Another reason for the existence of the collection server is a technical reason - updating the survey systems (Blaise DB and management system) directly from the interviewer's laptop will extend the duration of the transmission process, which could cause technical problems, and also for reasons of data security it is better to shorten the transmission time as much as possible.

- **Technical problems on servers**
  Rarely, various technical problems occur on collecting servers, such as a slow server, sometimes resulting in disruptions in data transmission process.

  As for today, a specific technical solution is given to any problem that arises locally, and the development team together with the technical staff restores files that have not been transmitted, in case it happens.

  There are plans in the near future that the development team and technical staff will find more global solutions to such situations.

7   Summary

Data transfer from the field to the CBS internal network and back in Blaise surveys with CAPI collection method is performed in a generic way, same for all surveys.
This generality is expressed in using the 'Generic broadcast' DLL on the one hand, and the set of generic applications developed by the Blaise development team and using the 'Generic broadcast', on the other hand.

Each new CAPI survey conducted on by CBS joins this infrastructure, which enables adding each new survey to the data transfer mechanism in a convenient and relatively simple way, with only minor changes between the surveys.