3 Survey Instruments in SHARELIFE

Marcel Das, Maurice Martens and Arnaud Wijnant

3.1 Electronic interview implementation

As Chapter 2 has laid the ground, this chapter turns to the actual implementation of the survey content into the SHARELIFE study. The SHARELIFE questionnaire differed in many ways from the questionnaires that were used in the two previous waves of SHARE. Questions and routing were completely different. In addition, there was the need to graphically display the Event History Calendar (EHC). This had an impact on the interface design as well as choice of software. The architecture was largely taken from ELSA, which used the EHC method before. All survey instruments in SHARELIFE were programmed by CentERdata, a research institute housed on the campus of Tilburg University, the Netherlands. The previous waves used the computer-assisted interviewing system Blaise, developed for the Windows operating system by Statistics Netherlands and designed for use in official statistics. Although SHARELIFE also made use of Blaise as questionnaire engine, it was decided to replace the default data entry programme used to display questionnaires written in Blaise by a new programme.

As in the two previous waves, the CAPI (Computer Assisted Programming Interview) application had a generic structure in terms of routing and setup. The main SHARE concept of *ex-ante harmonization* of the survey instrument was not changed. Only the language of the question texts, interviewer instructions and answer categories differed by country, while the generic questionnaire, basis for all countries, was in English. Translations of texts and the events could be entered in the Language Management Utility (LMU), which was originally designed for SHARE wave 1 (see Das et al., 2005) and updated for wave 2 (see Brouwer et al., 2008). Further fine-tuning happened for SHARELIFE, not only concerning the stability and handling of the LMU, but also some specialized sections were added to support the EHC.

The EHC CAPI application contained one additional country-specific element. The application had a "look up" area where one could search for general country-specific events, allowing the respondent to use landmark anchoring as discussed above. The events were specified directly by the country teams, covering occurrences that were commonly thought to be remembered by respondents and serving as a reference point. The events were grouped in social, economic, political, sports, natural and other events.

The EHC CAPI application enabled the interviewer to gather detailed information about important (personal) events in the respondents' lives. The individual's history of events was then shown on the screen. The graphical display of the calendar grid, the "look up" area for general events, and the display of the individual events required a different Data Entry Programme (DEP) than the default Blaise DEP. As the interface that the interviewer uses to go through the questionnaire is important for the success of the whole project, it was decided to

implement a new DEP in Visual Basic that could communicate with a compiled Blaise questionnaire. Figure 3.1 displays the DEP as it was used in SHARELIFE.
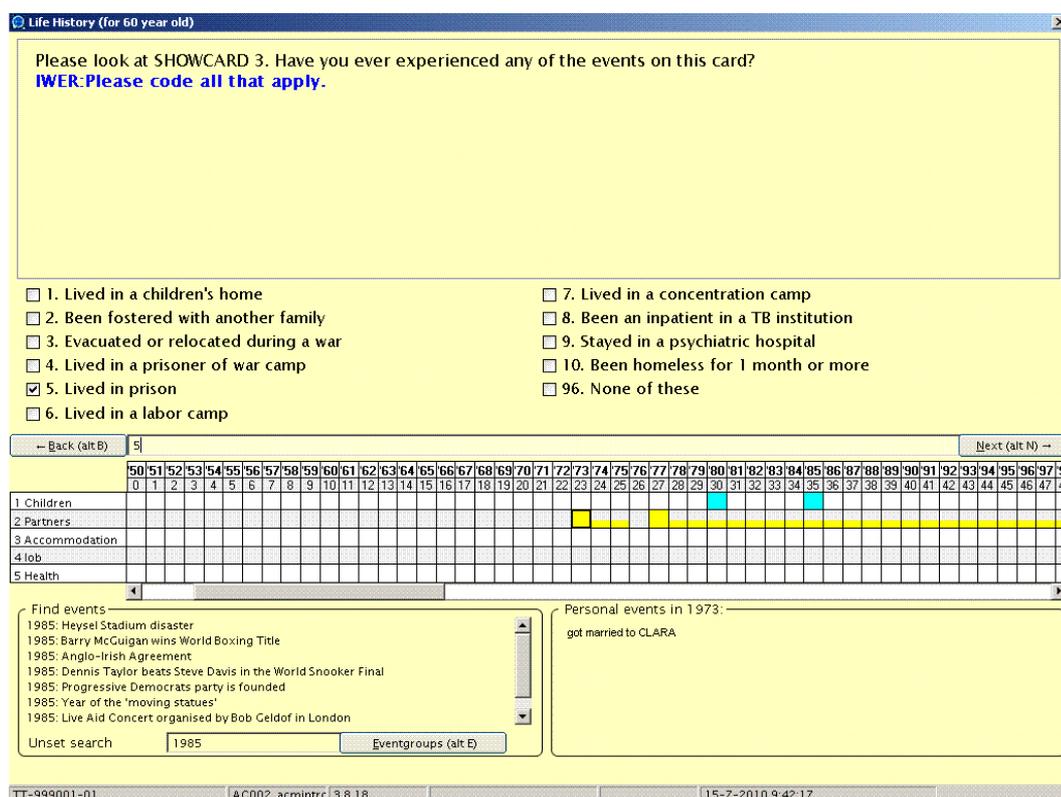

Figure 3.1: Screenshot of Data Entry Programme as used in SHARELIFE

The DEP-screen shown in Figure 3.1 is divided into six parts. On top is the question part, where the current text the interviewer is supposed to read out is placed in black font while interviewer instructions are set in bold and blue font. Below are the answer categories that are possible for this question – this part is empty if it is an open question and there is text to be entered. The answers can either be clicked on with the mouse, or they can be typed into the answer line below the answer categories, where also open questions are answered.

Below the answer line is the main new feature of the instrument, the calendar, displaying the timing of each of the life history events or modules. The following five "key events" in life were displayed vertically: children, partners, accommodation, job, and health. The top of the calendar section displays each year of the respondent's life with his/her corresponding age, starting from the year the respondent was born. (The respondent was asked for his/her date of birth at the beginning of the interview.) Cells were colored when certain events occurred in the specific year and module. As mentioned in Chapter 2, it was possible during the interview to quickly jump to questions previously answered in a different module. As can be seen in the screenshot, some personal events already have

already been entered for this respondent, for example two children in the first line, and two relationships in the second.

The "personal events" box on the lower right hand side of the DEP field is filled while going through the questionnaire. If the interviewer selected a cell in the calendar grid, a description of the personal events that occurred in the selected year was displayed. The events list corresponds directly to the filled squares in the calendar and serves as a tool to get more details about the event in the calendar. In Figure 3.1, the marker in the calendar is set to "1973", in which the coloured square shows that the respondent started a relationship. Looking at the "Personal Events" box, the interviewer could see that it was actually the marriage with Clara which began in 1973.

Below the calendar on the lower left hand side is the field labelled "Find events". Whenever a respondent could not remember the exact year in which a particular event happened, but knew it was around the time that some general event occurred, this search tool could be used. Country-specific events could be added by the country team; a basic list of events was derived from Wikipedia for each country separately.

This search tool is a completely new development for SHARELIFE. Before this implementation, the respondents had a list of events for each year, but here the problem was that either the interviewer or the respondent had to have the knowledge of when an event had happened for these lists to be useful. For example, if the respondent remembered that her move was a month after J. F. Kennedy was shot, then this does not help if neither she nor the interviewer knew when that happened. So this led to searches for events, which cost time and was at times frustrating for both the interviewer and the respondent. The new search tool provides a solution for this, because the interviewer can search for a specific event, as shown in Figure 3.2:
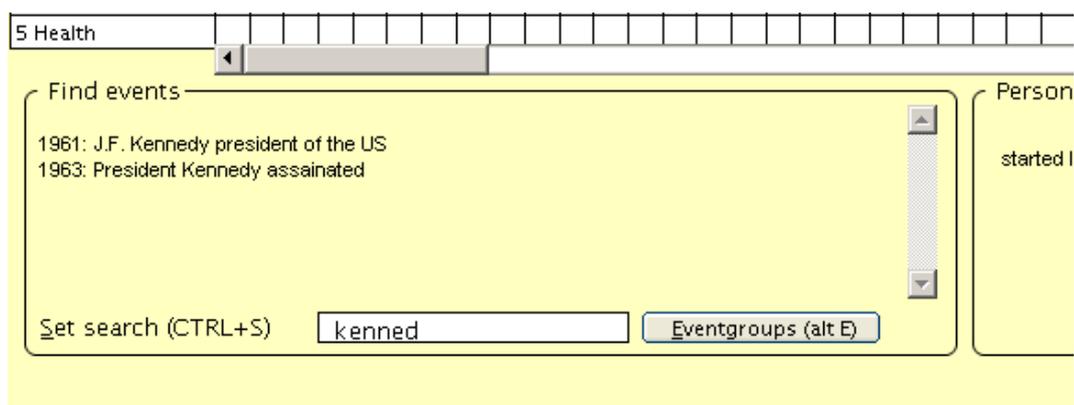

Figure 3.2: Screenshot of Search Tool

Here the interviewer entered "kenned" and a list of events that contain this string appears in the box. The current example shows two events, the inauguration and the assassination of J. F. Kennedy. In a similar way (see Figure 3.1), the

interviewer could enter a year, and then all events of that year appeared in the "Find events" box.

The electronic implementation of the Event History Calendar proved to be a successful tool in the end – both interviewers and respondents provided useful feedback for improvements during the testing phase. Not all respondents took interest in the graphical representation of their lives, but many of them actually asked to have a printout of the screen after the interview.

3.2 Sample management

With more than 30,000 cases to handle, a survey like SHARELIFE needs an efficient way to manage the sample electronically, especially if this sample is distributed across thirteen different countries and fourteen different survey agencies in Europe. CentERdata developed the SHARE Sample Management System (SMS) for SHARE's previous waves (see also Das et al, 2005, and Brouwer, 2008), and programmed a major upgrade for SHARELIFE. The system consists of two main components:

- a Sample Distributor (SD) installed on a server at each survey agency
- a Sample Management System (SMS Client) installed on the each interviewer's laptop

Both components were programmed in Java, an object-oriented and platform-independent programming language that, in terms of syntax, is largely based on the programming language C++. Both the SD and SMS Client have a similar user interface. The SD was created to monitor the complete fieldwork in a country (see also Chapter 4), to distribute the sample among these laptops in the fieldwork, and to gather the data from the laptops in the field. The SMS Client was used to record contacts, to decide who to interview, and to start the actual interviewing. An overview of how the two components work together is shown in Figure 3.3.

In contrast to SHARE wave 2, the SD and the SMS Client were the only means of accessing the data during the fieldwork. This means that corrections on the data were only possible via these programmes or via programmers at CentERdata who could create correction files which could be either loaded in the SD or the SMS Client application. This design-decision contributed to a harmonized data-processing procedure, enabling CentERdata to deliver the data of more than 30,000 interviews in thirteen countries in a reasonable short amount of time.

Another major change in the management software used in SHARELIFE was the introduction of a feature to make a copy of the household state after every action in one of the two components. This copy could always be used to go back to a previous state if any mistake was made. This mechanism also made it possible to track what happened to a household or respondent over time in order to analyze difficulties during the fieldwork.
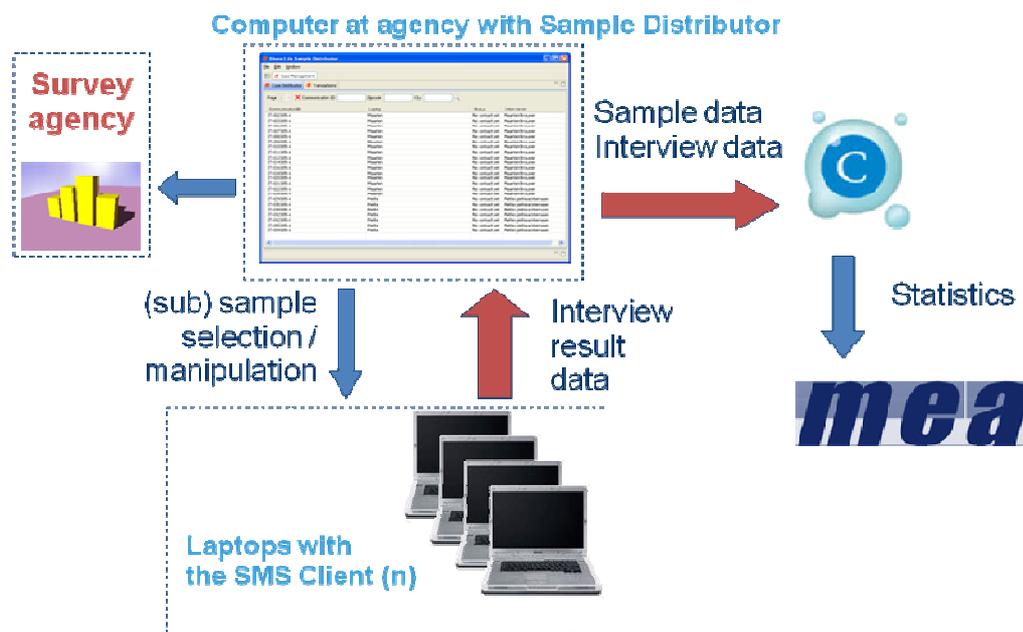
Figure 3.3: Relations between the Sample Distributor and SMS Client

Coverscreen wizard

The interview starts with a "coverscreen" that offers an introduction to the study and collects basic demographic information about everyone currently living in the household. The coverscreen establishes which household members are eligible to participate in SHARELFE. In SHARELFE a coverscreen wizard was used, which replaced the Blaise coverscreen interview for respondents who did not object to re-using data from an interview in a previous wave. Unlike the Blaise coverscreen, the coverscreen wizard is fully integrated in the SMS Client application, which has several advantages:

The first advantage is that the wizard has full access to the data stored for the sample as it was collected in the previous waves. With this information the wizard can check if there are any changes in the household-composition. The wizards helps the interviewer to determine whether a person's name and birthday are still correct, whether relations within the household are still as in previous waves, and whether people have deceased/moved out. The wizard also checks whether any new persons have joined the household.

The second advantage is that the wizard is able to correct the sample data inside the SMS Client application. New households can be created if a split-off (e.g. due to a divorce) occurs, names and birthdays can be corrected, but also new respondents (e.g. new partners of a previous respondent) can be added to the household. Household addresses can also be changed via the coverscreen wizard, if a whole household moved to a new address.

The third advantage is that certain information can be generated automatically after the coverscreen wizard is completed. This can explain why a respondent became ineligible for this wave. This automated information can help survey

24

agencies to analyze the states of specific respondents in the sample. Since it was necessary to use data collected in previous waves, the coverscreen wizard was only available for respondents who did not object to using previously collected data in the SHARELIFE interview. In case a respondent objected, the SHARELIFE SMS client application started the "old" Blaise coverscreen.

Synchronization

In order to synchronize data between the SMS Client application, the SD and CentERdata, the File Transfer Protocol (FTP) was used. FTP is a protocol that can be used to send files from a FTP-client application to a FTP-server application on a different server. The Internet provides the connection between the two components.

Both the SD and the SMS Client application have a built-in FTP-client. The client in the SMS Client was used for the synchronization between the SMS Client and the SD. The FTP-client in the SD was used for the synchronization between the SD and the CentERdata FTP-server. To set up the complete SHARELIFE software with automatic FTP-synchronization enabled, a FTP server had to be installed on the agency's server. An overview of the synchronization is presented in Figure 3.4.
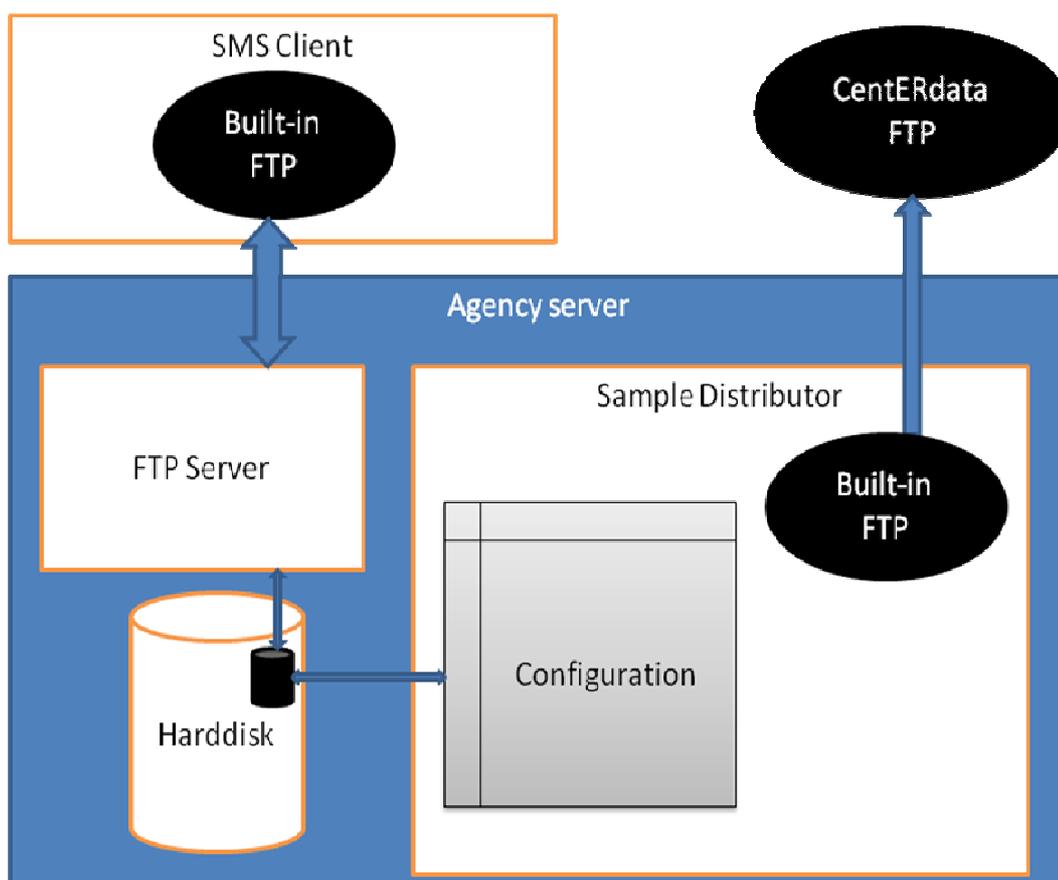

Figure 3.4: System overview for automatic FTP-synchronization

To carry out the actual synchronization, files were sent between the different FTP-clients and servers. There are two types of files: transaction files and CentERdata export files. Transaction files contain interview data, log files, and sample data of the households. These were used for the communication between the SD and the SMS Client. The SD is only able to *read* sample data from these transaction files, but can forward the log files and interview data to CentERdata. For this purpose the CentERdata-export file type was used. This file contains all interview and log data of the collected transactions plus a copy of all the data stored inside the SD (except for identifying information like the names and addresses of respondents). CentERdata used these files to create interview data files, which can be downloaded by the country teams to see the results of the interviews and to analyze the collected data.

3.3 Security

It is important to take security measures when working with personal data as in the SHARELIFE project. CentERdata created a protocol to minimize the risk of data abuse. Before and during the data collection phase, four actors were involved: Mannheim Research Institute for the Economics of Aging (MEA, Mannheim, Germany), CentERdata, the survey agencies, and the interviewers (in all participating countries). In some countries members of the country team actively participated in the data preparation phase. For simplicity's sake, these persons are considered part of the agency.

MEA provided the input from previous waves for the longitudinal households – the so-called preload database (see also Chapter 5) – which was necessary for use in the coverscreen wizard. CentERdata added names to the preload databases before the data were loaded into the SD. The following steps were taken, starting from the creation of the preload databases to the final processing of the collected data.

1. CentERdata received the processed data for the individuals in the SHARE study from MEA (preload data, collected in earlier waves).
2. CentERdata built country-specific SHARE SDs containing the preload data for each country. These country-specific SDs were made available for download from a password-protected website.
3. The different agencies downloaded the SD and installed it on a local machine/server. The security settings at this point were agency specific.
4. After its installation the address information was imported into the SD by each agency itself. Address information always stayed at the agencies; CentERdata and MEA did not have access to it. The imported address information was stored in a password-protected database and could be viewed by the agency's administrator. He/She could only log in to the application with an individual specific login code.
5. The agencies distributed the data from the households to different laptops. To initialize a laptop, a laptop-specific initialization (INI) file was created on the SD. This INI file only contained an ID to set up communications between the SD and the SMS Client (laptop). After the INI file had been

processed on the SMS Client it was automatically removed from the laptop.

6. The communication between the SD and the SMS Client was based on transaction files. These files could contain personal data. For security reasons, these files were encrypted using symmetric encryption, requiring a password as key (MD5 and DES). The transport of these files could be automated using an FTP-server at the SD side of the communication. The SMS Client had a built-in FTP-client to send the files to an FTP-server.

7. All personal information at the SMS Client side was stored in a password-protected database as well. Blaise files (containing the interview data) were stored on the laptop, and could only be read by Blaise itself.

8. The transaction files contained the contact database and Blaise files. These files were only available in an encrypted state at the SD side (using the MD5 and DES encryption methods).

9. Periodically the agency would export the collected data from the SD to CentERdata. This communication also used the encrypted (MD5 and DES) transaction mechanism. The data (except the Blaise files) were stripped of personal data before being sent to the CentERdata server. Only the encrypted Blaise files still contained personal data (names, gender, and year of birth).

10. CentERdata removed the personal data from the Blaise files as soon as they arrived at the server. This was done before their export to SPSS for processing (merging of databases, creating dummy variables, etc.). These data were only stored on an internal CentERdata server.

After the fieldwork was completed, the final data processing was performed by CentERdata. Processed data were sent to MEA for cleaning and preparation for scientific use. Users can download and use the data after signing a confidentiality statement in which the researcher agrees to the conditions for using the data (see http://www.share-project.org). After the signed statement has been received, a personal login code to access the data is provided. The public data contain scrambled identification (ID) numbers and do not contain any names of the individuals or households. CentERdata is the only actor who has access to the key to scramble the ID numbers.

3.4 Concluding remarks

Since the start of the SHARE project it has been a challenge to develop the survey instruments for the use by the agencies in the participating countries. The first survey instruments were developed in wave 1. Wave 2 could benefit from all lessons learned in the first wave, but an additional complexity was added due to the longitudinal dimension. . This required the programming of preload information collected in wave 1. Matching the correct individuals sometimes turned out to be a tough exercise. In SHARELIFE the setup of the survey instruments was such that after every action performed by the interviewer, a copy of the household state was made. The interviewer in the field, the agency, and CentERdata could revert to a previous state if a mistake was made. This