Blaise 5 in SHARE

Maurice Martens, CentERdata

1. Abstract

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of about 140,000 individuals aged 50 or older (around 380,000 interviews). SHARE covers 27 European countries and Israel in 40 locales and is currently fielding its ninth wave. The questionnaire is fielded in CAPI mode.

CentERdata scripts the questionnaire in Blaise and has developed several tools that connect to these questionnaires. For latest wave the questionnaire was migrated from Blaise version 4.8 to version 5.4.

This migration was not only a change within the Blaise environment but also forced an adaptation of all tools we developed that interfaced with Blaise in our architecture. This step meant a new case management system was developed, new interfaces with our Translation Management Tool (TMT) and a new design for the Data Delivery backend were created.

This paper discusses these solutions and will also address some positive and some negative findings. The Blaise 5 development environment, full Unicode compatibility, the introduction of roles were greatly appreciated. The limited functionality for triggering events, keyboard navigation and control over lookup tables is something that should be improved on. We will discuss the workarounds we implemented to get these features in.

2. Introduction

The Survey of Health, Aging and Retirement in Europe (SHARE)) is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of about 140,000 individuals aged 50 or older (around 380,000 interviews). SHARE covers 27 European countries and Israel in 40 locales and is currently preparing its ninth wave. The questionnaire is fielded in CAPI mode and is fully ex-ante harmonized, this means that the questionnaire is centrally programmed and its underlying data model is identical for all countries. This allows for a quick turnaround and easy access to the data during fieldwork.

CentERdata is an independent, non-profit research institute located on the campus of Tilburg University (NL). The institute is specialized in (online) data collection, data dissemination, survey methodological research, tailor-made software solutions, model building, and socio-economic (policy) research. CentERdata hosts a representative household panel, the Longitudinal Internet Studies for the Social Sciences (LISS) Panel, with excellent possibilities for representative surveys and (controlled) experiments. In addition, CentERdata has expertise in state-of-the art econometric and data science models and techniques and collaborates closely with (academic) researchers on national and international projects. CentERdata also offers support in the field of valorization and project management, for example to support companies to arrive at data-driven and evidence-based decision making.

CentERdata scripts the SHARE questionnaire in Blaise and has developed several tools that connect to these questionnaires. For the eight wave of SHARE, the questionnaire was migrated from Blaise version 4.8 to version 5.4. This migration forced us to adapt many of the tools we developed over the years that interfaced with Blaise in our architecture. This paper will discuss the software architecture as used in the SHARE study in wave eight. The SHARE study is ex-ante-harmonized. We develop a single source questionnaire and define a data model that is used in all participating countries. Only the translations differ per language. An overview of the SHARE dataflow depicted in the figure below.

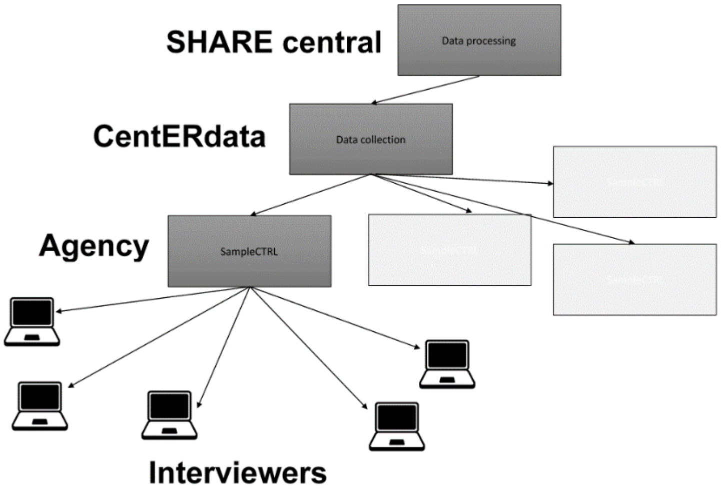


Figure 1: Architecture

We send out interviewers with laptops to households. There are two types of households: panel households; they have done at least one interview with a person in the household in a previous wave, or refreshers; new households that participate for the first time. On the laptop a Blaise questionnaire is installed, together with CaseCTRL, a program that can be used by the interviewer to manage the local subsample. It registers contacts and contact attempts, it has a build in agenda, and runs a household grid. When the interviewer has a connection to the internet, they can synchronize their data from CaseCTRL with an installation of SampleCTRL. Sample control is an application that runs at a server at a local agency that is hired to coordinate the fieldwork in a country. This agency can monitor the individual interviewers, can assign and (re-) distribute subsamples. They are requested to synchronise with a central server every two weeks. In this synchronization step the data is anonymized and send over to a Data Collection server. This server is hosted by CentERdata. In this Data collection environment, the data is joined together and some basic validation is done on the data. When the data passes the validation it is pushed to the SHARE central server in Munich, Germany where during fieldwork the data is processed into fieldwork progress reports and ultimately disseminated, cleaned and published.

The SHARE development workflow starts with a document that describes the questionnaire definition for the new wave. CentERdata will review this and builds a first version of the questionnaire in Blaise, this is tested and if needed adaptations in questionnaire definition or in the programmed source tool are done. Several of these cycles are done until everybody is happy with the programmed source questionnaire, and we freeze the development of the source questionnaire. Now the questionnaire is uploaded to the Translation Management Tool (TMT). The TMT is an online tool that coordinates the translation process. It is can be configured to support various translation processes, including the TRAP-D process. Up to wave 7 of SHARE, the TMT used the Blaise API to load in a compiled Blaise source questionnaire, detected changes in the source version from an earlier version it already had stored, and flagged those in the translation environment. The translatable elements are shown in context, together with texts used in the other tools, like the Sample Management System and are translated by professional translators. We will import the translated text in a Blaise questionnaire and several iterations start, where a version is tested, translations are adapted and new tools are generated. During the country specific testing phase, it can happen we find country or language specific requirements that force an adaptation in the source questionnaire. If the issue is deemed problematic enough this might force an adaptation to the original definition, which restarts this loop from start, for the involved items. Ultimately tools are generated that can be used to do fieldwork with. Share has three such development cycles per wave. One for a pretest, where new items are tested, next for the field rehearsal, which aims at generating a final instrument to be used in the main fieldwork phase, and last the version for the main fieldwork phase, in general only bug fixes and possibly some questions removals.

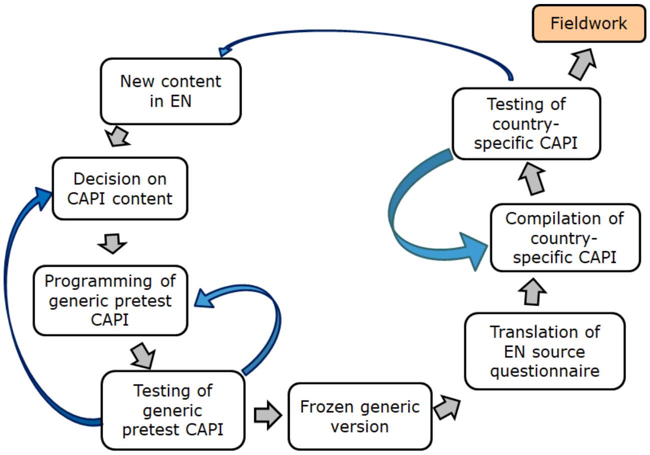

Fieldwork
New content in EN
Testing of country-specific CAPI
Decision on CAPI content
Compilation of country-specific CAPI
Programming of generic pretest CAPI
Translation of EN source questionnaire
Testing of generic pretest CAPI
Frozen generic version

Figure 2: Questionnaire development process

We intended to keep this development workflow for the wave eight instrument, but hoped to reduce the number of cycles due to bugs. Especially on versions, we generated for languages with Non-Latin scripts, often issues were reported with non-readable text, especially when fills are used.

The SHARE tools up to wave 7 used the TMT to load in texts into the CAPI and the SMS. The CAPI is in this context the Blaise questionnaire which is called via the DEP. It is started by the Sample Management System (SMS), this talks to a tool called Sample Distributer (SD), which transfers to the Data Delivery System (DDS). Finally, the data is ported to the Data Portals, which has three views on it: Internal, External and Questasy (see figure 3).

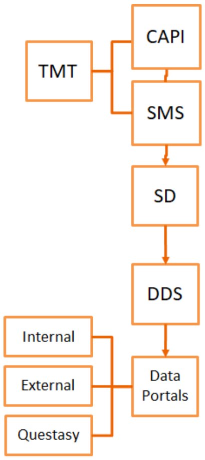


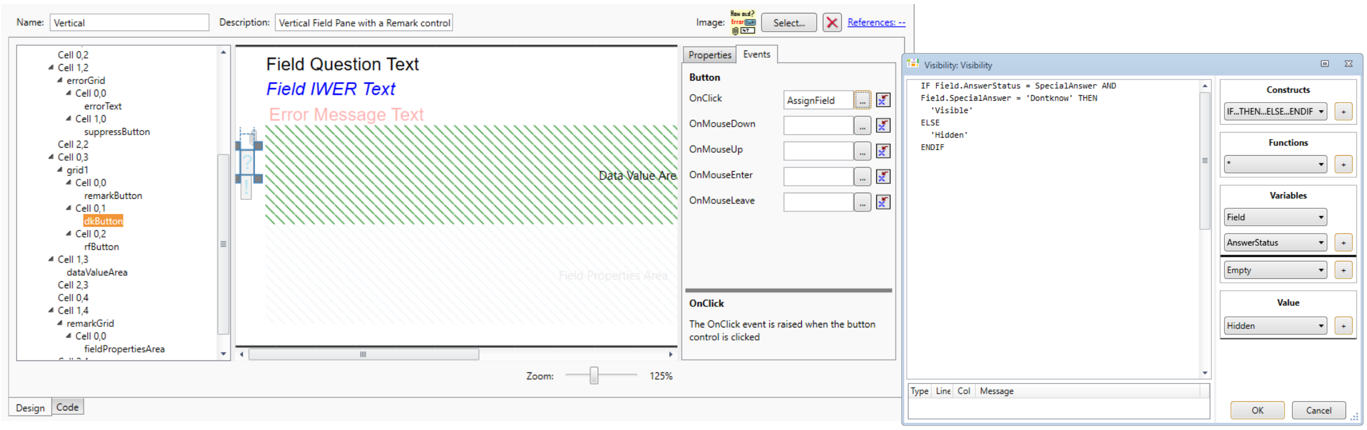
Figure 3: Previous waves tools

3. Migration

Many of these tools used the Blaise API or interfaced with the questionnaire's database, and might have to be adapted. We first draw up a migration plan to set up SHARE Blaise 5 CAPI, so we could test out how feasible a migration would be:

1. Import wave6/wave7 questionnaire; several items from the wave 6 and wave 7 questionnaires reoccurred in the wave 8 questionnaire. We imported them in Blaise 5.
2. Update to wave8 questionnaire; in the Blaise 5 environment, we adapted the imported questionnaire to the wave 8 definition.
3. Design default interface; we understood the questionnaire layout as used under previous waves, with the split screen, was no longer available, we had to build an interface that still supported several 'old' features, like keyboard navigation.
4. Implement non default features; in SHARE there are some questions that need counters, show wordlists or need more complex lookup tables, in phase 4 these were addressed
5. Export to TMT; the export to the TMT needed to be redefined, we decided to review this once we knew which of the features would be possible or when it would be clear if there were new concepts in the questionnaire.
6. Import from TMT; in earlier waves we copy pasted the translated texts in the source code; ideally there would be better processes we could use this time to generate translated questionnaires.

The import of the previous wave's questionnaire worked without any problems. And scripting the questionnaire worked perfectly. The new environment was useful. It sped up development time, the background parsing, helped finding problems immediately. Since the interviewers were used to using the split screen, which had the navigation paths displayed on the bottom of the screen, we tried our best to mimic this in the Blaise 5 environment ourselves. This failed; we tried several ideas but concluded it was not feasible. Another feature the interviewers liked were the icons that showed when a 'DontKnow', 'Refusal' or 'Remark' was attached to a question. We added these icons to a field pane, and determined their visibility on the status of the field.



Name: Vertical
Description: Vertical Field Pane with a Remark control

OnClick: AssignField

OnClick
The OnClick event is raised when the button control is clicked

Visibility: Visibility

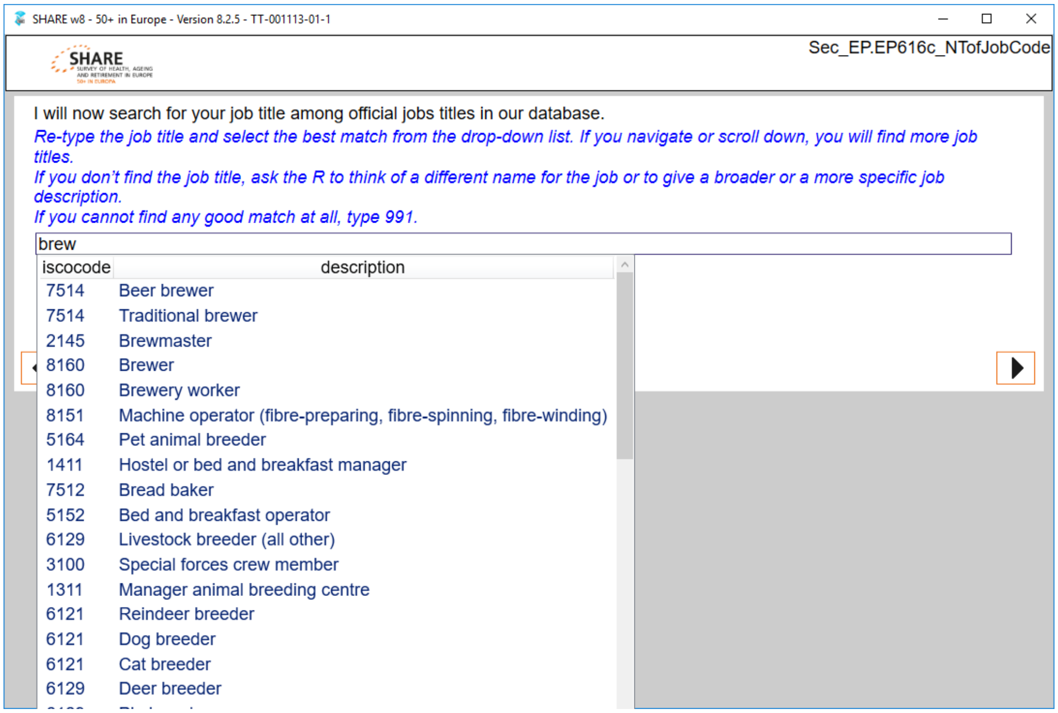
```
IF Field.AnswerStatus = SpecialAnswer AND
Field.SpecialAnswer = 'Dontknow' THEN
  'Visible'
ELSE
  'Hidden'
ENDIF
```

Constructs: IF...THEN...ELSE...ENDIF
Functions: *
Variables: Field, AnswerStatus, Empty
Value: Hidden

Figure 4: DK and RF

Several non-default features needed to be explored:

- Jobcoder; SHARE invested a lot of effort in setting up large databases of already classified job titles; in previous waves we wrote external apps to show these, so we could have full control over the algorithms and behavior. We hoped in Blaise 5 the way lookups behaved would be improved, to allow for alternative algorithms, and we hoped to show a list that also accepts whatever is typed in, so we could in go detect that there was no match, and still collect the response. This was unfortunately not possible, we ended up using the lookup as they are available in Blaise 5 with trigram search, but still hope this will be further improved.


SHARE w8 - 50+ in Europe - Version 8.2.5 - TT-001113-01-1
SHARE
SURVEY OF HEALTH, AGEING AND RETIREMENT IN EUROPE
50+ IN EUROPA
Sec_EP.EP616c_NTofJobCode

I will now search for your job title among official jobs titles in our database.

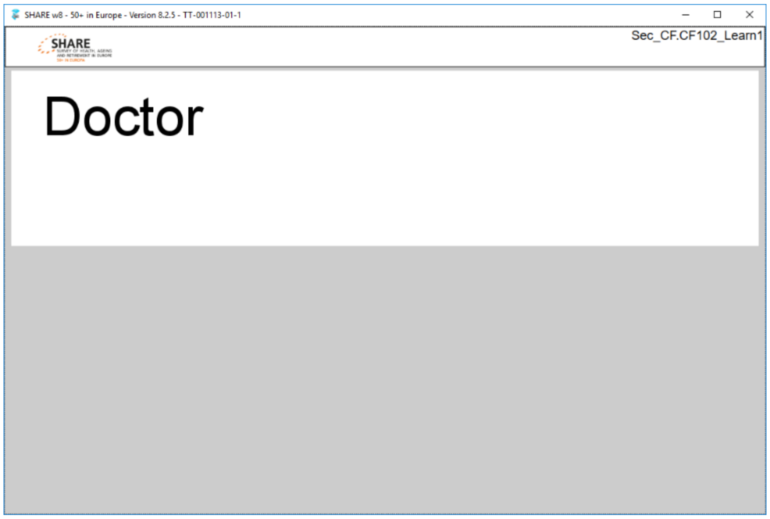
Re-type the job title and select the best match from the drop-down list. If you navigate or scroll down, you will find more job titles.
If you don't find the job title, ask the R to think of a different name for the job or to give a broader or a more specific job description.
If you cannot find any good match at all, type 991.

brew

iscocode	description
7514	Beer brewer
7514	Traditional brewer
2145	Brewmaster
8160	Brewer
8160	Brewery worker
8151	Machine operator (fibre-preparing, fibre-spinning, fibre-winding)
5164	Pet animal breeder
1411	Hostel or bed and breakfast manager
7512	Bread baker
5152	Bed and breakfast operator
6129	Livestock breeder (all other)
3100	Special forces crew member
1311	Manager animal breeding centre
6121	Reindeer breeder
6121	Dog breeder
6121	Cat breeder
6129	Deer breeder

Figure 5: Lookup table

- Wordlists/Counter; In SHARE, there are some cognitive measures, like a timed word recall, questions where a timer and a stopwatch are shown. In previous waves videos were used to display these. Unfortunately, at the time we developed the wave eight questionnaire videos were not yet supported in Blaise 5. We invested quite some time in using the timer, somehow feeding an array of texts to the timer that would then change the text of a label at each tick. This however did not work, changing labels at runtime via an action is not possible, and this also makes it impossible to develop a feature like a stopwatch. We ended up using animated GIF files, and disable navigation, this solution was not ideal since the GIF restart when enter is pressed. If only the Text property of labels could be changed by clicking a button or by a tick from the timer, we would have developed this differently.



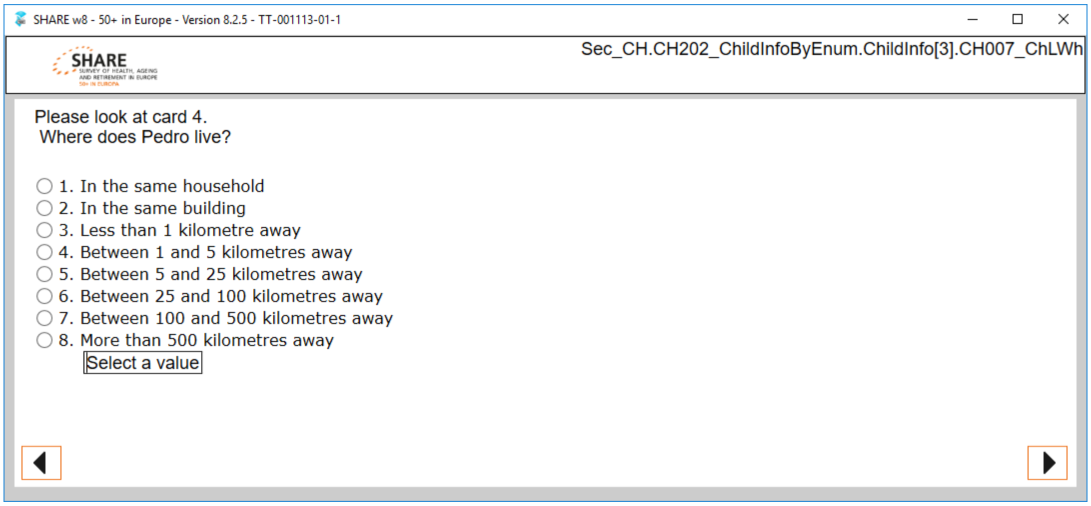
SHARE w8 - 50+ in Europe - Version 8.2.5 - TT-001113-01-1

Sec_CF.CF102_Learn1

Doctor

Figure 6: Disabled navigation while word recall list is playing

- Keyboard navigation; a key functionality, the complete questionnaire is in CAPI mode, the interviewers work quicker if they do not need to point and click. In version 5.4, this was not yet implemented. To get this working we introduced enumerationTextBoxes and setTextBoxes, which got the focus when a field pane is activated, one could type in the responses and the attached checkboxes or radio buttons would be checked. In addition, when the interviewer checks the radio buttons or checkboxes, the textboxes will show the value, like in earlier waves.



SHARE w8 - 50+ in Europe - Version 8.2.5 - TT-001113-01-1

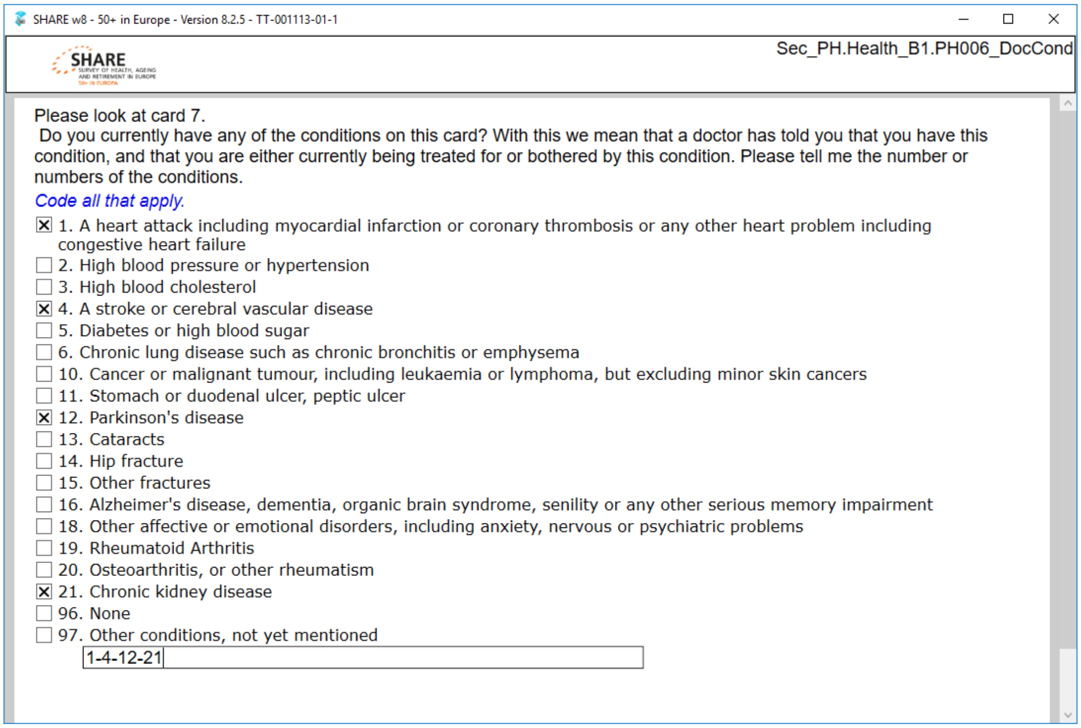
Sec_CH.CH202_ChildInfoByEnum.ChildInfo[3].CH007_ChLWh

Please look at card 4.
Where does Pedro live?

○ 1. In the same household
○ 2. In the same building
○ 3. Less than 1 kilometre away
○ 4. Between 1 and 5 kilometres away
○ 5. Between 5 and 25 kilometres away
○ 6. Between 25 and 100 kilometres away
○ 7. Between 100 and 500 kilometres away
○ 8. More than 500 kilometres away

Select a value

Figure 7: When nothing is entered, the setTextBox shows the message ‘Select a value’


SHARE w8 - 50+ in Europe - Version 8.2.5 - TT-001113-01-1
SHARE
SURVEY OF HEALTH, AGEING AND RETIREMENT IN EUROPE
50+ IN EUROPA

Sec_PH.Health_B1.PH006_DocCond

Please look at card 7.
Do you currently have any of the conditions on this card? With this we mean that a doctor has told you that you have this condition, and that you are either currently being treated for or bothered by this condition. Please tell me the number or numbers of the conditions.

Code all that apply.

☒ 1. A heart attack including myocardial infarction or coronary thrombosis or any other heart problem including congestive heart failure
☐ 2. High blood pressure or hypertension
☐ 3. High blood cholesterol
☒ 4. A stroke or cerebral vascular disease
☐ 5. Diabetes or high blood sugar
☐ 6. Chronic lung disease such as chronic bronchitis or emphysema
☐ 10. Cancer or malignant tumour, including leukaemia or lymphoma, but excluding minor skin cancers
☐ 11. Stomach or duodenal ulcer, peptic ulcer
☒ 12. Parkinson's disease
☐ 13. Cataracts
☐ 14. Hip fracture
☐ 15. Other fractures
☐ 16. Alzheimer's disease, dementia, organic brain syndrome, senility or any other serious memory impairment
☐ 18. Other affective or emotional disorders, including anxiety, nervous or psychiatric problems
☐ 19. Rheumatoid Arthritis
☐ 20. Osteoarthritis, or other rheumatism
☒ 21. Chronic kidney disease
☐ 96. None
☐ 97. Other conditions, not yet mentioned

1-4-12-21

Figure 8: Keyboard navigation to select multiple response options

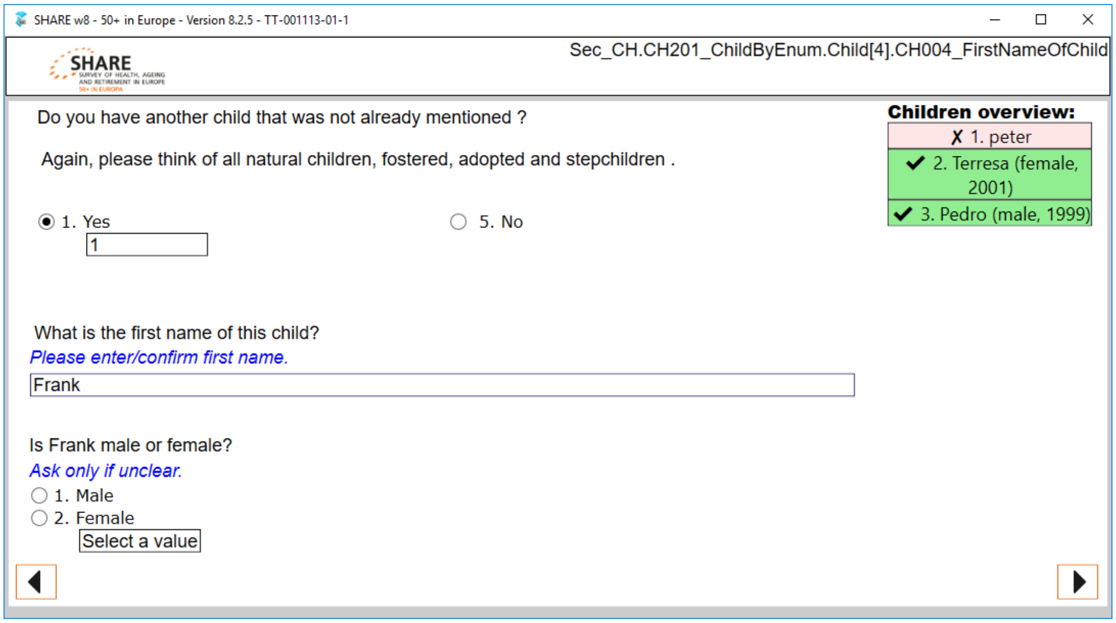
- Textroles; We like the text roles features very much, in earlier waves of SHARE this was already done implicitly, interviewer instructions and questions text were already defined as separate translatable items and during import we used layout we gave them different fonts and colors,.Now we can easily set this up.

- Unicode; Under Blaise 4 CentERdata developed several hacks to support the SHARE questionnaire in non-Western scripts like Arab, Hebrew or Russian. These hacks cost us frustration, and needed thorough checking, especially for fills that were cut off at 256 characters without any error message when fed through procedures. We really appreciate the full support of Unicode in Blaise 5.

- Images; Wave 8 of SHARE had some questions that used images on show cards, since the use of images was so straightforward, we also show them in the responses the interviewer sees.

- Child grid; since the split screen as common in Blaise CAPI questionnaires in Blaise 4, is not implemented in Blaise 5, it was not convenient to present a full overview of all children in a household, or at least not one compatible with the previous questionnaire routing and field. We chose to rephrase the questions, and develop a child overview on screen as a separate area in the field pane.

- We match children that were preloaded from a previous wave or from the responses of the partner, and children we may have detected in the Social Network section, and possibly add children that are not already mentioned. In this interface, the list of children is shown on the right, and will change to green and will add a check mark when a child is confirmed, if one of the children is mentioned twice or should not be in the list, they are colored red and a cross is added.



SHARE w8 - 50+ in Europe - Version 8.2.5 - TT-001113-01-1

Sec_CH.CH201_ChildByEnum.Child[4].CH004_FirstNameOfChild

Do you have another child that was not already mentioned ?

Again, please think of all natural children, fostered, adopted and stepchildren .

● 1. Yes ○ 5. No

1

What is the first name of this child?

Please enter/confirm first name.

Frank

Is Frank male or female?

Ask only if unclear.

○ 1. Male
○ 2. Female

Select a value

Children overview:

✗ 1. peter
✓ 2. Terresa (female, 2001)
✓ 3. Pedro (male, 1999)

Figure 9: Children overview

The keystrokes are different from before. We gained better insight in loading times on question level. This created quite the discussion, since the definition of interview duration is depending on this, do we measure the time a question is shown on screen, or also include the time between when the response is submitted and the next question is shown, and to which question belongs that time, maybe not relevant for most surveys, but for a survey that may take on average an hour, this definition can be quite relevant.

In earlier waves, we would use the API to walk through a compiled questionnaire, and determined the questionnaire structure from that, for example to import into the translation environment. Since Blaise 5 exports XML definition we appreciate that, no need to call the API, simply parse XML made our systems more stable, also paper version interface, Stata scripts, various excel files with metadata overviews.

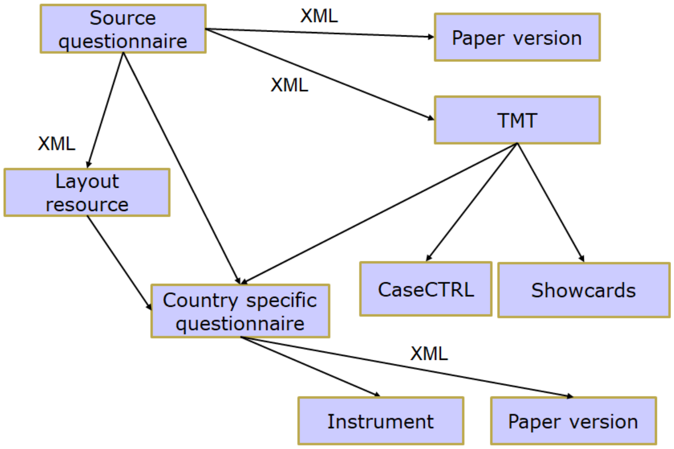


Figure 10: Use of Metadata XML

An extra feature we hoped for was to support the use of tablets for the SHARE system. Since Blaise 5 promised to support this, we knew that our SMS system would become a bottleneck in any future support. To overcome this a new SMS was developed, under the name CASE CTRL, this light weight program, can run on any device, it has a web interface and can link to many questionnaire engines. Linked to this we also replaced the SD in our systems and replaced it with Sample CTRL. This resulted in the tools described in figure 11.

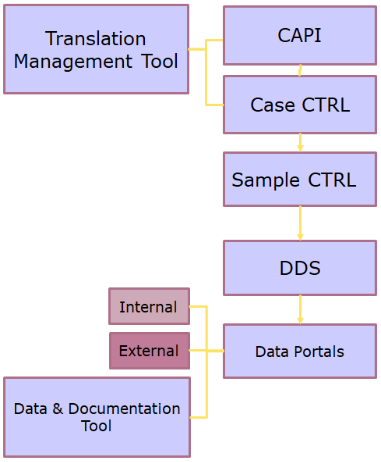


Figure 11: SHARE wave 8 tools

This resulted in the following changes for the SHARE wave 8 architecture:

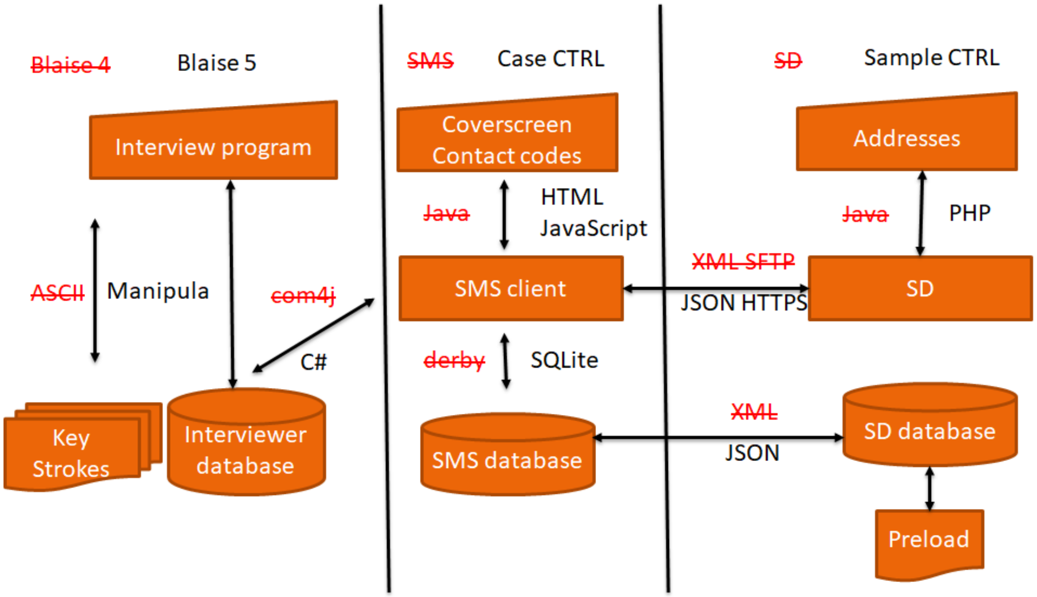


Figure 12: Changes for wave 8

4. Conclusion

The Blaise 5 environment is greatly improved compared to the Blaise 4 system. Its programming interface works much faster, which is convenient when developing a complex, long questionnaire like the SHARE questionnaire. The separation of layout and questionnaire definition is very useful. The introduction of text roles and the compatibility with Unicode solved many problems we had before.

There are some limitations on what we could do, which sometimes feels like they could be made possible if things were a bit more open. The burden of creating complete apps for some features is sometimes a bit overdone when you are very close to a solution within the Blaise environment. On the other hand, maybe we targeted the problems from a wrong angle.

The metadata XML export really gave us a solid reusable definition we used in various exports in pdf, html, excel and Stata script format. It would be perfect if from this definition a questionnaire could be build. This would allow us to generate instruments as an automatic integral part of the translation cycle.