

Converting Social Survey Blaise 4 questionnaires to Blaise 5; The Challenges of Multi-lingual Questionnaires and Challenging Collection Environments

Andy Watson and Steve Maurice, Office for National Statistics United Kingdom

1. Abstract

The Office for National Statistics (ONS) Social Surveys Division (SSD) runs numerous Longitudinal, annual and ad-hoc surveys. Most of these surveys are conducted as face-to-face interviews on Blaise 4. The decision has been made to upgrade these surveys to Blaise 5 so that future surveys can more easily incorporate mixed mode elements such as Computer Assisted Web Interviewing (CAWI) and Computer Assisted Telephone Interviewing (CATI).

The Blaise 5 uplift project was initiated to transform all the SSD surveys from Blaise 4 to Blaise 5 and all the associated legacy support systems. While collecting data via Blaise 4 ONS have used case management software 'Casebook', developed and maintained in-house. Now that we're moving to Blaise 5, we're looking to retire Casebook and rely more heavily on "out of the box" features of Blaise 5.

In this paper we report on our experiences moving the National Survey for Wales (NSW) and the International Passenger Survey (IPS) from Blaise 4 to Blaise 5.

1.1. The National Survey for Wales

The NSW survey is a study run by Office for National Statistics on behalf of the Welsh Government. The study gathers information on many topics including health, schools, sports, arts and culture. The statistics are used by Welsh Government, Sports Wales, Arts Council Wales, Natural Resources Wales, Welsh local authorities, charities and academics. They are used to inform progress against key indicators of the Future Generations Act, to inform progress on existing or the shaping of new policies within Wales.

There is now a follow up survey, the Welsh Language Usage Survey (WLUS) in the form of a paper questionnaire distributed by interviewers to eligible respondents at the end of NSW interview. This questionnaire is about the Welsh language. It is to establish if a respondent, who has indicated they may be able to speak Welsh in the main questionnaire, can speak Welsh, and if so, how often, where and when they use it.

1.2. The IPS

The IPS is a continuous survey that has been conducted by ONS since 1961. Between 700,000 and 800,000 interviews are conducted every year at major airports, seaports and tunnel routes covering both departures and arrivals terminals. This is the main source of information used by the UK governing bodies for planning, monitoring and informing decisions on tourism and immigration policies. Having successfully moved IPS data collection from paper to tablet^[1] using Blaise 4 as the collection tool^[2], the Office for National Statistics (ONS) are now ready to move collection from Blaise 4 to Blaise 5 so collection can continue as Blaise 4 is phased out.

2. Behind the surveys

2.1. NSW

The NSW was started in 2016 and has covered a variety of subjects. The questionnaire object is currently written in Blaise 4.8 and collected face-to-face. The Welsh Government stipulated that the questionnaire be available in both English and Welsh for all respondents. A specification was provided by Welsh government, in English with all questions and routing specified. The questionnaire utilised some of the standardised ‘Core blocks’ from other Social surveys already created. When this completed single language version of the questionnaire was finalised, a separate specification was then provided by Welsh Government with all questions and routing translated into Welsh.

Table 1: NSW Selected Language of interview

Language	Frequency	Percent
English	11,756	98.6
Welsh	159	1.3
Other	7	0.1
Total	11,922	100.0

2.1.1. NSW Methodology

The annual sample size for the NSW is 12,000 face-to-face interviews with a required response rate of over 56%. This sample is split into 12 field periods over the year. Each year Welsh government request different subject modules for the questionnaire, depending on their requirements for informing decisions to be made on new policies. Due to the number of modules created each year, the length of the survey can differ greatly. To keep a level of consistency, some modules are ‘Sub-Sampled’ meaning that certain modules only get asked to a randomly selected percentage of the main sample.

2.1.2. NSW Collection

Data collection for the NSW is done face-to-face by field force interviewers in the homes of the respondents. Advanced letters are sent to all sampled respondents and follow up contact is made by the interviewers to arrange a convenient time for the interview. An In-house case management system, Casebook, is used to manage all the interviewer’s cases every month.

2.2. IPS

2.2.1. IPS Methodology

The IPS is conducted at various Airports, Seaports, and at the Eurostar/Eurotunnel. The survey is not pre-sampled as it’s impossible to know who’ll pass through the ports in advance. Each location has a different sampling interval for each direction of travel, e.g. At Gatwick it might be that every 20th person to pass is selected and every 40th for departures. A new form is created for each attempt at an interview.

The questionnaire takes two different routes according to subsampling. One route being the full questionnaire with all questions asked (full form), the other being a much shorter route with the potential to turn into a full questionnaire if certain criteria are met (filter form). As with the sample intervals, each port has a different sub sampling for both arrivals and departures e.g. At Gatwick arrival shifts might be 2 out of 5 presented with the full form, while the other 3 are presented with the filter form. In a shift with both full and filter forms, research were keen for the order these forms are presented to be random, e.g. for a 2 out 5 shift, the first form of every group of 5 forms has an equal chance to be full or filter.

All sample intervals and sub samples are decided ahead of collection by the research team and stored in a Blaise database to be used by the questionnaire as a lookup.

2.2.2. IPS Collection

As IPS interviewers are standing in busy ports and need to be mobile to approach those sampled, it was decided that interviewers would collect using tablets rather than laptops. Tablets are far lighter and more suited to longer shifts so the questionnaire had to be designed to work well on a Tablet^{[1][2]}



IPS data are collected offline due to a lack of available connectivity at the ports our interviewers operate in. In Blaise 4 currently the questionnaire is downloaded, at home before the shift is due to start, via Casebook and is launched offline via Casebook. The data for each interviewer on each shift is packaged into a randomly named compressed file and transmitted back to the office where it is smashed into a master file. Reports are created from this file, flagging cases that interviewers have indicated may need additional editing and a team of coders and editors work through each item resolving issues.

3. NSW Questionnaire

The questionnaire is written in in Blaise 4.8 in a Computer Aided Personal Interviewing (CAPI) format, with the interviewer leading the respondent through the questionnaire. In the NSW there is a section of more personal question blocks about religion and sexual orientation, where greater discretion maybe required. These blocks have been designed in such a way that they are available in a Computer Aided Self interviewing (CASI) format, as well as the standard CAPI format should the respondent feel they are unable to complete the section unaided.

3.1. Current Multilingual elements on the NSW

In the current Blaise 4 version of the NSW questionnaire, the multiple languages are handled by the Languages setting in the data model. The NSW questionnaire has three languages; English (ENG), Welsh (CYM) and the Help language (HLP). The language can be set at the beginning of the questionnaire, asking the respondent which language they would like to complete the interview in. The interviewer can use the next form language in the menu to toggle between the languages in case reference to the other language is required. This does not change the language permanently however, when the questionnaire is moved on to the next question, it reverts to the 'Set' language.

3.1.1. Character Encoding

Most of the characters in the Welsh alphabet are 'standard', apart from all the vowels can have a circumflex (to bach) as well as the letters 'w' and 'y' which can also have a circumflex. Examples of the special characters - â Â, ê Ê, î Î, ô Ô, û Û, ŵ and ŷ

To enable this within Blaise 4.8, UTF-8 encoding must be applied to the language in the projects datamodel properties.

The development environment, the Blaise 4 control centre is not capable of receiving these special characters directly from a copy and paste. A third-party text editor has to be used to open the block requiring the special characters, the text inserted in the correct location, saved and then re-opened with the Blaise control centre.

3.1.2. The Helpfile

The NSW questionnaire has an accompanying helpfile, defined as language in the datamodel and the accompanying Q-by-Q help file specified in the Mode Library file, this also must be available in both languages. There are 2 separate pages for each help page, one for each of the available main languages. To get to the correct page, once a language has been selected, a text fill has been utilised that can be changed in the rules for each question that has a Q-by-Q help page. This text fill is then the address for the page within the Q-by-Q helpfile, according to the language required.

3.2. Converting those Multilingual elements into Blaise 5

The same approach can be taken in Blaise 5 as in Blaise 4, by creating languages within in the datamodel and then having multiple entries for each question, description or message. When using the conversion tools within Blaise 5 converting from 4 to 5 these language options are converted directly across and no additional work is required to enable them. This solution is just as limited as the Blaise 4 however, in that the multiple languages all must be defined for each questionnaire object at the datamodel level. There is no re-usability to the additional settings and the encoding

```
IF (QLanguage.LangInt = CYM) THEN
  EduHelp := 'EducationWelsh'
ELSEIF (QLanguage.LangInt = ENG)
  EduHelp := 'Education'
ENDIF
```

3.3. The future Blaise 5 solutions

With the onset of the recent upgrade from Blaise 4 to Blaise 5, ONS has been developing its own standardised Resource Database. Creating ONS themed templates and styles that can then be used across different resource sets. The ability to create a suit of layouts for the standard questions within ONS should make vast improvements to the flexibility, transferability, adaptability, consistency and quality of our questionnaires.

Creating a multilingual questionnaire in Blaise 5 there appears to be a great many more tools to aid the user in doing so. The languages can be added to the datamodel in the same way as in 4, but now there is a Datamodel text editor, and Language selector control and you can define a culture for each language. For the definition of each language you are allowed to define language specific Font Definitions, Styles, Media, and Texts. These can then be mapped from the resource Database to whichever datamodel it is applied. The culture can be defined by a language or the country. For example, English-United States, French-France, French-Belgium. The culture is used to represent the following attributes aligned to a specific culture or language such as numbers, currency, dates and times. The defined culture can even be used to alter the direction in which texts are presented on the screen either right-to-left or left-to-right. NB: The Q-by-Q helpfile is a challenge yet to be resolved!

4. IPS Questionnaire

As the IPS has already gone through a long transformation project moving from paper to tablet our goal is to move from Blaise 4 to 5 changing as little as possible, to keep interviewer training to a minimum.

As with the work on NSW, converting the Blaise 4 IPS questionnaire into Blaise 5 was relatively straight forward. The source files and existing lookup files were converted to Blaise 5 using the ‘Source Converter 4 -> 5’ and ‘Data Converter 4 -> 5’ applications that are provided with Blaise 5.

There were some issues that needed to be resolved before the questionnaire could be compiled and run, and a few things that interviewers had requested during development of the Blaise 4 instrument which had to be considered while developing in Blaise 5. These main challenges are detailed in this section.

4.1. Current IPS complexities in Blaise 4

4.1.1. Subsampling

In the Blaise 4 questionnaire, subsampling is carried out by reading the previous form. This is done using a technique we call ‘double compiling’ because it requires that the questionnaire be compiled twice, first without referencing the questionnaire to create an initial metadata file (BMI), followed by a second time once reference to that BMI has been added. As Blaise 4 doesn’t check to see if a database exists when compiling, we can reference the database file (BDB) that’s created when the questionnaire is run, allowing us to search the BDB with the Primary Key minus 1 to read the previous forms data. Each form holds fields which act as a current total of full/filter forms, as well as group sizes and target amounts for full/filter forms, so reading the previous forms data allows us to determine which type the current form should be.

4.1.2. Creating a new form without losing shift information

Running the questionnaire through Casebook allowed us to first ask interviewers to complete a shorter “shift details” questionnaire to fill in key shift information that can be read in by each form of the main questionnaire. This saves the interviewer from having to re-type the same information multiple times but creates extra databases containing data duplicated in the main questionnaire that in turn are stored on our servers by case management processes. It’s been a goal of the office to avoid having a separate shift information questionnaire when moving to Blaise 5 as it leads to lots of un-necessary files in several locations.

4.1.3. Duplicate case identifiers.

In the Blaise 4 questionnaire, shift information is keyed in manually by the Interviewer. The key variables are Shift Number, Form Number and Interviewer number which make up the primary key, and Port Route which describes which port the interview took place in. As Casebook gives each set of data a random name when packaging it for transmission back to the office, if any of these details are incorrect it doesn’t matter as the data are set aside and not smashed until the issues are resolved by the coding and editing team. This prevents data with an incorrect primary key overwriting other legitimate data.

4.1.4. The navigation panel

One of the features of the Blaise 4 instrument that Interviewers gave a lot of feedback on was the navigation panel which was designed in the Blaise Menu File (.BMF). During development we reduced the number of panels, added the option for a panel on the other side of the tablet for left-handed interviewers and removed the flags as some were offended by having to select certain flags for language that didn't match their country. This panel was the most contentious part of Tablet development and it was difficult getting a consensus amongst interviewers how best to implement it.

4.2. How these complexities were handled in Blaise 5.

4.2.1. Subsampling

We have been unsuccessful replicating the 'double compile' process using Blaise 5 so struggled with this while working with early versions of Blaise 5.

However, once the 'Actions Setup' feature was added in a later version, allowing a Manipula procedure to be called during the running of a questionnaire, the subsampling problem was easily overcome.

On the first form once the shift information has been filled in a Manipula procedure is called that writes out all the necessary data a separate file. That file is then searched and read in when the next form is running so that code in the rules can dictate whether it should be a Full or Filter form.

4.2.2. Creating a new form without losing shift information

As with the subsampling issue, when the 'Actions Setup' feature was added, this became a very easy problem to solve and we were able to get rid of the separate shift details questionnaire completely.

First we established how many of the 167 fields in the shift details questionnaire were essential to the main questionnaire.

There were only 8 which were essential, so when the form number is 1 these data are collected and then written out to the same file that the subsampling information is written out to (as described in 4.1.1). These values are read in at the key page of the next form to save the interviewer some typing (increasing form number by 1) and throughout the questionnaire as needed.

4.2.3. Duplicate case identifiers

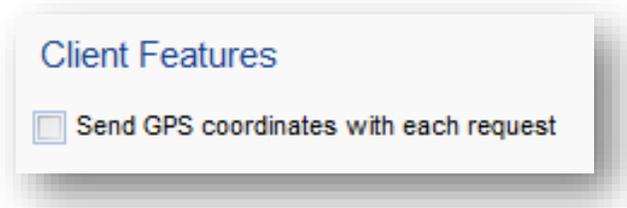
As it's vitally important that the key shift information fields are accurate in Blaise 5 to prevent data being overwritten, we decided quite early on in Blaise 5 development that we'd use lookups for everything. The research team already had two separate files containing...

- a) Shift numbers, Port routes and Shift dates
- b) Interviewer numbers, names and the port they work at.

These files were converted into lookups in Blaise 5 and are used at the Primary Key pages to fill in the information. The first Primary Key page uses lookup 'a', to set the Shift number and port route, and Shift date. The second Primary Key page displays the information that's been selected on the first page, highlighting the shift date in red if it doesn't match the system date asks the interviewer to confirm that these details are correct by selecting their own name in lookup 'b'.

Unfortunately, due to unforeseen circumstances, shifts are sometimes moved around to different dates during the month, so we have to allow for shift dates that don't match the system date.

In the future we'll consider using the GPS features of Blaise 5 to set the Port details and further reduce error:



This will be dependent on hardware as we're aware that a new Tablet device is currently being considered. We will likely enable GPS-recording at first when a survey month is live to see what range of co-ordinates we get back for each port before automating the process.

To ease the transition from Blaise 4 to 5, we've subsequently implemented the Interviewer Name lookup in Blaise 4.

4.2.4. The navigation panel

The navigation panel was quite tricky to try and resolve. In one of the earliest versions of Blaise 5 we were able to recreate the 'Keyboard action' of the Blaise 4 menu file by adding a button to the page template in Blaise 5 that, for instance when pressing the '2' button, assigned the active field a value of: `'Page.ActiveField.ValueAsText + '2'` followed by a save.

This unfortunately was dependent on a bug that was fixed in a later version of Blaise 5. Since that fix we've been unable to recreate the Keyboard Actions of Blaise 4, but there's really no need. Our interviewers are already using the On-Screen Keyboard (OSK) that is part of Windows 10 for text input, and it is very straight forward to change a setting and make this appear with a Number-Pad. Showing the interviewers how to turn on the Number-Pad allows us to free up valuable screen space which would otherwise have been taken up by the clunky and un-necessary navigation panels.

5. Summary/reflections

Overall, moving the IPS from Blaise 4 to Blaise 5 has been very straightforward. The Data and Source convertors provided with Blaise 5 are excellent and save what could have amounted to many hours of work.

Moving the NSW from Blaise 4 to Blaise 5 has been similarly straightforward and the research team, field interviewers and Welsh government have all been positive in their feedback when consulted. The 2020 NSW questionnaire was produced in parallel in Blaise 4 and 5 to provide a comparative test for all concerned, as well as a development and testing plan for actual go-live in 2021.

The transition has been easier for IPS and NSW than some other surveys due in no small part to the hard work David Kinnear (ONS) has put in to make a very good Resource Database for our team to use. The resource database is full of applicability conditions that have meant all fancy layout features and field panes for all types of questions are applied automatically. That enabled us to focus purely on issues described in this report, and other "nice to have" improvements.

Being in a current state of transitioning from Blaise4 to Blaise5 gives a distinctly unique viewpoint. It is a good opportunity to review what worked well in the old system and what we would like to see in the new

system. What didn't work well and should be left behind. What we look forward to using in the new system, features that we feel we would benefit greatly from. As well as those new available features that we are not so sure about yet, but in keeping an open mind we may be convinced are an improved method or system.

6. Links and References

[1]. Lan Benedkit. Developing the UK Passenger Survey in Blaise 5 on Tablet Computer, Office for National Statistics, UK, Statistics. IBUC 2015, 16th International Blaise Users Conference

[2]. Mike Hart. Converting Paper Collection on the International Passenger Survey to Blaise 4 on a Tablet Computer, Office for National Statistics, UK, Statistics. IBUC 2016, 17th International Blaise Users Conference

<https://gov.wales/national-survey-wales>