ONS Web Development Project: Tackling the Difficulties of Social Survey Data Collection

Lucy Fletcher, Office for National Statistics

1 Introduction

In the context of challenging UK public sector efficiency targets and the increasing cost of traditional survey data collection methods (such as face-to-face interviewing), there is considerable interest in developing a capability for internet data collection on the social surveys run by the Office for National Statistics (ONS).

ONS aims to be an innovative, cost effective and considerate (in terms of respondent burden) producer of official statistics. As such, it is placing increasing emphasis on developing mixed mode survey designs, including exploring the use of the internet for data collection, to improve efficiency and reduce the burden on households who respond to its surveys.

2 Background

The Labour Force Survey (LFS) is one of the UK's largest and most important social surveys and a key ONS statistical source for economic and labour market measures. Through face-to-face and telephone interviewing, it collects a wide variety of questions ranging from employment to health and income. In terms of time and complexity, it is probably one of the most burdensome surveys for respondents. However, the data collection process is also one of the most costly within the ONS Social Survey portfolio.

Due to the high costs and complexity, plus also increasing pressure on the office to be innovative with its approach to collecting and disseminating statistics, the LFS was an ideal candidate for a web survey pilot.

In 2010, Business Statistics Division (BSD) of ONS obtained a corporate licence for Confirmit Professional, the web survey authoring software aimed at the market research sector. BSD collects data from businesses through paper questionnaires and telephone data entry, and was put under pressure to trial web based collection.

While BSD piloted an online version of their Capital Expenditure Survey, Social Survey Division (SSD) was able to share the licence, and also start piloting work. Although Blaise would have been the preferred software, as this is what is used for all ONS's social surveys, at the time there was no funding for setting up the necessary hardware for running live questionnaires. SSD carried out two LFS web pilots using Confirmit, and further work is in progress using Blaise IS.

This paper will outline some of the piloting work already conducted, and discuss the current project to test the feasibility of adding a web collection mode to one of ONS's most complex surveys.

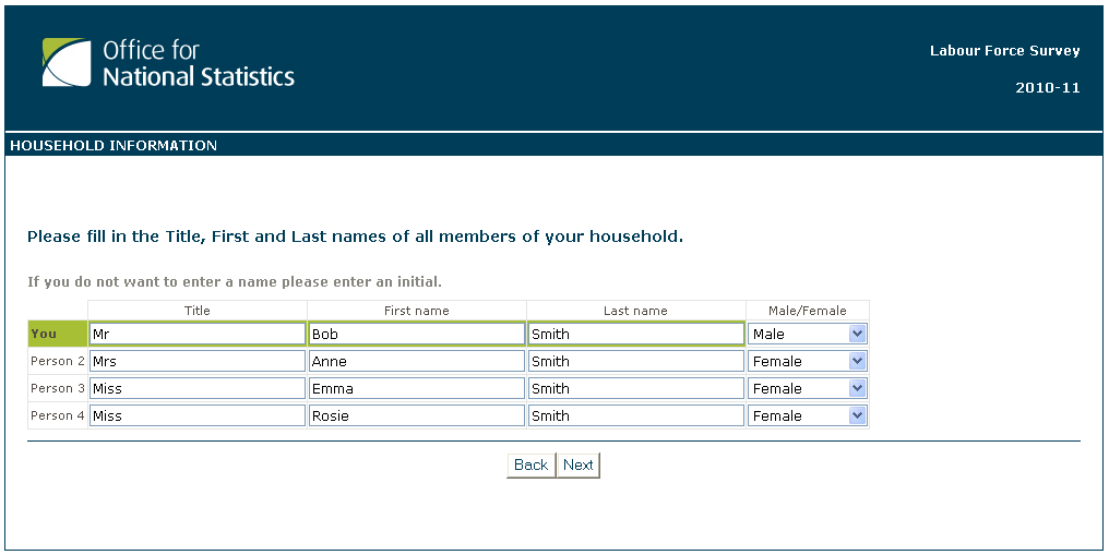
3 Completed LFS web pilot work

3.1 First LFS web pilot - 2011

In 2010 the Blaise Development and Support Team (BDSS) in Social Survey Division developed a questionnaire using Confirmit for the first Labour Force Survey web pilot. With around 200 questions, this was a shortened version of the LFS questionnaire that is used in face-to-face and

telephone modes. Only the key topics were programmed, such as employment and earnings, and some questions and onscreen guidance were modified to make them clearer to understand.

In addition, the online version was only administered to one member of the household, whereas the normal LFS is administered to all household members aged 16 or over (in person or by proxy). We didn't manage to produce a workable household questionnaire in Confirmit where members could record their own details in sequence, so we proceeded with an individual-level questionnaire. The questionnaire did however include a short household section, collecting basic socio-demographic information about each household member (see screenshot 1), and their relationships to each other.



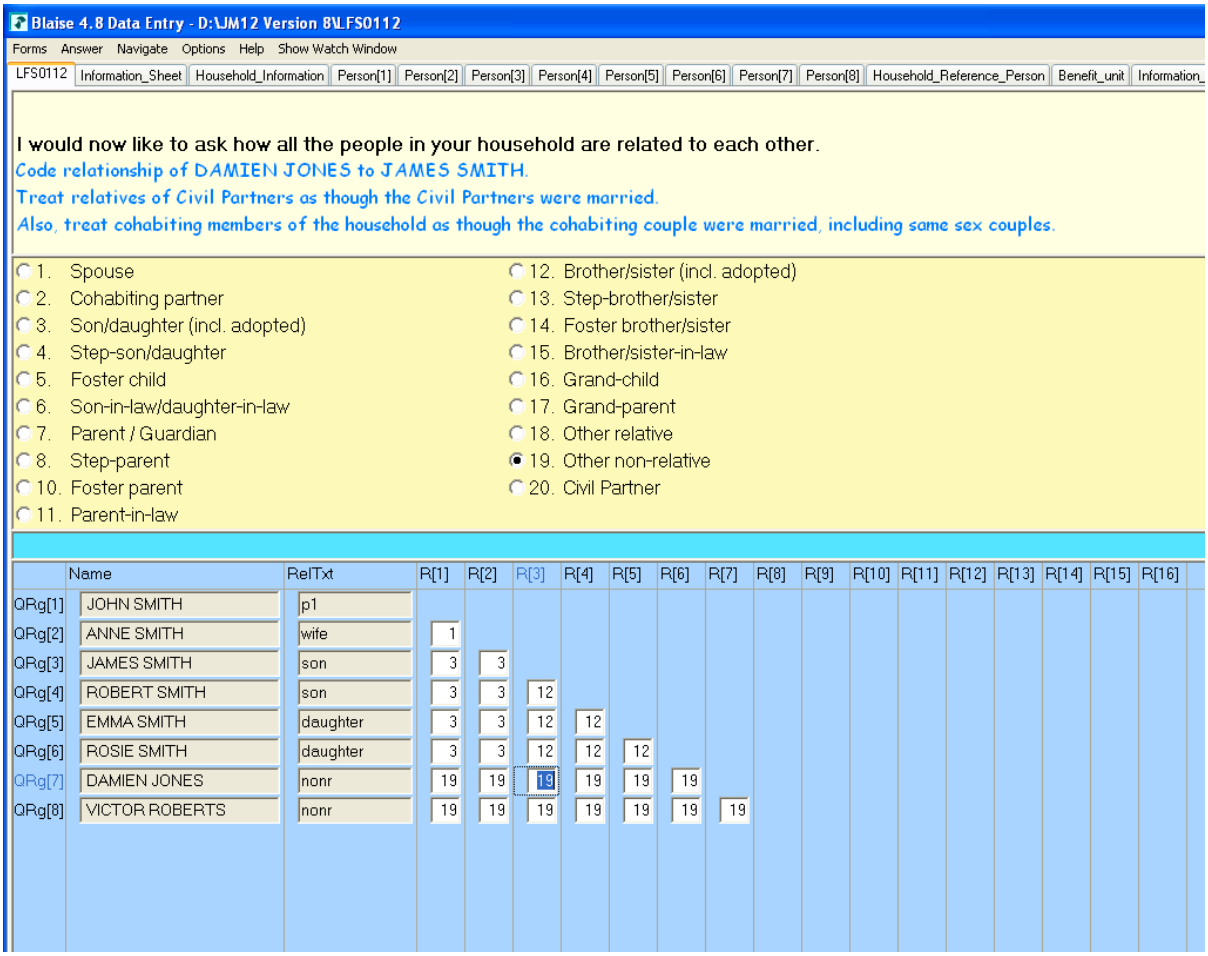
Screenshot 1: Collecting demographic data in a Confirmit web interview

The main aims of the pilot were to examine the following:

- Respondents' ability to answer LFS questions, and complete a whole interview;
- Respondents' attitudes to responding to the online survey;
- Internet survey design and usability, including presentation of questions and answer lists.

1,424 respondents who had previously taken part in the LFS were invited to take participate. The fieldwork took place in early 2011, and a quarter of respondents completed the full survey. On average, the online questionnaire was completed in 18 minutes. A further cognitive exercise was carried out with 21 participants (from a different sampling frame) to gather further feedback about the online facility. These respondents were observed completing the questionnaire to assess how well they were able to navigate through it, understand the questions and provide meaningful answers.

In general, we found that both sets of respondents had no problems completing the online questionnaire, and said they would prefer this method in future to a face to face or telephone interview. They did struggle however with some existing LFS questions which they found confusing, despite being familiar with the LFS already. This demonstrates that we face a big challenge in adapting the questionnaire to suit a mixed mode environment, but without losing the meaning of questions and compromising data quality.



Screenshot 2: The household relationship grid is one of the challenges that lies ahead when adapting the interviewer-led instrument for a web mode

In order to develop a full online LFS, further investigation needed to be carried out to assess acceptable and optimal questionnaire lengths. As we know from this research that respondents are willing to fill out a 20 minute LFS questionnaire, how would they feel about 30 minutes or even 40 minutes? As the Confirmit licence was still available, this requirement led to the development of a second pilot in 2012.

3.2 Second LFS web pilot - 2012

The main aim of the second LFS web pilot was to look at whether length of the questionnaire affected respondents' willingness to complete the questionnaire, and also investigate whether people would complete questionnaires for other members of the household.

The questionnaire instrument for the first pilot was replicated, and then extra sections of questions were added to increase the approximate length to 40 minutes. In terms of content it was almost as long as the interviewer-led LFS but without some of the ad-hoc modules.

30 respondents were assigned to three different questionnaire lengths – 20 minutes, 30 minutes, and 40 minutes. Those in the 20 minute group were asked the same questions from the original pilot, while for the other two groups, routing was added to the extra questions to provide an approximate 30 minute interview, and a 40 minute interview with the full set of questions.

All were familiar with ONS's social survey questionnaires having agreed to take part in future work, but none had taken part in the LFS. Interviews took place in the respondents' homes using their own computers and the respondents were required to complete the survey on their own with no guidance

from the interviewer. The interviewer observed how the respondent completed the survey and a cognitive interview was conducted once the respondent had finished.

There was a mixed reaction when respondents were asked to provide feedback regarding the length of the questionnaire they had completed. Whilst respondents were generally positive in terms of the 30 minute interview, they did at the same time suggest that 30 minutes was probably the maximum length of interview that they would be willing to complete. For the 40 minute interview the reaction was much more varied, however with many of the participants stating that they thought the questionnaire was too long.

There were also mixed feelings when respondents were asked whether they would be happy to complete the information on behalf of other household members. Those who were willing to spend time completing this information said they wanted to help where possible, and were confident that they could answer the questions in reference to family members. Those who were not willing to provide information suggested reasons of confidentiality, time spent completing the questionnaire and insufficient knowledge about other household members.

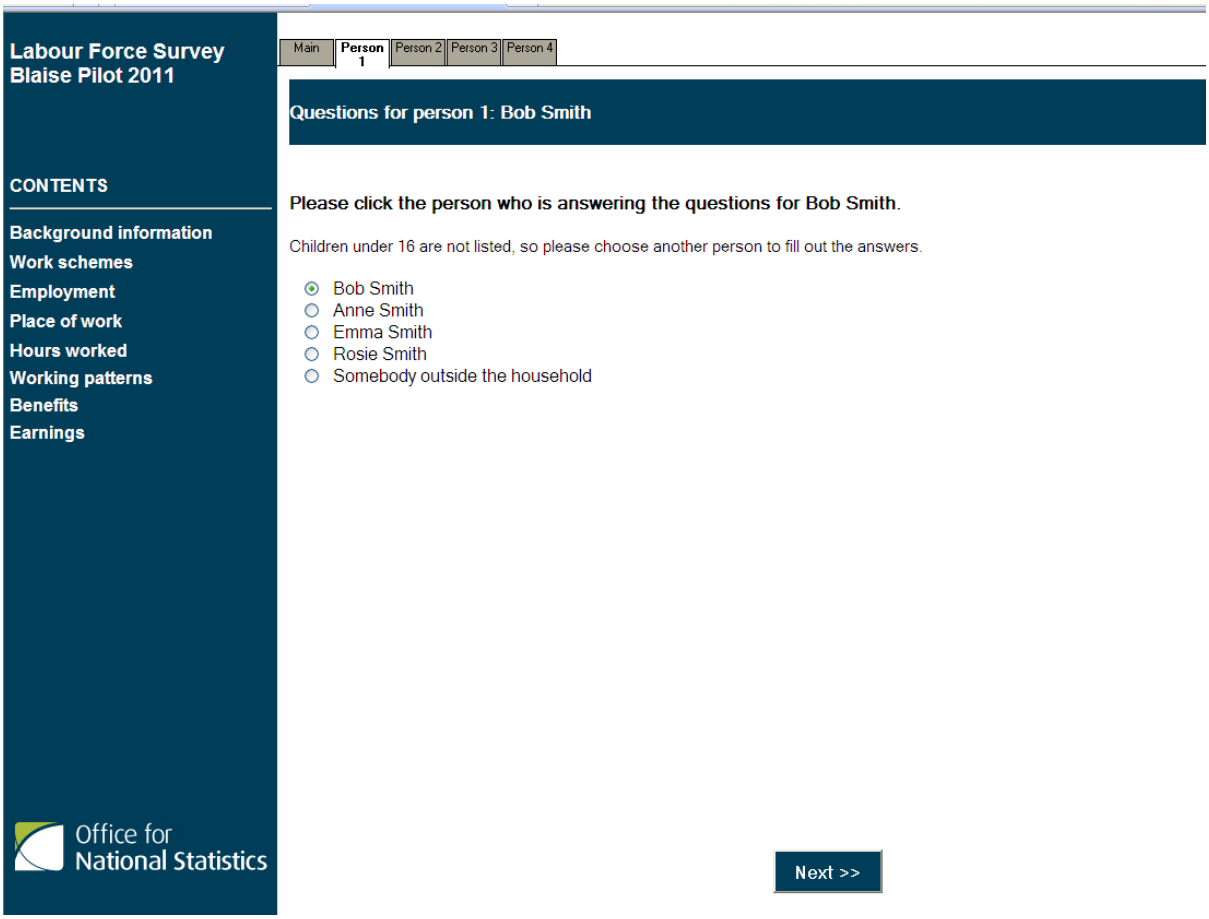
Both pilots provided many useful recommendations to inform future work, particularly in terms of questionnaire development and length. Confirmit Professional was useful in providing a fairly quick and easy way to set up and host a web survey, but the conclusion drawn was that it didn't fully meet the needs of Social Survey Division for complex household surveys and integrating with current systems. The best way forward was to spend time developing a prototype LFS household questionnaire in Blaise IS which would be ready to go live when the hardware was put in place.

4 Current work

4.1 LFS household web prototype (Blaise IS)

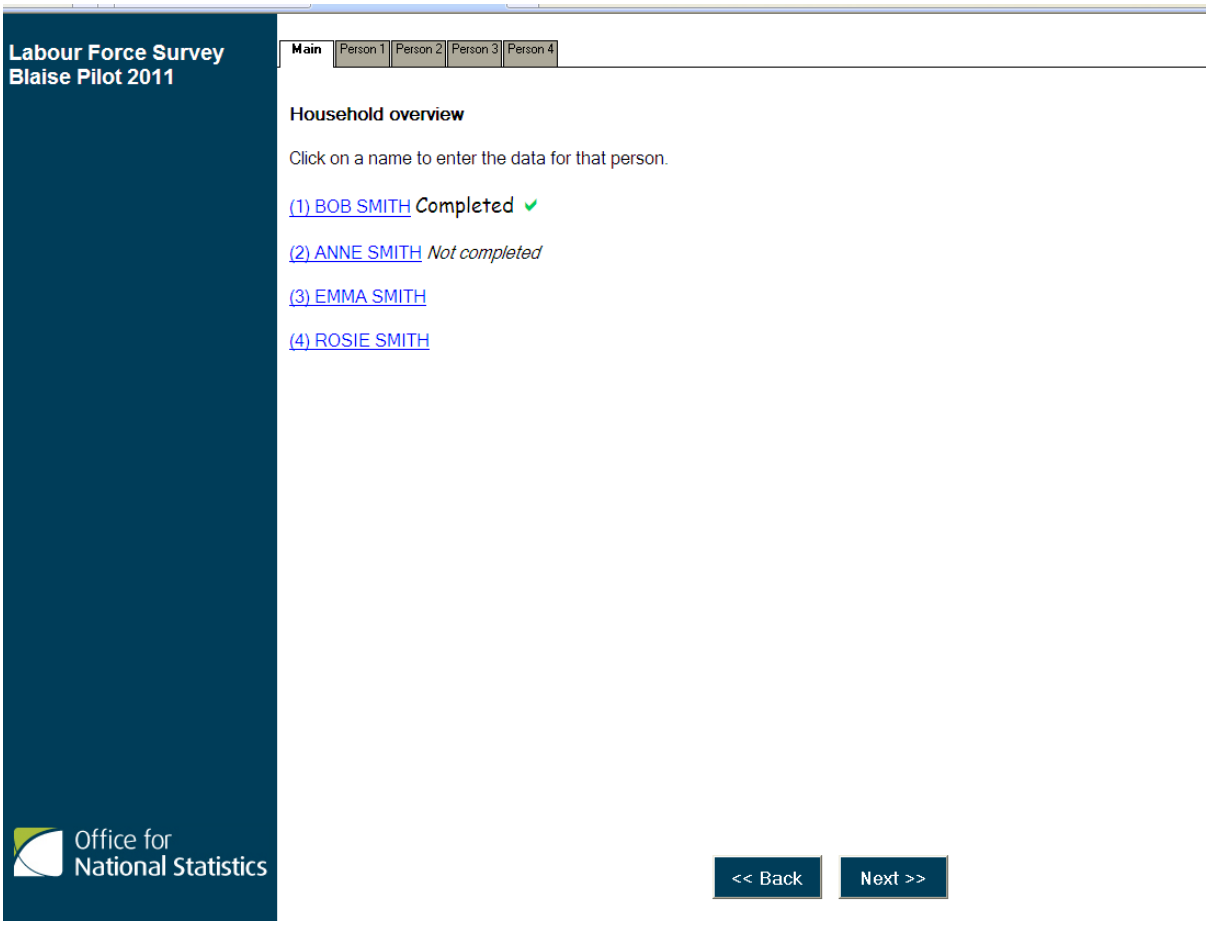
In 2011, the BDSS team started to program a web version of the LFS using Blaise IS 4.8.1.

This time we made it more similar to the interviewer-led LFS in comparison to the Confirmit pilot by programming it as a household survey. A tab will appear for each household member, which will jump to his or her personal questionnaire. Each respondent can fill in their own information, or one respondent can do this for themselves and then the others in the household. As with the face-to-face and telephone modes, it is possible to state that the questionnaire is being completed by someone else (by proxy; see screenshot 3) and appropriate textfills will then be used in the question text.



Screenshot 3: Collecting data from all members of the household

Once each person has attempted their questions (either in person or by proxy) the household selection screen (screenshot 4) displays the status of each person and allows the selection of another person, or submission of the whole questionnaire.



Screenshot 4: Household member selection screen with status

We have not yet been able to pilot this work as we do not have the hardware in place to host externally, so the questionnaire is currently hosted on the ONS intranet for internal testing. Until we obtain a hosting solution, we are still working on the questionnaire so it will be ready to go live in the future.

As the LFS is a very lengthy and complex questionnaire, there is still a lot more work to do to make it user-friendly and engaging for respondents. As there will be no interviewer to guide respondents through the questionnaire, the wording of questions, placement and wording of checks and onscreen guidance will need to be revised. Also, the LFS uses many coding frames, and so far we have only tested out the shorter, simpler ones such as the country coding list for country of birth.

As well as visual improvements, we are concerned with ethical considerations such as the situation where more than one household member is interviewed within the same questionnaire, and individual responses may be visible to other household members. Also, following a household's first interview, a lot of its data will be 'rotated' into the next wave and used in question text, response options and questionnaire checks to cut down on interview time and decrease the burden on the respondent. As well as the ethical issues this may cause, it also raises security concerns as personal details are being stored on an external server.

The earlier internet piloting work focussed on the survey stages up to and including collecting data from respondents. The data processing stage was outside the scope of the pilots. Work is taking place to identify the requirements and agree the procedures to process data from an online mode and integrate it with data from the face-to-face and telephone modes. This is further complicated by the inflexible, aging systems in place that currently collect and process LFS data.

4.2 Electronic Data Collection project

4.2.1 Background

Towards the end of 2011, a project was set up in Business Statistics Division (BSD) of ONS to look at modernising ways of collecting data from businesses. Currently, this division collects data from businesses largely from paper questionnaires, or for the simpler questionnaires, telephone data entry (TDE). The systems that have been built up to support this method are antiquated and expensive to maintain. Paper questionnaires place an unnecessary burden on respondents, and take time to dispatch, collect and process. Responding to ONS's business surveys is a legal requirement under the Statistics of Trade Act (1947), so the office is facing increasing pressure from businesses to modernise the data collection process.

The Electronic Data Collection project (EDC) was set up to address these concerns, with the specific aim of delivering web based business data collection through a web-based portal. The definition of portal for the purposes of this work is:

> "a web site that functions as a point of access to information in the World Wide Web. A portal presents information from diverse sources in a unified way…Portals provide a way for enterprises to provide a consistent look and feel with access control and procedures for multiple applications and databases, which otherwise would have been different entities altogether." (Wikipedia, 2012).

It became clear early on that the aims of BSD and the EDC project were similar to those of Social Survey Division, and so both divisions started to work together with the help of specialist contractors to ensure that the project delivered a solution that met the needs of the two areas.

The vision is to provide a common and secure web data collection platform, extendable to delivering a variety of secure data collection and messaging services to ONS's business and household respondents (and potentially for consulting users of statistics), that integrates securely with ONS's back-end systems where much of the complexity lies.

4.2.2 Liferay and Blaise: 'Proof of concept' phase

The first phase of the project started in November 2011 and was completed in March 2012. The aim was to provide a 'proof of concept' to demonstrate how web data collection might operate within ONS, while meeting the list of requirements from both work areas. This phase did not include actual data collection, though the aim is to demonstrate this in the next phase.

Various software packages were assessed which provide a web-based portal, and Liferay 6.1 (Community Edition) was selected by the contractors as the package they would use to demonstrate the proof of concept.

The Proof of Concept Report (2012) describes the features of Liferay:

> "Liferay Portal is a free and open source enterprise portal written in Java and distributed under a GNU Lesser General Public License and a number of proprietary licenses. Liferay has been design for delivery of intranets and extranets, for organisations of any size. Liferay Portal allows users to set up features common to most websites without requiring large amounts of custom development. Liferay is constructed from functional units called portlets assembled in a content management framework and web application framework. Liferay has achieved a solid track record across a number of industries, providing real world performance with for intranets and extranets across verticals. In the context of the electronic data collection project Liferay offers a range of building blocks with the capability to accelerate development and deployment to meet the objectives of the programme in both the business and social contexts."

The main purpose of the portal was to deliver these functions:

- **Registration** – Allow a user to register using a unique identifier and password. Following successful authentication, a user will be able to confirm his or her company/personal details and create a unique password and user id as well as entering required information such as email address. This email address will be used to activate a user's account.
- **Login** – Handle the day to day authentication and authorisation of registered users.
- **Portfolio management** - Allow a user to view all completed, in progress and 'not started' surveys for that company/user.
- **Web survey integration** - On selection of a questionnaire or following a link to a questionnaire, the portal will call a specific web survey application.

Although both divisions used Confirmit Professional for early pilot work, Blaise IS 4.8.2 (and later, 4.8.3) was chosen as the web survey application to integrate with the Liferay portal. The main reasons for this were functionality, as we know Blaise IS can produce the complex surveys required; financial, in that ONS already has Blaise licences; and support, with many expert Blaise users plus the available help from Statistics Netherlands. For Social Survey Division, it makes consistency across modes as we already produce Blaise datasets for telephone and face-to-face. The portal can however work with a variety of applications, so in the future, BSD could use a different web survey package, and SSD could make the transition to Blaise 5.

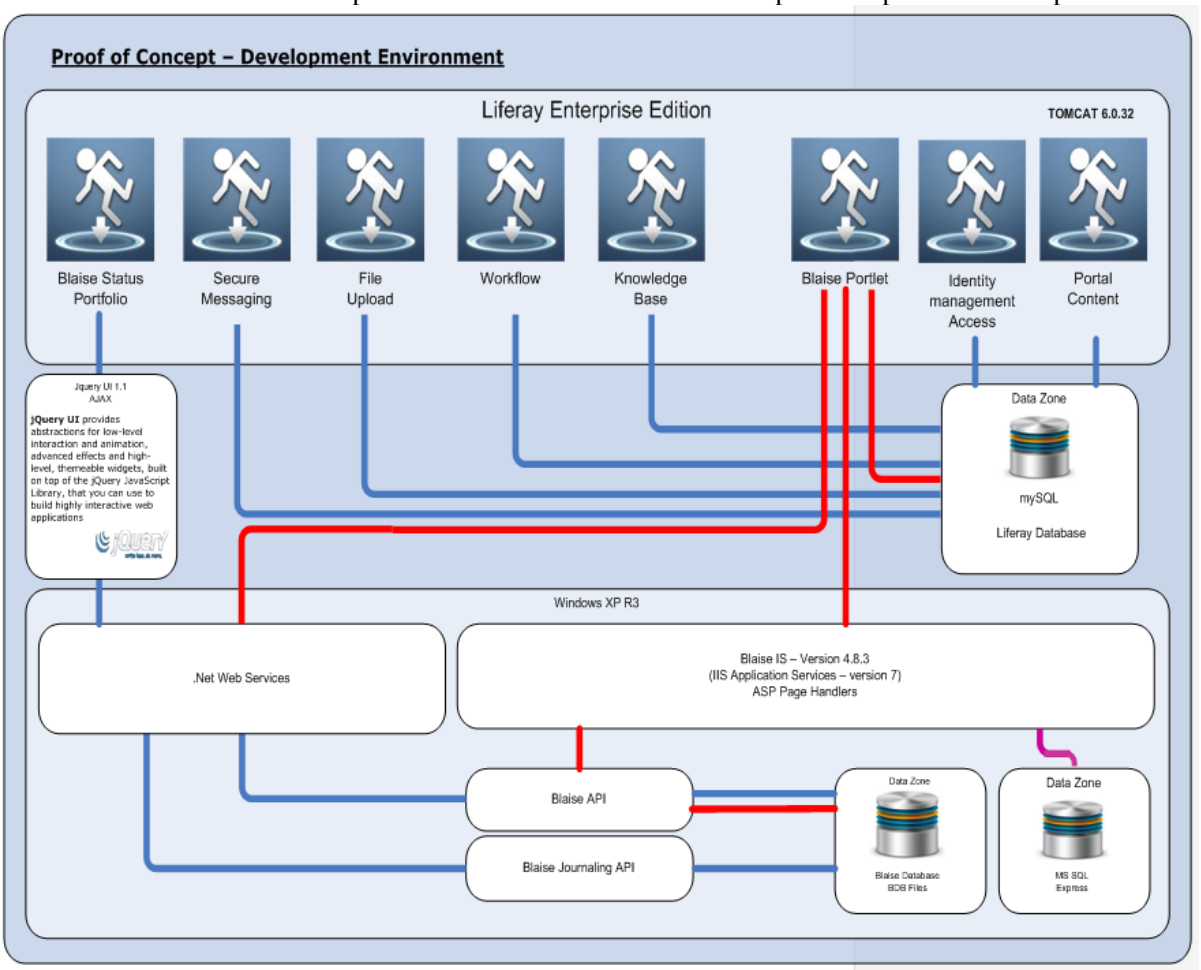
The Blaise team in SSD produced an online version of BSD's Capital Expenditure Survey (Capex), which was the first attempt to program a business survey using Blaise. The team also provided the code for the LFS prototype web questionnaire. The two surveys were then integrated into Liferay.

Some work was necessary to try and pass external values into Blaise. To test this, we amended the BiInterviewStarter.asp page using hard coded values. We input a variety of different values such as language, currency and business reference number into Blaise (see screenshot 5), and this allowed us to control who viewed certain questions in the routing, ensuring that respondents were viewing relevant questions and information. Eventually the API was used instead to dynamically pass values between Liferay and Blaise. Both methods allowed us to change the values within Blaise, so for example a respondent could select Welsh as their language of choice when they register an account on Liferay, but change it to English part way through the questionnaire. This preference would pass back to Liferay and then be remembered next time they log in.

```
Call AddField(xmlDoc, FieldsNode, "QBParameters.Lang", "ENG", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.Currency", "GBP", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QID.RefNum", "49900123456", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.Respondent", "1", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.QValidQues[1].Q1AquValid", "Yes", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.QValidQues[1].Q2AquValid", "Yes", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.QValidQues[1].Q3aAquValid", "Yes", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.QValidQues[1].Q3bAquValid", "Yes", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.QValidQues[1].Q4AquValid", "Yes", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.QValidQues[1].TotAquValid", "Yes", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.QValidQues[1].Q2DisValid", "No", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.QValidQues[1].Q3aDisValid", "No", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.QValidQues[1].Q3bDisValid", "No", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.QValidQues[1].Q4DisValid", "No", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.QValidQues[1].TotDisValid", "No", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.QValidQues[2].Q1AquValid", "No", biResponse, "")
Call AddField(xmlDoc, FieldsNode, "QBParameters.QValidQues[2].Q2AquValid", "No", biResponse, "")
```

Screenshot 5: Modification to BiInterviewStarter.asp with hard-coded values

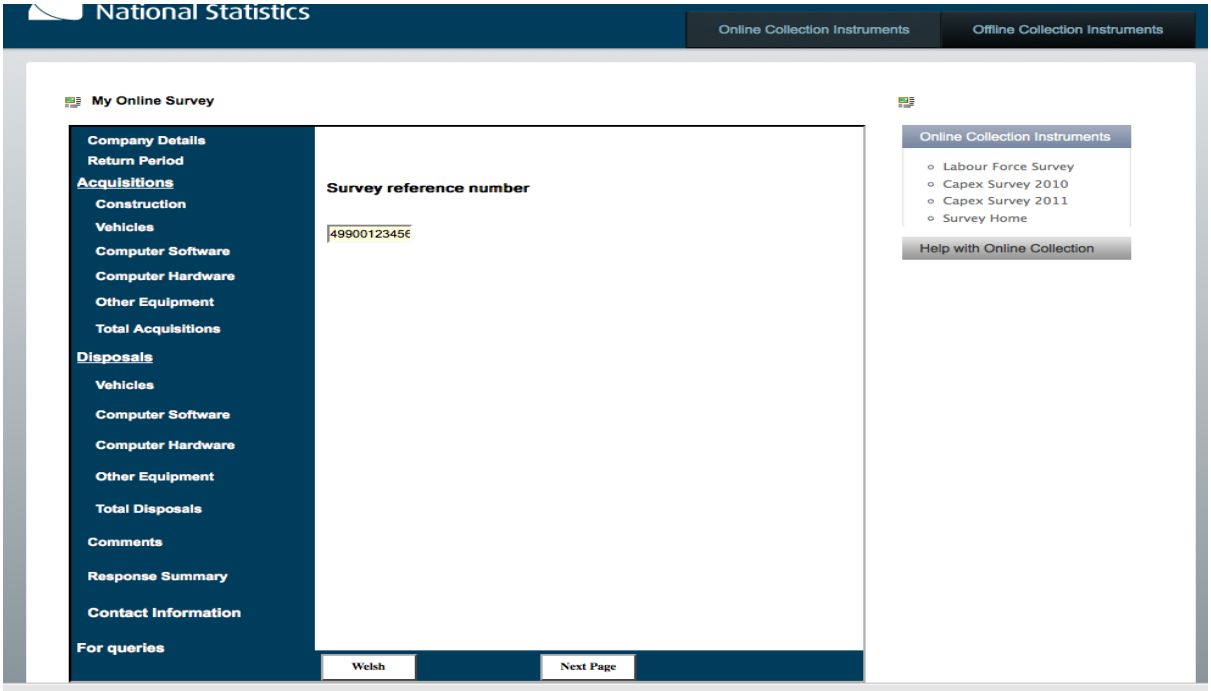
Screenshot 6 shows the development environment which was set up for the proof of concept:



Screenshot 6: Proof of concept development environment

Following the setup of the Liferay portal, a series of stakeholder demonstrations were held, showing how the proposed solution would work for both business and social surveys. Using two different scenarios for a business survey respondent and a social survey respondent, the demonstrations showed the process of registration through Liferay (where the user can create a username and password), then logging in to a survey, and completing the survey.

Screenshot 7 demonstrates the Blaise IS Capex survey running within the portal:



Screenshot 7: Blaise IS running the Capex survey through the web portal

The proof of concept exercise proved that Liferay integrates well with Blaise IS to produce a solution that will help ONS make a step forward in modernising its data collection processes. A series of performance test are being conducted (results not available at the time of writing).

Although Blaise IS appears to function well with Liferay, the portal allows different packages to be slotted in alongside or instead of Blaise if necessary. It can be updated and programmed without any coding knowledge, making it easy to maintain.

The portal has other features that will be useful to engage with respondents such as news feeds and secure messaging, plus the ability to show relevant information based on respondent id such as when they need to complete surveys and relevant news items. These features may also benefit other areas of ONS, and not just the social and business data collection areas.

5 Thoughts about using Blaise IS

Before designing the LFS prototype questionnaire, the BDSS team has little experience in using Blaise IS. The team drew upon the samples in the Blaise examples folder and training course materials which were helpful. Some aspects of the work were more challenging, in particular layouts and journaling.

5.1 Layouts

We found that the layout options were not user-friendly. The questionnaire must look professional, and enable the respondent to fill in their details accurately and swiftly, maximising response and minimising error. While it was fairly straightforward (although time-consuming) to transfer the LFS questions into the web version, the layouts were fiddly, and the results were sometimes down to trial and error and much recompiling! We struggled in particular with custom edits and grouping.

5.2 Journaling

One of the tasks for the EDC project was to set up a system of collecting paradata. Blaise IS has a journaling feature which allows paradata to be collected, such as browser type, length of time spent on each page, interview status, and so on. In the short amount of time we had to look at this feature, we managed to demonstrate that we could collect some basic information, but not as much as much as what was specified by the project. This would have involved tweaking ASP pages and style sheets, and there was no technical expertise in the team to enable us to amend these in any depth. Liferay can also collect paradata and it looks likely that a final solution would involve either just using Liferay, or using a combination of the two.

6 Future plans

Much of the future work around electronic data collection is reliant on funding, and obtaining funding is becoming increasingly difficult due to government efficiency targets, despite the obvious gains that would be made in the long term. Assuming that ONS secures funding to make further progress with electronic data collection, then there are a few areas that we plan to explore.

6.1 Develop the Blaise LFS prototype

BDSS plans to develop the Blaise LFS prototype based on the recommendations from the two LFS pilots. Ideally, this would be integrated into Liferay for a full-scale pilot.

6.2 Offline electronic data collection

ONS hopes to be able to offer an offline form of electronic data collection, for example spreadsheets or e-questionnaires, which respondents can complete offline, but are still received and sent via the web portal. BDSS have just started looking at the BASIL component of Blaise as a possible solution.

6.3 Blaise training needs

At the moment, there is only one Blaise support team in ONS, and this sits within Social Survey Division providing support to social survey researchers. If Blaise IS becomes the future tool for electronic data collection, then it will be vital to consider the training needs within both divisions. The support team will face the challenge of not only helping to set up many different surveys, but also training one division in how to use Blaise, and another (who already have Blaise skills) in how to use the IS component.

6.4 Develop standards for web surveys

ONS social surveys currently have a series of standards in place to ensure consistency across surveys. These standards cover both programming and screen standards, for both telephone and face-to-face modes. A similar set of standards (particularly relating to display of questions, answer lists, checks and guidance) will need to be devised for web surveys to ensure consistency.

7 References

Aubrey-Smith, S.A. and Thatcher, B (2012). *Labour Force Survey Online Pilot 2012: Usability/Cognitive testing report.* Office for National Statistics, Newport (unpublished internal document).

Burgis, A. and Hamblin, J. (2012). *Proof of Concept Report*: *Electronic Business Data Collection and Social Surveys Internet Data Collection.* Office for National Statistics, Newport (unpublished internal document).

Portanti, M (2011). *Results and recommendations from the 2010/11 internet pilots: Summary report.* Office for National Statistics, Newport (unpublished internal document).

Wikipedia (2012). *Web portal*. http://en.wikipedia.org/wiki/Web_portal. [Accessed 9 Mar 2012].