The ONS Integrated Household Survey: The next installment

Alessio Fiacco, Office for National Statistics, UK

1. Introduction

The UK's Office for National Statistics (ONS) is engaged in an extensive modernisation programme. This includes a statistical modernisation project that will bring about the integration of some of its continuous household surveys.

The aim of the project is to deliver better value for money while increasing the value of statistical outputs and providing a flexible means of meeting future survey needs (Bumpstead 2004 and Burrell 2006).

The process of 'integration' includes the whole of the survey process, including data collection, processing and delivery of outputs. This paper will consider some of the Blaise issues faced in the implementation of the ONS Integrated Household Survey (IHS), focusing on the survey instrument strategy adopted.

2. Current plans

Original plans for the IHS envisaged a January 2008 'go live' date for all component surveys. More recently, it was decided to have a phased implementation of the IHS in 2008, in order to better manage the risks associated with the 'big bang' approach.

From January 2008, the General Household Survey (GHS), the Expenditure and Food Survey (EFS) and the ONS Omnibus Survey will be integrated into one household survey.

The integration of the Labour Force Survey (LFS) into the IHS and the introduction of the unclustered sample design (for all non-LFS survey streams)[1] will occur later in 2008, to coincide with the delivery of the new survey case management system. This is the system required to handle the flow of survey information and manage field operations on a fully unclustered sample.

Earlier this year, ONS made a successful bid to carry out the new English Housing Survey for the Communities and Local Government department. This housing survey will come within the IHS framework from April 2008, in conjunction with the implementation of the LFS into the IHS.

[1] The LFS is already based on an unclustered sample design.

Table 1. IHS component surveys and related acronyms

IHS Survey	Formerly known as	IHS topic (acronym)
Labour Force Survey and associated boosts in England, Wales and Scotland	LFS / LLFS	Work (WRK)
Expenditure and Food Survey	EFS	Living Costs & Food (LCF)
General Household Survey	GSL	Lifestyle (GLF)
Omnibus Survey	OMN	General Opinions (OPN)
English Housing Survey	n/a[1]	Housing (HSG)

[1] This a new survey. It is a merger of two existing surveys, the Survey of English Housing and the English House Condition Survey.

The IHS questionnaire will include all the information required by the component surveys, which will therefore be discontinued in their current form. The questionnaire consists of a common ('Core') module, providing information on census-type socio-demographic variables. The Core will be administered to the whole sample, while different topic modules will be administered only to parts of the sample.

In July 2007, ONS also won the contract for the new Longitudinal Disability Survey of Great Britain. This longitudinal study, which is commissioned by the Office for Disability Issues, is likely to be brought into scope and launched as an IHS survey, when fieldwork starts in April 2009.

Table 1. Brief history of the IHS

Date	Event	Blaise strategy
2004-05	Consultation exercise (talks, presentations and seminars) Meetings with key stakeholders; publication of articles and presentation of papers to national and international conferences	n/a
December 2004	Prototype test	Single survey instrument
Spring/Summer 2005	Field trials carried out	Single survey instrument
February/March 2006	First pilot survey	Single survey instrument
April 2006	Integration of Field Force and Field managers	n/a
January 2007	Successful bid for English Housing Survey	n/a
February/March 2007	IHS systems test/'Dry-run' (in-house)	Four separate survey instruments[1]
June/July 2007	Second pilot survey of over 3,300 households	Five[2] separate survey instruments
September/October 2007	Longitudinal modules second wave (Work) and 'dry-run' (Lifestyles)	Separate survey instruments
April 2009	Longitudinal Disability Survey	Separate survey instrument

[1] The Blaise datamodel were built from both common and survey-specific blocks
[2] Additional datamodels were used (e.g. LCF Editing and Coding version; LCF expenditure Diary etc.)

3. Survey instrument strategy adopted

Previous IHS field trials and pilot work were carried out using a single Blaise instrument that included a Core set of questions and all questions for all topics. The routeing to the survey-specific sections was based on sample record information.

In more recent questionnaire piloting, a multiple instrument approach was adopted, which involved using five Blaise datamodels (one for each interview type). The same common

code (e.g. Core questions) was included in each questionnaire to avoid duplication of effort and obtain consistency of data.

The technical decisions made in relation to the IHS were influenced by the constraints imposed by legacy systems, as well as other considerations such as the need to minimise the impact of change.

There were several factors that were taken into account, as the IHS is an end-to-end process and the choice of the survey instrument strategy has an impact on other aspects such as field management, data processing, business organisation to name a few. These factors are explored in detail below.

3.1 Field management

Data that informs field management operations (Case Management System reports) currently rely on the use of separate Blaise instruments in order to better monitor response and the assignment status on the IHS component surveys.

For historical reasons, there has been a strong culture of survey independence in the ONS, which resulted in its case management system being geared to the needs of its many diverse projects (Hofman and Gray 1998).

Even with the launch of the IHS there will still be a need to provide data on topic, as well as the IHS as a whole, to meet different survey requirements.

Response rates on the Core will be monitored by the use of additional Core-only outcome codes.

3.2 Data processing

The need to reproduce current outputs was identified as one of the main concerns in terms of changing the survey instrument approach, compared to early IHS pilot work. At present, various processing systems are used by different IHS component surveys. In addition, the component streams have diverse processing and data turnaround cycles.

For example, the WRK stream is a weekly survey that produces rolling monthly outputs and has a very short period for interviewing reissues (so-called ‘hangover week’). Non-WRK modules, being monthly surveys, have generally longer field and reissues periods and - as in the case of LCF – the whole survey year could potentially be used to carry out reissues.

The WRK module is also unique compared to the other IHS streams, in that it uses SIR-based processing in the production of its derived variables. Non-WRK streams perform edits, derived variables computation in SPSS or Manipula. The use of a single Blaise datamodel for the IHS would require all survey streams to be processed in the SIR database-package at some stage.

For WRK, to continue to provide current survey outputs, all redundant variables (i.e. non-WRK-related) would need to be dropped before going into SIR. Non-WRK surveys would require more effort to reproduce current outputs, as their data are not currently processed in SIR. On both accounts, the multi-instrument approach in Blaise was

deemed preferable.

It was thought that having a single IHS Blaise database, created from the single instrument, could potentially introduce a bottleneck when processing the data. It would also cause the processing cycles for all survey streams to become interdependent.

3.3 Business organisation

In preparation for the recent IHS pilot work a Core IHS team has been established. This team is responsible for the delivery of Core outputs, such as tables and microdata files. The Core team will also be responsible for delivering and coordinating changes in the Core section of the Blaise questionnaire.

Existing survey teams will continue to exist after the launch of the IHS and will be responsible for implementing changes to topic-specific sections in Blaise and output production.

To ensure that inputs/outputs are in line with standardisation and best practice, it is envisaged that the Core team and the Blaise Development, Standards and Support team will work closely, being organisationally linked to each other.

3.4 Legacy/Structural factors

Certain blocks - used for storing sampling or geographical data – may require tweaking on different survey streams. Due to the panel element of the WRK, more fields form part of its primary key compared to the other IHS modules. For instance, as part of the WRK serial number, there is a field that identifies the wave at which the address first came into the sample. This field is not present on the other IHS modules.

This prevents identical blocks from being used where the ‘key’ for searching an external file is the serial number. Thus, two versions of the same blocks may be required to accommodate searches on different primary keys.

Although, a technical solution could have been found, the use of multiple datamodels provided a suitable workaround.

Table 3. Primary key fields on IHS survey streams

Survey stream(s)	External data file search
WRK	IF PAFData.SEARCH (QID.Quota, QID.Week, QID.W1Yr, QID.Qrtr, QID.Address)[1] THEN PAFData.READ SURVEYYEAR:=PAFData.QGEO.SURVEYYEAR (…)
GLF, OPN, LCF, HSG	IF PAFData.SEARCH (QID.Area, QID.Address, QID.Hhold) THEN PAFData.READ SURVEYYEAR:=PAFData.QGEO.SURVEYYEAR (…)

[1] Three additional fields form part of the WRK primary key. These fields are not used when matching sample record information

Another legacy issue is related to the reference period on different interview types. Historically, the survey design for the UK LFS has been based on 'fieldwork quarters'.

This means, for example, that the first week of any quarter is the first week of fieldwork for the quarter and the data collected actually refer to the last week of the previous quarter (the question text reads: "in the seven days ending Sunday {*the week before*}…").

For the Core section of the IHS, the reference period on non-WRK surveys is based on the last seven days from the date when the interview started, whereas on WRK the date is set in advance of the case being sent out to interviewers. On the topic-specific section of other IHS component surveys the reference periods differ by question.

3.5 Time series

A range of new outputs from the whole sample Core module will be produced. However, there is a need to continue to provide survey outputs that are currently produced from the separate surveys.

To preserve the integrity of key time series, it was necessary to take a different approach on some IHS survey streams, as in the case of the Education section of the questionnaire.

The opinion module, for instance, does not require information on Educational Qualifications to the same level of detail as the other component surveys (i.e. WRK and GLF). Thus, a shorter, Core education block is used on OPN, whereas a longer, survey-specific, version is used on GLF.

4. Issues and considerations

The survey instrument decision had to take into account emergency situations, such as discovering a routeing error (in any of the interview type) that requires the interviewers to download an amended version of the Blaise questionnaire.

In the design of the Blaise instrument, importance was placed on the need to minimise the impact of rescattering on all survey streams in view of potential overload on the current data transfer system.

4.1 Handling of errors

The IHS component surveys are generally fairly stable and changes are mainly introduced to incorporate new requirements at the beginning of the survey year or reflect changes in legislation, such as the introduction of new State benefits in the new financial year. The Core of the IHS is also fairly stable and the chances that the Core questions are incorrect and need rescattering are deemed very low.

The IHS includes the ONS Omnibus, a multipurpose survey that has a fast turnaround and a monthly changing set of questions/topics. Prior to data collection, the Blaise instrument is tested, however, it is a fair assumption that it is not as stable compared to the other continuous survey part of the IHS.

Using a single datamodel means that, if a routeing error is found in one of the survey stream, a new (amended) instrument would have to be rescattered. Thus, the impact of a routeing error, say on OPN, would affect all other survey streams.

The impact will have implications not only in terms of rescattering but also from a data

processing point of view, as extra Blaise databases would have to be merged into a single datastore, hence requiring additional computing effort.

One of the benefits of the multi-datamodels approach lies in the fact that, if errors are found in the module-specific section, only the stream affected would have to be downloaded again by interviewers and the impact on other IHS modules is thus minimised.

4.2 Context

Interviewers at present transmit their work to headquarters via internal laptop modems through a standard telephone line. The WRK is a quarterly survey whereas the other IHS survey streams are monthly surveys. Non-WRK surveys require a new questionnaire to be downloaded by interviewers every month.

Having five separate datamodels that needed to be scattered at the same time, caused some bottlenecks due to the long download times at the recent pilot (June-July 2007). The size of the questionnaires compounded the transmission problems.

Together with the large questionnaires, ONS transmits to interviewers a large WinHelp-based Help file and a large number of look-ups/coding frames. Files are sent to interviewers compressed through WINZIP.

The table below gives an indication of file sizes in kilobytes (KB) at the June/July 2007 pilot:

- IHS Lookups: 1768 KB
- IHS Help files: 1152 KB
- GLF*: 1614 KB
- LCF*: 1417 KB
- WRK*: 1323 KB
- HSG*: 631 KB
- OPN*: 519 KB

* This included Blaise metadata (*.bmi, *.bxi and *.bdm), survey-specific lookups and sample record.

Having multiple datamodels (built from a combination of common and survey-specific blocks) may put a heavier load on the transfers system if scattered at the same time. However, it also enables a staggered transmission of the questionnaires to interviewers, as a means for reducing download times for interviewers.

ONS is working on a project, which aims to deploy the use of Broadband or possibly '3G' technology[2] for field interviewers. Broadband has already been implemented for Field and Regional Managers to allow faster connection to Head Office, which will be required for work allocation processes.

The Northern Ireland Research and Statistics Agency (NISRA) shares ONS's concerns regarding file sizes. NISRA is currently responsible for carrying out the Northern Irish component of three of the five IHS component surveys. Using separate Blaise

[2] The third generation (3G) mobile phone technology allows high-speed data transmission services in a mobile environment.

instruments was also seen as preferable for our colleagues at NISRA.

4.3 Survey-specific requirements

Some IHS component surveys require additional Blaise instruments. For instance, the LCF has a version of the datamodel used for editing and coding purposes, as well as an expenditure Diary datamodel. The GLF also has a separate instrument for its keeping-in-touch exercise (Setchfield 2007).

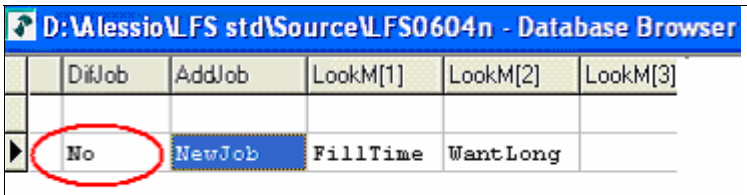
A single IHS mega-instrument could have been implemented to cater for all these purposes. However, such an approach would be in conflict with one of the requirements in terms of Blaise programming at ONS: to keep things simple wherever possible.

The issue of the size of the questionnaires (mentioned above) also caused a rethink on the data rotation method to be used on the WRK stream. The process of data rotation is another survey-specific requirement, which may have an impact on the data transfer system, as well as add to the complexity of the Blaise code.

For WRK, it was originally envisaged to use a Blaise database as an external file for moving the data from one wave of interviewing to the next one. This solution suited the multiple-instrument approach for the IHS.

However, it also aggravated the issue of file sizes for the WRK questionnaire. In addition, rotation issues - caused by the appropriate rules not being re-executed - were experienced, which resulted in the data being ‘fed-forward’ incorrectly.

Table 4. Data rotation issues when using an external Blaise database



D:\Alessio\LFS std\Source\LFS0604n - Database Browser

	DifJob	AddJob	LookM[1]	LookM[2]	LookM[3]
▶	No	NewJob	FillTime	WantLong	

{The field *AddJob* should only be on route if *DifJob=Yes*, but due to the rotation method used this did not happen}

```
RULES {Block level}
DifJob
IF (DifJob = Yes) THEN
    Addjob
    IF Addjob = NewJob THEN
        LookM
    ENDIF
ENDIF


RULES {Table level}
{The following code is within the household array and is placed after the Block QLookWrk has been called.
QHolding is the block within the Datamodel that receives data from the external file}
IF QLookWrk[i].DifJob = EMPTY THEN
    QLookWrk[i].DifJob:=QHolding.QHHHold[i].xDifJob
ENDIF
IF QLookWrk[i].AddJob = EMPTY THEN
    QLookWrk[i].AddJob:=QHolding.QHHHold[i].xAddJob
ENDIF
{etc.}
```

To resolve the difficulty with the size of the questionnaire and the incorrect data rotation, a different strategy was proposed: pre-loading the fields in Blaise using a Manipula script that will be run in the WRK production line.

This method will also allow better testing of the questionnaire, reduce the amount of complex Blaise code and minimise the impact on the data transfer system.

Table 4. Pros and Cons: Single Vs Multiple Blaise instruments

Issue	Strategy	
	Single IHS instrument (February/March 2006)	Multiple IHS instruments (June/July 2007)
Impact of errors discovered after scattering:		
- Error in the common Core section[2]	Rescatter once (every stream affected)	Rescatter five times[1] (every stream affected)
- Error in the survey-specific question blocks[3]	Rescatter once (every stream affected)	Rescatter once (one stream affected)
Scattering/transfer system:	Staggered file transfer is not feasible	Staggered file transfer is feasible
Download times:	2.5 hours	30 minutes per instrument
Data processing of survey stream data:	Processing cycles become interdependent	Processing cycles need not change

[1] Five datamodels were used in the second IHS Pilot (June-July 2007)
[2] The Core section consists of questions administered to the whole sample
[3] This section is administered to parts of the sample

5. Options considered

The use of Prepare Directives to reduce the size of Blaise instruments and streamline a large number of Blaise instruments into one has been used in other organisations (e.g. Schou 2004). This option was considered for the IHS, however it was not deemed to meet the requirements of the survey.

One of the reasons for not adopting the Prepare Directive was that switching to a different directive would not create a separate metadata file (*.BMI) and data file (*.BDB): these would simply continue to be named according to the *.BLA file name.

This is a major drawback in terms of some of the survey-specific requirements mentioned above. For instance, the LCF instrument used by Coders and Editors has a different data shape (in terms of attributes etc.) from that used by interviewers.

Another reason why the use of Prepare Directives was discounted was that there may need to delete a Blaise database if the questionnaire was previously compiled using a different conditional define. This seemed to add a potential element of confusion.

6. Conclusions

Seen together with the two preceding IBUC papers on IHS (Bumpstead 2004 and Burrell 2006), this third paper in the series shows that the IHS is still evolving and how ONS has begun to adapt some of its earlier theoretical approach in the light of constraints posed by business processes.

There are various considerations when deciding which survey instrument design to follow. In the case of the IHS, end-to-end survey processing influenced some of the technical decisions on which Blaise strategy to adopt. Reducing the impact of change, in view to safeguard the ability to reproduce current outputs, for instance, had attached a high priority at ONS.

The phased-implementation of the IHS in 2008 should bring benefits in terms of the lower risk that this approach brings. The IHS will consist of five different questionnaires that are designed on a modular basis, being built around a common Core questionnaire. New survey questionnaires can be easily added on to the Core questionnaire and brought into scope of the IHS.

The introduction of the HSG module, which was not part of the first pilot (in 2006), is an example of the flexibility of this system and shows the feasibility of adding modular surveys to the IHS. As the IHS becomes established and new surveys are brought into scope, the IHS importance as a key household survey for UK official statistics is also being confirmed.

The IHS data collection design does not result in duplication of code maintenance as far as Core questions are concerned. In the longer term, the IHS should also bring about greater harmonisation on some topic-specific questions across different interview types (e.g. income questions on LCF and GLF).

Although, five different Blaise questionnaires were used in the recent pilot work, they all shared the same Mode Library to ensure consistency of screen layout. In addition, the use of the Question-by-Question Help files on IHS ensures that same working practices extend to the interviewing process on all the IHS component surveys (Setchfield 2006).

The chosen design for the IHS means that version control to ensure consistency of changes across the IHS-based surveys is of paramount importance. Version control needs to be managed using configuration control methods, such as using the latest version number as part of file names. The use of the network path for Core/common include files in Blaise is a possible way, although this was not tested in the June-July 2007 pilot. Having a dedicated Core Blaise team should ensure that version control and changes are managed in a coherent way.

Technical solutions to some of the Blaise issues linked to the use of a single Blaise instrument exist. This paper has merely scratched the surface in terms of the complexities of dealing with such a mega-instrument.

There is no doubt that some fine-tuning is required before IHS goes live in 2008. Work to address some of the issues arising from the recent pilot work is ongoing. By the time the next IBUC meets the IHS will be live and running. At that point we should have a clear overview of the whole development and, perhaps, one more but final paper on the subject.

7. References

Bumpstead R. (2004) Integration of UK government household surveys: Development of a modular Blaise instrument for a proposed Continuous Population Survey. International Blaise User Conference 2004

Burrell, T. (2006) Developing an Integrated Household Survey (IHS) Blaise Questionnaire from Four Major Social Surveys. International Blaise User Conference 2006

Hofman, L., and Gray, J. (1998) CAPI Survey Management Systems: Case Management on Laptops. In: Couper et al. (eds.), Computer Assisted Survey Information Collection.

Schou, R. (2004) Moving 43 Generated Unique Instruments to One National Instrument. International Blaise User Conference 2004

Setchfield, C. (2006) Training and Learning Opportunities using Blaise. International Blaise User Conference 2006

Setchfield, C. (2007) Coping with people who just won't stay put: The Use of Blaise in Longitudinal Panel Surveys. To be presented at the 11th International Blaise Users Conference, Annapolis, September 2007