Integration of UK government household surveys: Development of a modular Blaise instrument for a proposed Continuous Population Survey

Robert Bumpstead, Office for National Statistics, UK

1. Introduction

The Continuous Population Survey (CPS) is an innovative project to integrate the five key government surveys on which the UK Office for National Statistics leads into a single modular survey system. The proposed new survey involves a comprehensive integration of the entire survey process: from the creation of a unified field force of interviewers administering a common modular questionnaire, to the processing and production of outputs from a single common source.

The proposal to integrate Office for National Statistics (ONS) household surveys builds on a successful and ongoing programme of harmonisation and instrument standardisation which began in the 1990s. This paper first sets the context for the CPS by exploring the drivers to further integration, and then the costs, benefits and challenges entailed in an integrated household survey.

The paper also describes the main features of the proposed survey and discusses the implications of these for building an integrated questionnaire and modular Blaise instrument. Development of the instrument itself is still at a relatively early stage, but issues already scoped by the project and examined here include: harmonisation and standardisation; modular structure; routing; dependent interviewing; concurrent interviewing; interview mode, and; use of audit trails to inform design.

This paper also considers aspects of the wider programme to modernise ONS statistical tools and systems of which the CPS is a fundamental part and is, in turn, dependent upon. The paper concludes with an overview of the next steps in the survey's development.

2. Context

2.1. Background

The widespread development and application of Computer Assisted Interviewing (CAI) over the last two decades is one factor which has prompted consideration of comprehensive survey integration. The existence of electronic questionnaires and the computerisation of other elements of the data collection process has removed some of the practical difficulties in designing and administering combined surveys.

Among National Statistical Institutes (NSIs), perhaps the most notable development was the creation of an integrated household survey, *POLS*, by Statistics Netherlands in the late 1990s. Proposals for integration were previously investigated in the UK around the same time. One result was the successful launch of a merged Expenditure and Food Survey (EFS), bringing improvements in quality, coherence, and efficiency by replacing two separate surveys with one integrated survey.

However, for a variety of reasons, no further major integration took place. Since then, a stronger base for integration of government household surveys has developed, together with the

availability of improved technologies with which to deliver it. In addition, the drivers for integration have significantly increased.

2.2. Drivers

At the heart of a proposed Continuous Population Survey is a recognition of the need for UK National Statistics to produce better information on key social and economic variables between decennial censuses, for a range of policy purposes, and to meet the increasing demand for regional and sub-regional information.

In common with many other NSIs, ONS has experienced a rapid growth in the demand for small area statistics in recent years. In the UK there has been particular pressure for information about income and ethnicity, but also a range of other outputs. Cross-government initiatives to improve public access to official statistics, and to improve the availability and precision of local and regional statistics, has added to this demand.

There is also increasing demand for a range of new surveys, including those under European Union (EU) regulations which Member States are obliged to fulfil. Furthermore, there is a need to maximise value from ONS continuous surveys and improve the coherence of National Statistics. These demands cannot be met within the current survey arrangements.

2.3. Existing survey arrangements

Summary descriptions of the five surveys to be integrated are provided in Annex A. They are:

- the Labour Force Survey (LFS)
- the Annual Population Survey (APS);
- the General Household Survey (GHS);
- the Expenditure and Food Survey (EFS), and;
- the National Statistics Omnibus Survey (OMN).

An ongoing programme of harmonisation of questions, concepts, and classifications has improved the coherence of survey outputs in recent years. Household surveys have increasingly adopted common standards across not only the ONS but also the wider UK Government Statistical Service (GSS). Standardisation of Blaise instruments among the ONS household surveys has established a range of common authoring practices from naming conventions to screen layouts. Harmonisation of questions and classifications has increased the overlap between survey content. There is now more consistency between the surveys, but scope for improvement remains.

For example, estimates of the same variables across the different surveys cannot be combined and, despite the use of common questions, small but statistically significant differences between those estimates sometimes occur. Different fieldforces of interviewers serve different surveys, with differences between training programmes, work patterns and practices. The existence and maintenance of separate fieldforces, together with separate Blaise instruments, and a range of different survey processes from sample selection to weighting and imputation, imply a duplication of effort and sub-optimal use of limited resource.

2.4. Objectives

A Continuous Population Survey aims to build on and improve standardisation. As a coherent, effective survey system the CPS will include a range of components which will, together, best describe society as a whole and the individuals within it. Moreover, ONS believes that only through the combined resource of an integrated survey can demands for improved reporting at national, regional and local levels be addressed.

The central objectives of the CPS are to:

- develop a world class modular survey system and accompanying modular Blaise instrument, better able to meet the information needs of the 21st century;
- provide more coherent, better quality information on which Government, stakeholders and the wider user community can base their decisions;
- increase the precision of statistical outputs at national, regional and sub-regional levels;
- create a range of new outputs, including inter-censal estimates of key socio-demographic variables at the sub-regional level;
- develop a survey system with the flexibility to accommodate other surveys at a later stage;

And, to achieve these objectives:

- while maintaining continuity of outputs and preserving the integrity of key time series;
- without the need for resource additional to the combined budgets of the component surveys, and;
- while delivering further efficiency savings from economies of scale and increased value of statistical outputs.

2.5. Current position

The development of an integrated household survey to replace the existing ONS-led continuous household surveys is at a relatively early stage. Following an initial broad scoping study and informal consultations with stakeholders, funding was agreed in 2003 for a development project to establish more detailed proposals. At this time the project was known as the Integrated Social Survey (ISS). Later, it was renamed the Continuous Population Survey to better reflect its purpose and outputs.

ONS is currently engaged in a round of formal user consultation which started with publication of consultation document at the end of July 2004. [1] Work on the overall modular design of the survey is ongoing, while an extensive programme of development and testing has been proposed. On satisfaction of a range of agreed quality measures, the CPS would proceed to a start date no earlier than January 2008.

[1] http://www.statistics.gov.uk/about/consultations/ons_consultations/downloads/CPS_Consultation.pdf

3. A Continuous Population Survey

3.1. Principal features

The proposed form of integration involves a single sample of addresses and a modular survey instrument comprising all the existing topics to meet the information needs currently met by the separate surveys.

A short core module will be administered to the whole sample, while different topic modules will be administered to selected parts of the sample. Interview combinations will be composed of both core and selected topic modules. Interviewers in an integrated field force will each be responsible for delivering all the required interview combinations in their area.

From these main survey features stem a series of potential benefits and risks. These make the integrated survey highly attractive to data users in terms of increased value of statistical outputs, but also present a series of challenges in terms of managing change and minimising its unwanted effects. These elements are central in informing the objectives and implementation of a modular questionnaire and Blaise instrument for the CPS.

3.2. Benefits

There are large potential gains in the precision of many estimates from an integrated survey without the need for additional cost. These result from a series of factors, including improved weighting and better local representation. Two principal factors are further discussed here - sample size and design. In addition, improvements in flexibility and coherence are also considered.

3.2.1. Sample size

The CPS sample will be, in principle, composed of the cumulative total of addresses sampled by the existing surveys. Based on the survey reference year 2003/04, the CPS would have an annual independent achieved sample of 270 thousand households and more than 500 thousand adults, making it the largest ever continuous survey to be conducted in Great Britain. A full breakdown by survey of the number of achieved interviews in the survey reference year 2003/04 is given in Annex A.

All outputs derived from the core module will benefit from gains in precision associated with the new, larger sample. In comparison with the current LFS family of outputs these gains will be modest. In comparison with outputs derived from the other surveys, they will be very large indeed.

3.2.2. Sample design

The CPS will adopt an unclustered design, similar to that used by the LFS in Great Britain. Currently, all other Government household surveys use clustered designs, where addresses in a sample are selected from particular areas grouped in small ‘clusters’. Traditionally, this was necessary for all but the very largest surveys for reasons of economy. Clustering reduces interviewer travel time and costs, and makes fieldwork practical and affordable. However, a clustered sample leads to less reliable estimates.

By combining all the survey samples and fieldwork into one overall design the CPS will, for the first time, deliver unclustered samples for all interview combinations. An unclustered design will lead to substantial precision gains for many of the outputs currently produced from the GHS, EFS and Omnibus surveys.

3.2.3. Coherence and flexibility

A single modular survey approach also aims to improve coherence by delivering a range of outputs from a single data source. It reduces the incidence of 'competing estimates' where different surveys sometimes produce different estimates for the same variable. The process of harmonisation would be furthered as core module questions will be, by definition, common across the survey.

A flexible modular design would enable the requirements for new surveys to be met more readily and cost-effectively than at present. The expense and development time associated with launching a new stand-alone national survey are considerable, adding a module to an integrated survey much less so. The CPS also provides the flexibility to plan topic modules on the basis of sample size for the level of precision required, rather than constrained by whichever existing survey vehicle happened to be available, as at present.

As well as developing an appropriate design for the planned 2008 launch, the new survey is being built with future change in mind. The expectation is that new modules will be added as policy needs change over time, and that requirements for some new surveys will be met through the CPS. In addition, sponsors of other surveys may opt to switch their existing continuous surveys to this vehicle. Each survey joining the CPS would contribute to the overall sample size and further boost the precision and analytical power of the survey data.

3.3. Risks

There are a number of risks to the development and successful implementation of a Continuous Population Survey. Principal areas referred to relate to aspects of quality and continuity.

3.3.1. Quality

It is essential to a cost-effective CPS that each interviewer is responsible for all the interviews in their area. All interviewers, including new interviewers, will need to cover a large range of topics; larger than many will have covered before. Rather than work on a single survey, or a different survey each month, interviewers will be required to cover all interview topics on a weekly basis. New e-based learning methodologies and techniques must be devised as part of a programme to develop an integrated field force, and be validated in field trials.

Equally, the proposed new fieldwork design and modular questionnaire will need to be tested in practice to ensure they are acceptable to respondents and interviewers, help optimise response rates, and minimise respondent burden. Field trials must demonstrate that outputs can be delivered within the timescales necessary and to the required quality.

3.3.2. Continuity

Ultimately, any decision to implement the CPS rests on a demonstrable ability to deliver the benefits anticipated while maintaining the integrity of key time series, for example the UK

headline rate of unemployment. [2] In some instances time-series data span decades. An unduly large or unexplained step change in estimates would be unacceptable. A large scale experiment will be conducted to run a model CPS alongside the existing surveys. Benchmarked outputs from the two sources will be published and compared, and subsequently used to inform the decision regarding the go-ahead of the CPS, further development, or discontinuation.

3.4 Instrument requirements

Good questionnaire and Blaise instrument design will be central to the success of the project. To realise the statistical benefits implied by the new sample design, the questionnaire must be modular in structure, flexible, and consist of all the topics necessary to maintain fulfilment of information needs.

At the same time, the questionnaire must minimise and simplify the requirements placed on the interviewer (and respondent) in as far as is possible. Given the increase in the diversity of interviewer workload, improvements in the usability of Blaise instruments are highly desirable. Equally, a reduction in the time necessary to conduct each interview is also a priority, together with any other innovations which might help maximise response rates.

These improvements need to be delivered while maintaining sufficient comparability between the old and new designs so that unacceptable discontinuities to key outputs do not arise.

Finally, an essential element of CPS instrument design involves 'future-proofing' the survey. The modular structure of the survey instrument must be designed so as to readily accommodate new topics, and ensure that a range of survey types and features can be incorporated within the CPS survey system. These include:

- interview combinations with cross-sectional, quarterly or annual panel designs;
- diary components and other self-completion elements;
- telephone or personal interviewing, and;
- provision for interviewing with all household members, or with individuals sub-sampled within a household, or combinations of both.

Meeting these various requirements within a single instrument are further discussed in the following section 4.

[2] The headline unemployment rate as defined by the International Labour Organisation is a key output of the quarterly Labour Force Survey.

4. A modular questionnaire and instrument design

The CPS questionnaire will be designed as a single modular survey instrument comprising:

- a **core module** administered to the whole sample providing information on key variables for all CPS households and persons;
- **topic modules** administered to parts of the sample providing information on variables for which sufficient precision to meet policy needs can be obtained from a portion of the CPS sample;
- a small number of viable **interview combinations** formed from combining core with selected topic modules so that all topic modules are covered.

The CPS will involve design of the new core module and a re-examination of the combinations of topic modules represented by the current surveys, with the aim of improving coherence in reporting and optimising the overall use of modules.

4.1. Core module

Core module questions will collect information on census-type variables and other key socio-demographic and labour market variables. For example, household composition, tenure, employment status and economic activity, educational attainment and income, ethnicity and national identity.

The core questionnaire needs to be relatively short and straightforward so that total interview length for core and topic modules remain viable. Indeed, it is essential to the success of the CPS that the core questionnaire does not become unduly burdensome.

There are advantages to fixing key core questions over a number of years to provide an uninterrupted time series to a high level of precision. However, continuous reporting is not essential for all outputs. To allow greater flexibility and scope for inclusion of a wider range of questions, it would be desirable to include some questions periodically. Therefore, it is proposed to sub-divide the core module into two parts:

- A **fixed core** module with questions normally included for at least a five year period (or longer);
- A small **rotating core** module with questions normally included once every three years. For example, where a question might be asked first in 2008, then in 2011, and so on.

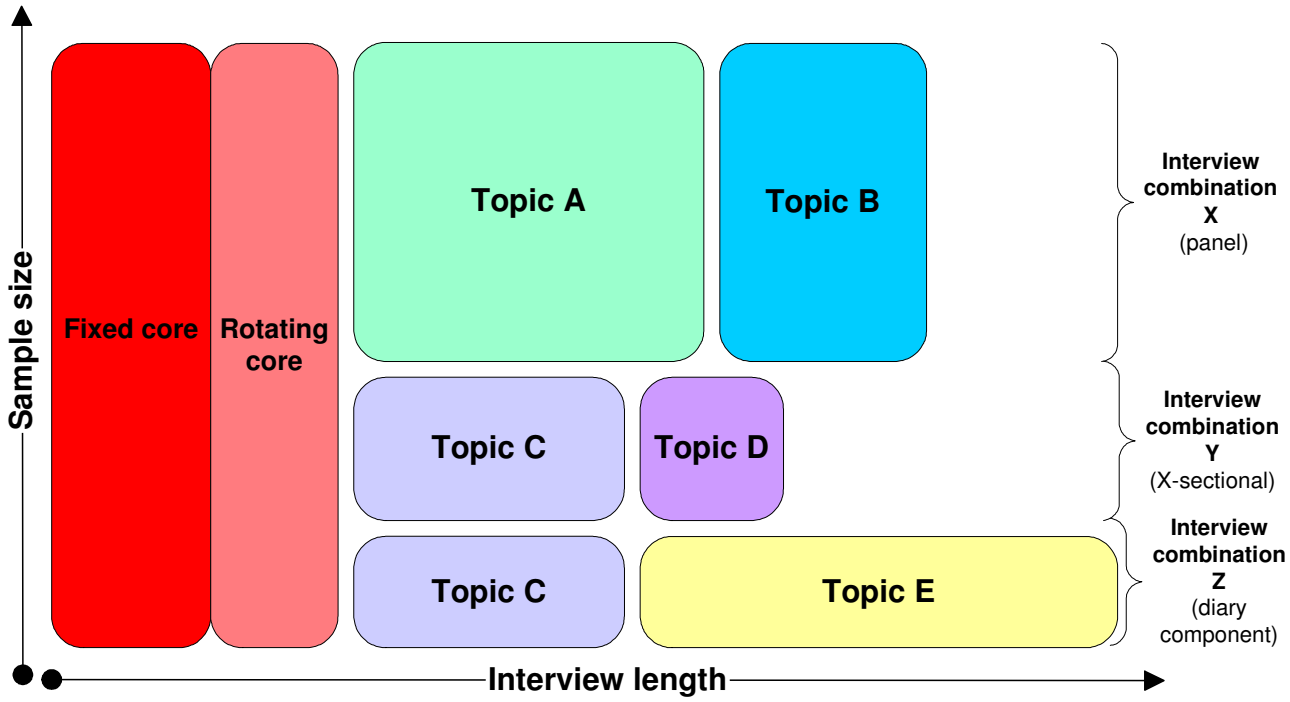
4.2. Topic modules

Questions in topic modules will collect detailed information on the various subjects covered by existing surveys where a continuing need for this information is identified. For example, health, education, labour market, income, and expenditure.

It is possible to run some modules across more than one interview combination to generate a sufficient sample size. It is also be possible to use the core questionnaire to identify rare or hard-to-find groups so that additional questions could be administered, or respondents followed-up at a later stage. A modular design would enable the quicker and more effective introduction of new modules, and amendment of existing modules, as customer output requirements change.

Figure 1 below shows a simplified abstract representation of how core and topic modules can be combined to form interview combinations.

Figure 1: Abstract representation of a modular survey instrument



A detailed illustrative diagram of a modular CPS questionnaire is contained in Annex B.

4.3. Interview mode

The CPS will be a mixed mode data collection instrument, combining elements of Computer Assisted Personal, Telephone, and Self Interview (CAPI, CATI, CASI).

Where an interviewer records an initial non-response, households may be 're-issued' to an alternative interview mode for a further attempt to secure an interview. For example, a field interview (CAPI) may be re-issued to the telephone unit where a telephone number can be obtained, or a telephone interview (CATI) re-issued to a fieldforce interviewer.

An integrated fieldforce, combined with new data handling systems, will improve the flexibility and speed with which re-issuing of interviews between modes can occur. [3] An efficient and systematic approach to mixed mode interviewing will help maximise response.

Different topic modules will have different rules for the use of personal, telephone, or self-interview - as with the current surveys. However, it is proposed that all questions in the core

[3] A proposed new Survey Case Management System (SCMS) is further discussed in section 5.

module should be suitable for both CAPI and CATI. Therefore, at a minimum, in the event of interview non-response in the field, the core module could be re-issued to the telephone unit.

This element of the survey demands an improved instrument design more sensitive to the particular mode of interview. Whereas current survey instruments are designed with a primary mode in mind, the CPS requires a mode-sensitive instrument that routes question type according to interview mode.

For example, in CAPI visual 'show cards' can be used to help respondents select an answer from a long list of options. In CATI the same information is best collected via a series of branched questions, each offering fewer answer options. Potential benefits include reductions in item non-response and respondent and interviewer burden, and, improvements in data quality.

4.4. Harmonisation and standardisation

All ONS computer assisted household survey instruments are written using Blaise, based on blocks sharing common standards. These are comprised of harmonised modules of questions (e.g. employment, ethnicity, education) and standard code (e.g. a block for the interviewer to record details about the administration of the interview, such as outcome and number of calls). Survey researchers also play the role of instrument developers, building their questionnaires using these standard blocks, together with blocks of survey specific questions.

These standard blocks are revised on an annual basis and are ready for inclusion on questionnaires at the beginning of the survey year, although essential minor changes may be made throughout the year. Instruments also conform to other common standards (e.g. screen design and layout) and are authored to common programming standards (e.g. naming conventions, datamodel structure). The ONS Blaise Development and Support Team (BDSS) monitors application of these principles.

This structure provides an excellent basis for full survey integration which may be seen, itself, as a logical extension of the harmonisation project. Indeed, without a preceding programme of harmonisation and standardisation, any move to any integrated survey would involve a step change of such magnitude that large discontinuities in outputs would be inevitable.

An important aspect of CPS development will be to take this process a stage further. Differences between survey specific questions can be minimised and, where duplication exists across surveys, blocks and questions will be removed. All blocks within the survey will, in essence, become standard. So for example, where it is necessary to collect detailed income information for different purposes in two interview combinations, the two income blocks will be developed to maximise similarity. Reducing the number of different questions (and rules of application) across the CPS will also reduce the interviewer learning burden.

4.5. Datamodel structure and routing

The CPS instrument is being developed as single datamodel, with elements of the existing survey instruments grouped together under an additional level of hierarchy. This approach minimises disruption to established survey specific naming conventions, but amalgamates those elements of blocks where a single survey approach is sufficient (e.g. top-level blocks for defining answer types, auxiliary variables, local variables, coding libraries etc.).

The routing structure currently being developed involves a linear progression through the instrument, in the sense that the entirety of the core module precedes sequential topic specific modules for an individual interview. This approach eliminates differential order effects in the core module, as core questions are always asked in the same sequence regardless of the interview combination which follows. It also provides a good working basis for topic specific researchers to develop blocks of code which can be inserted into the overall modular structure, overseen by a central design authority with responsibility for the instrument as a whole.

4.6. Audit trails

In recent years ONS has made greater use of the Blaise Audit Trail facility. Using the ATLAS and WesAudit tools, provided by Statistics Canada and Westat respectively, developers have been better able to analyse the vast array of data that an audit trail produces.[4]

The CPS will collect audit data in field trials planned for 2005 and 2006 to produce accurate timing information for individual questions, question modules and interview combinations. This will help inform design and optimise interview length across the different interview combinations. However, audit data from the existing surveys already provides CPS developers with useful information. By adding block and question times together from the current surveys to resemble the intended interview combinations in the new survey, good assessments of likely interview length can be made before the CPS instrument even enters the field. This helps to improve and accelerate the design process itself.

Audit trail data will also be used to validate the success of the CPS in reducing interview length compared with the current surveys. It could also be used to assess other aspects of the new survey's design. For example, the extent to which interviewers make use of extended context sensitive help functions or greater flexibility to use concurrent interviewing (see 4.7 below).

4.7. Other aspects of instrument design

Other on-going aspects of CPS instrument design include the standardisation and extension of dependent interviewing, concurrent interviewing and use of context sensitive help.

Dependent interviewing occurs when questions are asked which are dependent on a respondent's answer from a previous interview. The CPS will include at least three interview combinations which involve panel designs and elements of dependent interviewing. A standardised approach to designing and administering these questions, together with the processes used to 'feed forward' data from a previous interview is, therefore, required for the CPS.

The CPS will also incorporate wider and more flexible application of concurrent interviewing. This permits an interviewer, for example, to ask the same question of all household members in turn as an alternative to asking the whole questionnaire from start to finish for each respondent. Where appropriate, allowing individuals in households to be interviewed consecutively or concurrently (or in combinations of both) across all modules will improve standardisation and operational flexibility. It could also help to reduce interviewing time and respondent burden, particularly where information is collected from a single proxy respondent on behalf of all the household members.

[4] We gratefully acknowledge the help of colleagues at Westat and Statistics Canada

Finally, the CPS instrument proposes to expand the use of context sensitive help. This occurs where context sensitive information is made available in the Blaise questionnaire to the interviewer (or respondent) to deal with particular problems or queries as they arise during the interview. Benefits include improvements in data quality, reduced training costs, and reduction in paper instructions and documentation.

All changes in questionnaire design will be thoroughly tested and validated as part of the development programme outlined in section 5.

5. An evolving social survey system

5.1. CPS and ONS statistical modernisation

Delivery of the CPS is dependent on the development of a modular questionnaire and single Blaise instrument. However the CPS should also be understood in the wider context of the ongoing modernisation of ONS statistical tools and computing infrastructure. The CPS is dependent on the standard tools and systems that this programme will deliver.

The transition from a complex portfolio of surveys to a streamlined household survey system is illustrated in the diagrams below, where process 1 might represent the organisation and training of interviewers, process 2 the design and administration of questionnaires, and process 3 the editing, imputation and weighting of data.

Figure 2 provides an illustration of the extensive system of separate, but related, processes involved in each survey.

Figure 2 : Process model of existing surveys

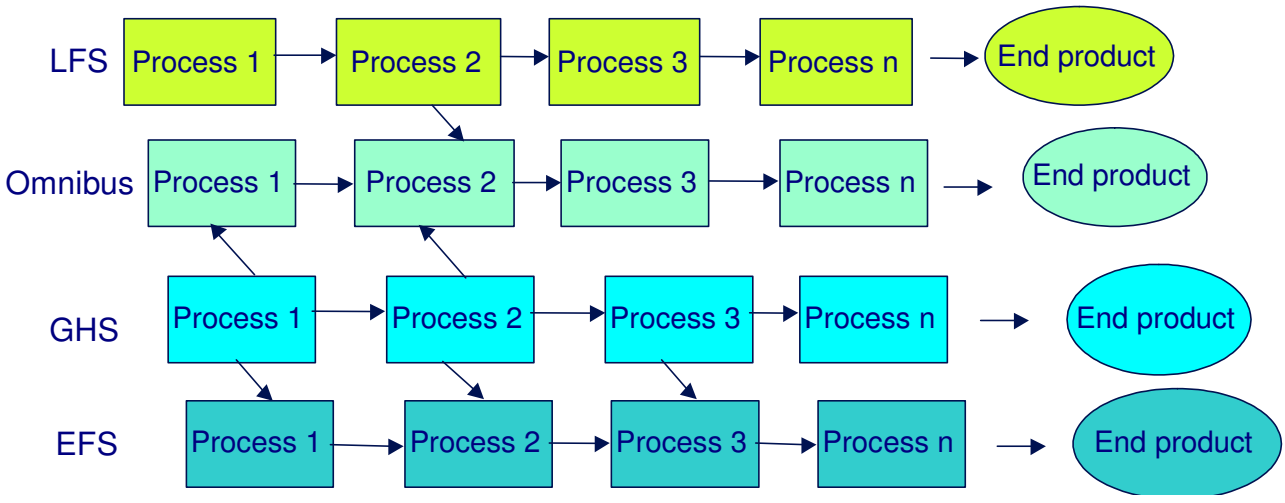
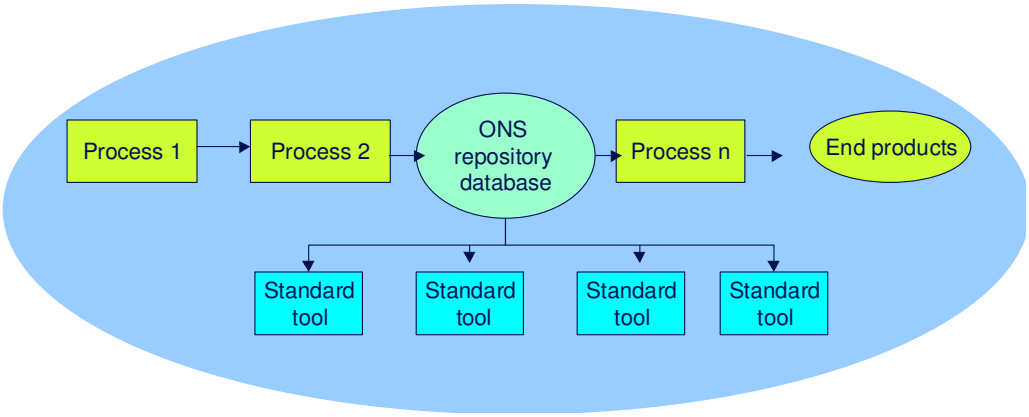


Figure 3 illustrates how the above model could be replaced with a single and efficient production chain. Data are derived from one questionnaire administered by a single field force, and stored in a Common ONS Repository Database (CORD). Processes like editing, imputation and weighting

are carried out using common tools to consistent standards. A range of end products are still derived, but from a common source.

Figure 3 : Process model for Continuous Population Survey



The CPS will depend on the delivery of these standard tools and the central database. In addition, it requires the delivery of a new Survey Case Management System (SCMS) to handle the flow of survey information and manage field operations more effectively.

5.2. Survey Case Management System (SCMS)

The new SCMS will support the modernisation of fieldwork operations that are necessary to provide successful data collection for the CPS. It will replace a range of older tools, with the aim of providing a more responsive and integrated system, together with improved information about data collection.

The SCMS will start with the assigned survey sample and cover every step in the process of information handling up to delivery of data to the central database, CORD. It will deal with allocation and re-allocation of addresses to all interviewers across all modes of data collection, and will manage the consolidation of disparate sources of data post-collection. It will also provide new systems to support the management of interviewers and, possibly and most significantly, real-time case management information. The design of the system will also ease the adoption of newer technologies, particularly remote connectivity, as they become available. Initially, the system will support household surveys but will be later expanded to support business surveys and the 2011 census too.

5.3. Development programme

The CPS development programme is designed to bring all these elements together to produce a comprehensive redesign of UK government survey taking and new a social survey system.

Currently, work is ongoing to consider a range of methodological issues regarding question development and data collection. Combinations of qualitative and quantitative techniques are being employed to investigate new questions, assess output quality and comparability, and, inform instrument design. A prototype Blaise instrument will be produced before the end of 2004.

A series of trials in 2005 and 2006 will test, develop, and validate all aspects of the proposed survey. These will begin with a small scale feasibility study involving a few hundred interviews,

and culminate in a large ‘parallel run’ involving many thousands. This trial will be central in taking the decision to go ahead with a CPS and discontinue the component surveys. For the CPS to proceed in January 2008 this decision can be taken no later than April 2007.

Annex A : Surveys for integration

The General Household Survey (GHS), the first UK multi-purpose household survey, started in 1971 and covers a wide range of social and socio-economic topics. The main aim of the survey is to collect data on core topics including housing, employment, education, health and family information.

The Expenditure and Food Survey (EFS) started in 2001 bringing together two surveys, the Family Expenditure Survey (FES) and National Food Survey (NFS), that were both well established and important sources of information, charting changes and patterns in Britain's spending and food consumption since the 1950s.

A Labour Force Survey (LFS) has been carried out in the UK since 1973 and in its present form since Spring 1992, providing a wide range of data on labour-market statistics and related topics such as training, qualifications, income and disability. In recent years the quarterly LFS has been supplemented by a series of annual boost samples in first England, then Wales and Scotland, known collectively as the Annual Local Area Labour Force Survey (ALALFS).

The Annual Population Survey (APS) is a boost sample in England which, when combined with results from the Labour Force Survey and LFS boost samples, will provide better local authority district estimates for key social and socio-economic variables than is possible from existing survey sources. The APS started in January 2004.

The National Statistics Omnibus Survey (OMN) is a regular, multi-purpose survey that started up in 1990 in order to provide quick answers to questions of immediate interest and information on topics that do not require a full, in-depth survey.

All the surveys attempt to conduct interviews with every household member, with the exception of the Omnibus survey where interviews are administered to one randomly selected household member only. The surveys differ in their use of Computer Assisted Personal and Telephone Interview (CAPI and CATI) and the extent to which proxy responses are permitted.

A breakdown by survey of the number of achieved interviews in the survey reference year 2003/04 is shown in Table 1 below.

Table 1 : Number of households interviewed in 2003/04

	thousands
Labour Force Survey	**86.5**
Annual Local Area Labour Force Survey Boosts	**87.0**
Annual Population Survey Boost [1]	**66.4**
Expenditure and Food Survey	**6.4**
General Household Survey	**8.6**
Omnibus Survey [2]	**14.0**
All	268.9

[1] Estimated [2] Currently one selected adult per household

Annex B : Illustrative diagram of a modular CPS