

# **Integrated Re-development of the ONS's Business Surveys and Business Register through the Business Surveys Integration Project and Register Re-engineering Project**

**Mark Pont, Pam Tate, Paul Smith and John Perry**

Office for National Statistics  
Cardiff Road, NEWPORT, UK, NP10 8XG. Email: mark.pont@ons.gov.uk

## **1 Introduction**

The present portfolio of business surveys of the Office for National Statistics (ONS) has grown up piecemeal in response to different demands and concerns, at different times, and within different organisations. These have been subject to various reviews and re-developments over time, most notably the creation of the consolidated Annual Business Inquiry in 1998 (Jones 2000), but most of the reviews have been specific to individual existing surveys, and have tended to focus on the detail of existing processes. ONS still has around 70 different business surveys, using some 800 different survey questionnaires.

ONS is currently modernising its infrastructure under the banner of the Statistical Modernisation Programme (SMP). ONS is also moving to common data management systems, and at the same time standardising its methods and implementing common tools with which to apply them, and systematising its processes. Through this programme, ONS expects to make considerable gains in both the quality of its outputs and the efficiency with which they are produced.

But we can say with fair confidence that further gains can be made, in both quality and efficiency, through rationalisation and integration of the wide and disparate range of our existing business surveys. Two major projects are involved in realising these additional gains: re-engineering the UK's business register (the Inter-Departmental Business Register (IDBR)), and the Business Surveys Integration Project (BSIP).

The BSIP has the broad objective of developing a redesigned and integrated portfolio of business surveys to meet the major requirements for economic statistics. This includes:

- revising survey methodology so that it applies current best practice in a consistent way and facilitates optimisation of survey design and coherence of outputs;
- meeting emerging needs such as those for extended and improved outputs at regional level, and for improved detail of the service industries sector; and
- increasing the efficiency and value for money of the programme of business surveys by making the best possible use of administrative data sources.

The success of this project is in turn heavily dependent on the extent to which the re-engineered IDBR supports these objectives, for example in providing a sampling frame with a capability to support a variety of stratification and estimation methods, and a unit structure that supports flexible data collection and can make maximum use of information supplied by businesses at varying levels of detail.

The paper describes the development of the two projects, and the ways in which the objectives of integration of the various surveys, and of consistency and mutual support between the integrated surveys and the Business Register have been addressed.

## **2 The Business Surveys Integration Project**

### **2.1 Dimensions of Integration**

The first step towards a Business Surveys Integration Project was to conduct a scoping study. This identified several dimensions over which integration should occur in surveys:

- over sector/industrial classification;
- over time, for example between annual and sub-annual data collection;

- by variable, for example combining questions relating to different variables in the same data collection instrument;
- between data from surveys and administrative sources;
- between national and regional level outputs;
- of methods;
- a more integrated approach to customer consultation; and
- a more integrated approach to the measurement and management of respondent load.

Within this framework, drawing on the findings from investigations of customers' needs and the experiences of other National Statistical Institutes (NSIs), along with potential uses of administrative data sources, the study identified a number of possible ways of improving integration along each of these dimensions.

## **2.2 The Integration Project in the Medium Term**

Each emerging possibility was tested against the following criteria, to check that it appeared to have the potential to contribute to ONS's strategic objectives, which are to:

- improve value for money for ONS;
- minimise the cost to respondents; and
- improve the fitness for purpose of the outputs in terms of the customers' needs.

Drawing together the most promising of the possibilities for improving integration, the following three elements emerged.

First, maintenance arrangements to support the IDBR would be redeveloped. A major part of this work is to develop a redesigned business register survey, integrated as far as possible with the collection of annual employment data, to form a Business Register and Employment Survey (BRES). This would be formed from two existing surveys:

- the Annual Register Inquiry (ARI), which collects employment and classification information about businesses' local units and is used to maintain business structures on the IDBR; and
- part of the Annual Business Inquiry (ABI), which produces detailed annual employment estimates from survey data relating to whole businesses.

Second, a Monthly Business Survey (MBS) is required to meet the needs for monthly data from businesses contributing to sub-annual indicators of economic performance. This will also provide timely measures of annual regional GVA growth. The MBS will be formed by merging several of ONS's current short-period surveys, and making better use of administrative data.

Third, a redesigned Annual Business Survey (Regional) – ABS(R) – will meet the needs for annual data from business surveys contributing to GDP estimates at both national and regional levels. A modular survey design is envisaged, with a core or spine of key variables that are gathered for the whole sample, and with more specialised or detailed modules each gathered from a different subsample, though the full realisation of this design will need a further phase beyond the period of the BSIP.

## **2.3 Design Issues and Methodological Implications**

A major factor influencing the design of all three of these future surveys is the increasing demand for regional level outputs. A fundamental requirement for meeting this is the availability of more good quality data pertaining to the individual sites or local units of the more complex businesses that have locations in more than one region.

The first step in achieving this is to establish what data such businesses have and are able to provide about their local units. This is likely to vary by industrial sector and from business to business. It will be important to capture all available data, irrespective of such variations, so that the information can be used both directly in regional estimation, and indirectly in constructing models for use in imputation of local unit data where this level of detail is not directly available.

To support this, the redesigned Register will need to hold a wider range of variables to support more stratification and estimation methods than currently, and a different definition of businesses' structures that supports flexible data collection

and can make maximum use of information supplied by businesses at varying levels of detail. In turn, the design of the BRES survey must enable it to keep the structural information about complex businesses up to date.

Finally, the increases in survey samples suggested by the need to produce more precise regional estimates will need to be offset as far as possible by making maximum use of administrative data sources, and the Register will need to support the hybrid survey/administrative data arrangements that this implies.

### **3 Business Register Design Issues**

#### **3.1 The Context of Register Redesign**

The IDBR is the list of the UK's businesses that is used for selecting samples for surveys of businesses and to produce analyses of business activity. It covers all businesses registered for value-added tax (VAT) or operating a Pay As You Earn (PAYE) income tax scheme. Businesses in the agricultural sector are recorded on the IDBR, although the Department for Environment, Food and Rural Affairs (Defra) is responsible for collecting agricultural statistics. Construction businesses are also on the IDBR, and both ONS and the Department of Trade and Industry (DTI) publish data relating to the construction sector. While the IDBR covers the whole of the UK, separate legislation covers the parts relating to Great Britain (GB) and Northern Ireland (NI). ONS is responsible for maintaining the British part of the Register and the Department of Enterprise, Trade and Investment (DETI) in Northern Ireland dealing with the Northern Irish part.

The two important statistical units in the present context are the *enterprise* and *local unit*, each of which is defined precisely in the European Council Regulation on the statistical units for the observation and analysis of the production system in the Community – EEC 696/93 (EEC 1993). In addition, we have the *reporting unit*, which holds the mailing address for the business and is the unit for which businesses report their survey data to ONS. In general, the reporting unit is the same as the enterprise. In some of the more complex cases, enterprises are subdivided into reporting units according to activity or geography, and are defined according to the appropriate local units from within an enterprise.

Sampling is carried out using a Permanent Random Number (PRN) method (Ohlsson 1995), with stratified simple random samples taken. Most business surveys in ONS are stratified by the dominant industry as specified by the Standard Industrial Classification code (SIC(2003)) and the employment of the business. In some surveys there is stratification by country within the UK and legal status (whether the business is in the private or public sector). The current selection system permits only one size variable to be used within a survey, and definition of strata using non-consecutive SIC(2003) codes is provided only by complex clerical procedures. Therefore, the current stratification system, whilst meeting users' needs in the early 1990s when it was developed, does not offer sufficient functionality to meet today's wider information demands.

The unit chosen for sampling in most surveys is the reporting unit. In general this equates to a whole enterprise, but in some complex cases, enterprises are partitioned into two or more reporting units where this enables: (a) easier completion of our questionnaires for the businesses concerned; and (b) easier analysis in ONS because the data provided are more homogeneous with respect to industrial activity. About 600 enterprises are split (into about 3000 reporting units) according to these criteria. A further 700 enterprises are split into separate reporting units for Great Britain and Northern Ireland because of the legal requirement for GB and NI data to be collected separately by ONS and DETI respectively.

These arrangements mean that there is some inconsistency between the levels used for selection and data collection, and restrict the scope for setting up different data collection structures for different surveys. Furthermore, the system is sensitive to businesses restructuring themselves, in which case potentially important linkages between the old and new structures may be lost.

ONS is part way through its Statistical Modernisation Programme to update its computing and statistical infrastructures. As part of this, the business register system is migrating from Ingres to Oracle. An improved interface for users and greater compatibility with other systems are expected to be the main benefits. The new Register will also allow greater flexibility for sampling and data collection, and will be better geared towards meeting new legal requirements of the European Union. The exact scope of the re-engineering work is currently being defined with users of the Register and methodologists.

### **3.2 The Implications of Integration**

ONS's Methodology Directorate (MD) was already aware of various methodological issues where development was needed to support the potential for greater flexibility, responsiveness and efficiency offered by the register re-engineering project.

One of these was the need to split the sampling and data reporting functions of the Register, in order to minimise the disruption to register linkages caused by restructuring of businesses and to increase the flexibility of data collection arrangements. This dovetailed with BSIP's need to maximise the use of the full variety of information that businesses could supply. The common solution includes a more complex and flexible model for businesses on the IDBR, and a flexible system of data collection or observation units, separate from the business enterprises which are expected to be the default sampling units.

Register structures and the linkages between the various business units are represented by a units model. As part of the re-engineering of the IDBR, and in response to changes in the European Regulations and the availability of data from new sources, the units model underpinning the Register is to change. Both the existing and proposed models are included as an appendix to this paper. The proposed use of the different units for sampling and data collection are described in the next section of this paper.

Another issue was the need for greater flexibility in stratification, in order to make best use of the more varied data in estimation, and to respond to emerging customer needs for more detailed estimates in an efficient and robust way. The main capabilities identified were the ability to use more stratification variables, including new size and/or structure classifications, and more flexibility in setting the detailed levels of stratification variables.

A third area was the need for improved procedures for updating the Register, to balance the needs for accuracy, timeliness and stability of register data, and minimise the effects on time series of 'shocks' arising from updating.

A collaborative effort between MD, the Business Registers Unit and the Designing Sources Division, (DSD - responsible for the BSIP) developed the specifications for the re-engineered Register to be able to support these methodological developments.

### **3.3 Data Collection Entities**

Part of the basis for re-engineering the IDBR is a revised model for the structure of businesses. The changes are in response to a range of developments. The major ones are described next.

First, the need for the IDBR to migrate to new systems will require some reorganisation of the data held on the IDBR and the way in which it is organised.

Second, the IDBR operates under European legal constraints and a new business registers Regulation is being prepared. The main changes compared with the current Regulation are the compulsory inclusion of agriculture and public administration and the requirement to record the structures of enterprise groups. Within agriculture, the IDBR currently holds only administrative data but a project is under way in the UK to link data from the separate farms holdings register that is maintained by Defra. ONS's development of a quarterly survey of public sector employment towards the end of 2004 (Hicks and Lindsay 2005) has resulted in improvements to data for public administration on the Business Register. The IDBR holds comprehensive data on enterprise groups, based mainly on the commercial source of Dun and Bradstreet, and the new regulation would require only the re-organisation of the information already held. Thus, ONS is in a strong position to respond to these new demands.

Although the IDBR currently meets most of the proposed requirements, it will be important to separate multi-national enterprise groups from their domestic truncated parts. A better operational definition of the enterprise has also been agreed, which combines administrative units under common control to form the truncated enterprise group and then derives the enterprises within this group by profiling, while recognising the constraint that legal units (company registrations in the UK) should not be split across enterprises.

Third, there are growing demands for statistics at the regional level and lower, boosted by the findings of the recent *Review of Statistics for Economic Policymaking* by Christopher Allsopp (Allsopp 2004). One feature of this is the increasing interest in new types of analytical units that group the local units of an enterprise within a specific geographical area or economic activity. Allsopp's Review also recommended that ONS should make better use of data from administrative sources for the production of business statistics. The new Business Register will therefore play an increasingly important role as the gateway for administrative data relating to businesses, and the hub to facilitate the linkage of administrative and survey data.

Fourth, there is an increasing national and international demand for data on business demography based on statistical business registers. One consequence of this is that register coverage is now being increasingly defined in terms of numbers of businesses rather than just the proportion of economic activity. It is estimated that there are around 1.8 million 'businesses' not on the IDBR. These are very small – they have no employees and their turnover is below the level at which they need to register for VAT purposes – and some may not meet the full criteria of autonomy required for enterprises. The new model, however, recognises this, and allows for the addition of such businesses where they are corporate. In the longer term, links to self-assessment tax registrations indicating income from self-employment, could facilitate coverage of the remaining non-corporate enterprises.

At the same time it is proposed to enhance the flexibility of data collection structures to meet changing user requirements. As described in section 3.1, the current sampling unit is the reporting unit, which does not always correspond exactly with the enterprise. Under the proposed arrangements, the sampling unit will be the enterprise. As noted from section 3.1, the IDBR's reporting units serve a dual purpose; they are both sampling and collection units. At present, around 1300 of the largest enterprises are split on this basis, over half of which have a GB/NI split for legal reasons. The introduction of data collection entities (DCEs) will ensure that one set of units is dedicated to data collection. These units will be built to maximise the usefulness of data requested from surveyed enterprises, while ensuring that sampling takes place from a consistent unit level for all businesses (usually at enterprise level). The use of a standard unit for sampling is an important aspect of the co-ordination leading towards data integration. However, enterprises are a creation of the statistical agency - operating entities are the real source of data. Furthermore, although each statistical entity is linked to an operating entity, the operating entity may not be able or wish to provide all the data required by a survey for the corresponding statistical entity, or may wish to provide data in different ways for different surveys, for example. Thus, for each individual survey, there is a need to identify the arrangements by which the data items are to be acquired for each enterprise in the survey sample. These arrangements are the attributes of a DCE. The attributes are divided into three sets:

- i. **Coverage:** defining the relationship between the entity from which the data are being acquired and the statistical entity for which the data are required;
- ii. **Collection Instrument:** the means of obtaining the data, eg questionnaire, telephone interview, administrative record, etc; and
- iii. **Contact:** the respondent name, address, and telephone number within the business operating structure.

In addition, DCEs provide the standard channel by means of which frame data obtained during the course of a survey process can be fed back to the central database. The DCE is also the vehicle for monitoring respondent contacts and assessing respondent burden.

DCEs are created only for selected businesses and are survey-specific. Ideally, we want to generate as many as possible automatically by an algorithm based on statistical entities of the IDBR, associated administrative links, and how data are to be collected from the respondent. They can be modified manually, if need be, to take into account information fed back from respondents. Further work needs to be done to evaluate the extent to which it will be possible to automate the system, given the diversity of structures of businesses in the UK.

DCEs will range in complexity. Most enterprises operate at a single location, so the DCE structure in most cases will be very simple to reflect this. In many of the more complex cases, we expect DCEs to be defined either on a regional or an industrial basis, or by region and industry. In the most complex cases, more flexibility may be required in order to define DCEs along a mix of industrial, regional, organisational and other lines.

The increased importance of data collection structures places greater emphasis on the process of identifying and maintaining the structures of such businesses, known as profiling (see section 3.5). Profiling needs to be sufficiently well-resourced and responsive to deal with changing reporting requirements of businesses.

### 3.4 Auxiliary Variables for Stratification and Estimation

Stratification in ONS has through custom been restricted to rectangular strata, almost always a strictly rectangular array, and stratification by a very restricted range of variables. As a result of the historical focus on activity and the proliferation of activity codes in coding systems such as NACE and SIC, we have used very detailed industrial stratification, which in some surveys has meant that the population of businesses in many strata is very small. Regional stratification, which has only been used in a few surveys, has had to use the *dominant region* (that is the region in which the largest part of the activity falls) for stratification. This means that stratification does not cleanly divide the population into regions, because the activity covered by the sampling units is often spread across several regions.

In order to implement the recommendations from Allsopp's Review, region becomes a more important stratification variable. Neither the population nor the sample size is big enough to support adding regional strata to the current detail, so we need a new strategy. We therefore need to ensure that the re-engineered IDBR is sufficiently flexible to accommodate the range of stratification options that may be required in the future. The basic planks of the new strategy are as follows.

- Strata of sufficient size. There are various stratification options, but the target is to ensure that strata have a large enough population size to support robust sampling and estimation based on adequate sample sizes. Cochran (1977) recommends around 30 units for large-sample methods to be valid, but in practice such large sample sizes are impractical with the current structure of ONS's business surveys. In recent redesigns (notably for the Retail Sales Inquiry) we have targeted an achieved response (that is, after expected non-response has been factored in) of 10 questionnaires per stratum, as an improvement over previous designs which might have a minimum as small as 5 *sampled* (not responding) units. This doesn't really approach Cochran's rule of thumb, although the use of a combined ratio estimator can ensure that the number of responses used to estimate ratios is large enough. It is our intention under the new scheme to ensure larger minimum sample sizes.
- Reduced detail of industry stratification and a more versatile use of the different levels of the SIC according to size of populations, homogeneity of SIC classes etc. In particular, the 2007 revision of the SIC will have more detail in the services sector than we have had previously. This is likely to be closely related to the sample size requirements in the previous point. The main message is that reduced industrial stratification is the price for additional control at a regional level, and this will mean increased variability in some estimates and the use of domain estimation techniques.
- The availability of additional variables for stratification and estimation. Ideally these will be easy to create, so that stratification is not limited to a small set of variables. Rather, any information, or combination of information, should be available for stratification as long as it is available on or through the Register.
- Taking into account the complexity of businesses in stratification. Given the increased demand for regional statistics, it may be more important to ensure appropriate coverage of businesses that span more than one region or more than one industrial sector.

### 3.5 Business Profiling

Business profiling is the method through which the ONS analyses the legal, operational and accounting structure of an enterprise group, in order to establish the statistical units within that group, their links, and the most efficient structures for the collection of statistical data in ONS's business surveys. It gives a better understanding of complex unit structures than the statistical surveys. It is a relatively expensive process that needs to focus on those businesses most at risk and requires experienced staff. With the inevitable resource pressures experienced by the ONS, the profiling plan has to be a compromise between quality and quantity.

Business Profiling in ONS is performed through a dedicated profiling team comprising 14 staff. It processes around 500 enterprise group structures each year, including some in the public sector, through a mix of desk, telephone and visit profiling. Approximately 3 per cent (65,000) of enterprises on the current IDBR have more than one local unit, and most of these have a relatively simple structure. Though complex units are small in number, they do, however, tend to be important players in the economy and have a large impact on business statistics outputs. By focusing on those that have more than twenty employees and either a presence in more than one region, or more than one significant economic activity, the number that would not fit the default structure and would thus require profiling is around 20,000. While this would not be possible through normal business profiling, the Business Register and Employment Survey (see section 2.2) will provide, at least annually, the necessary check of the structures of the additional businesses.

### 3.6 Register Updating Issues

A perfect IDBR should reflect the real business world. This means that its structures (statistical and administrative), classifications (eg industry, geography), linkages, data content (eg employment), activity status (live, dead, and dormant) and contact information should be as close to reality as possible. However, this is unrealistic and a register will suffer from quite a few deficiencies. Some of them include: lags in identifying births and deaths of businesses, inaccurate classification, outdated information, and inappropriate linkage of data collection unit to the sampling (statistical) unit.

Administrative sources are the main sources for updating small enterprises. Large enterprises are also based on administrative sources, but industrial classification, structure and size indicators are all based on data collected in the ARI at local level, and aggregated together. The new BRES (section 2.2) will fulfil the same function. Births and deaths of businesses are updated on the IDBR as soon as ONS is notified of them, although there can be quite lengthy delays between the event happening in the real world and ONS being notified of the event. In addition, there are major updates of the main size variables – turnover and employment – each year, although some changes are made as they occur.

A feature of the IDBR is the use of ‘frozen’ variables for SIC (2003), turnover and employment. Historically, industrial classification codes were updated immediately. However, this resulted in instability of sample membership, with reporting units moving in and out of samples because of reclassification, rather than because of sample rotation. To address this, extra variables were created. In addition to the existing ‘current’ data which would continue to be updated continually, a new ‘frozen’ variable was introduced that would be updated only once a year, apart from in exceptional circumstances. The intention was that the frozen variables would be used to determine strata for sampling in monthly and quarterly surveys. Thus, reporting units would remain in the same stratum throughout the year. Where it was known that a business had been reclassified during the year, the current field would be updated. This would permit questionnaires appropriate to the new industry to be sent and to assist in domain estimation. The main consequence of this is that short period surveys have to deal with one major shock to the Register once per year, rather than smaller shocks on each survey occasion.

Administrative data sources provide a smooth, continual update of births on the IDBR. There are time lags between the physical start-up of a business and its appearance on the IDBR. The classification of business information relates to items such as industry, institutional sector, size (turnover and/or employment) and location. These items are often used in the survey design for stratification and selecting the sample for a survey. The impact of inaccurate classification is that it could lead to units being placed in the wrong stratum and contributing to the wrong industry (or even survey), units being sent the wrong questionnaire, and/or imputation being based on incorrect information. If a large business is misclassified, this can add both volatility and bias to the outputs, and the business will often need to be treated specially. Finally, if corrections have to be made to classifications, this could result in revisions to previously published statistics.

Size classification (number of employees) is updated for the local units by the ARI. These are aggregated up to associated enterprises. These updates occur for all larger units selected with certainty. However, for the remaining units only a sample is updated on an annual basis. This implies that the size stratification for the remaining units has a mixture of currency, from one to ten years out-of-date. Data from the PAYE system provides a standard measure of size that is available for updating employment on the IDBR for those businesses that are employers, but this is not available for local units.

## 4 Data Availability from Businesses

As noted in 2.3 above, the first step in capturing and using more data on local units of multi-site businesses, and hence laying the foundations for improving the quality of regional estimates, is to establish what data the businesses have available. During early 2005, ONS conducted about 100 interviews with a range of businesses to study the availability of data at local unit level, for the following variables: employment; turnover; capital expenditure; inventories; purchases; employment costs; profits and the level and cost of production.

Quite surprisingly, the study found no real differences between economic sectors or the size of business. The main determinant was the availability and use of computer software, particularly accountancy software. In general, the study found that turnover, employment and employment costs were fairly widely available from businesses, or were easily obtainable upon demand, with the ease of production being related to the level of computerisation. Data were available from a minority of local units relating to purchases; profits; inventories and capital expenditure. Again, availability depended on the level of

computerisation. Finally, information about the costs of production was largely unavailable for local units since records tended to be kept centrally.

This study was quite small, and didn't cover the largest businesses. Furthermore, the study didn't go as far as collecting data from businesses and the practice may be different from the theory. However, this work gives a reasonable platform upon which to base a system of the collection and production of regional statistics. Since data appear to be largely available at the lowest level, they can be aggregated to any appropriate regional definition. Further work is needed in this area to get a clearer understanding of these issues, particularly for the largest businesses.

## **5 Use of Administrative Data Sources**

### **5.1 Current Uses in the Business Register**

As described in earlier sections, the IDBR is currently based mainly on inputs from two administrative sources: traders registered for VAT, and employers operating a PAYE scheme. Both sources are administered by Her Majesty's Revenue and Customs (HMRC), which provides details of 1.7 million businesses registered for VAT through daily updates and 1.1 million PAYE employers through quarterly updates. Other administrative sources currently used include data on company registrations from Companies House, which are mainly used to aid matching, and data on enterprise group linkages and foreign ownership from Dun and Bradstreet.

VAT and PAYE are used to identify new businesses and to provide size and industry classifications for them. Company registrations are used primarily to help with data matching. VAT also provides a regular source for updating the turnover of businesses. VAT traders generally report quarterly, although some make monthly or annual returns. PAYE employers are obliged to inform HMRC when an employee joins or leaves a scheme and HMRC then supplies this information quarterly to the ONS. In addition to using the data for direct updating of register information, the IDBR acts as the gateway for others to access this information.

### **5.2 Existing and Potential Use as Auxiliary Variables**

ONS's business surveys have used administrative data in estimation to increase accuracy for many years. This has been mainly turnover from VAT, but employee figures from PAYE are also used in some cases where employment values are unavailable from other sources. Wider use of the PAYE source as a stratification or auxiliary variable has been proposed, and this will be researched as part of this project.

### **5.3 Use for Supplementing and Substituting for Survey Data**

There is potential to make more use of existing data to replace parts of the survey collections. This is likely to result in some loss of quality (perhaps definitional, perhaps timeliness), but with the potential to reduce the burden on respondents and perhaps to increase the amount of data available.

This approach is already common practice in other major NSIs, for example Statistics Canada, where tax data is being used to replace survey data in both annual and sub-annual surveys. For the main annual business survey, the typical approach is to use data from business income tax returns, and especially from the business accounts data submitted in support of those returns, from some or all of the small, simple businesses in the survey sample. For the sub-annual surveys, the typical approach is to use data from VAT/GST returns to substitute for some or all of the small, simple businesses in the survey sample. ONS is currently evaluating company accounts data that may be used instead of survey data for small or medium-sized businesses in annual surveys and the possibility of using the VAT turnover data that ONS already receives to substitute for survey data in the proposed Monthly Business Survey. The use of the PAYE employees data will also be explored.

In some cases it appears that legislation may be needed to increase the availability of such data to ONS, but at a time when we are rebuilding ONS's survey systems, now is the opportunity to make sure that those systems are sufficiently flexible to permit the easy incorporation of administrative data and associated methods.



## 5.4 New Sources

Some of the changes to the units model for the new Register that we described in section 3.2 reflect the emergence of new sources of data for the Register:

- the elevation of the company registration to the same prominence as data on VAT and PAYE registrations as a source for register units and variables;
- more flexibility to incorporate data from other sources in the future; and
- links to a wider range of satellite registers.

Satellite registers are envisaged mostly in the financial services area, where ONS surveys currently rely on data from various regulatory bodies to provide sampling frames. It is proposed to match such sources directly to the IDBR to create satellite registers. This will allow samples for these surveys to be selected directly from the IDBR, providing auxiliary variables to aid stratification and estimation. Furthermore, integration of the financial services surveys into the Register allows better control of sample overlap between all surveys.

This approach is in line with ONS's administrative data strategy, which sees the IDBR as the gateway for bringing all administrative micro-data relating to businesses into ONS. Work on validation and matching of business micro-data from external sources will be centralised around the IDBR. This will have the benefit that administrative data will be linked directly to statistical units, and can this be more readily linked to data from other sources, enhancing the possibilities for further analysis.

In the context of the implementation of the recommendations of Allsopp's review of economic statistics, work has recently started to evaluate potential sources of local unit data from both the public and private sectors. A major source that ONS is currently investigating is data from the UK's Valuation Office Agency about floor-space and rateable values (a figure that reflects the amount of rates that is chargeable) for each business premises. Such data may be useful in recasting enterprise-level survey responses to local units for use in regional estimation.

There is currently only limited information available from Companies House but annual accounts will soon be provided in electronic form. Access to corporation tax and self-assessment tax data for the self-employed held by HMRC is not permitted under current legislation.

## 6 Future Development

The difficulties for methodologists and survey designers relate to the trade-off between future-proofing systems when they are redesigned, and a very restricted budget for re-design. It is therefore inevitable that the flexibility required in the future has to be anticipated to some extent, and implemented in as generic a way as possible. Not least of our concerns is that many of the methodological changes suggested need a good deal of research and evaluation before they can be applied in practice.

A paper presented to ONS's Methodology Advisory Committee in May 2005 (Hidioglou et al 2005) identified the following as major issues for consideration:

- the frequency of updating the Register with new structural information, and whether sub-annual surveys should be based on continuously updated or 'frozen' information;
- whether to use a model to update employment values for the non-surveyed portions of the Register;
- the periodicity and the priorities of various sources in register updating, and the evaluation of the effects of frame updating;
- the approach to dealing with deaths and accounting for unrecorded births;
- the trade-off between finer level industrial stratification, and broad stratification together with domain estimation;
- the relative merits of different approaches to small area estimation; and
- the optimum approaches to the use of data from administrative sources, especially regarding the extent to which these data could replace survey data.

## References

- Allsopp, C. (2004) *Review of Statistics for Economic Policymaking*. Norwich: HMSO.
- Cochran, W. (1977) *Sampling Techniques*. New York: Wiley.
- EEC (1993) Council Regulation No 696/93 on the Statistical Units for the observation and analysis of the production system in the Community.
- Jones, G. (2000) The Development of the Annual Business Inquiry. *Economic Trends*. 564, Nov., pp. 49-57.
- Hicks, S. and Lindsay, C. (2005) Public Sector Employment. *Labour Market Trends*. Vol 113, no. 4, pp. 139-147.
- Hidiroglou, M., Perry, J., Pont, M., Smith, P., Tate, P. (2005) *Methodological developments to support Allsopp's objectives for economic statistics*. Paper presented to the 8th ONS Methodology Advisory Committee meeting. ([http://www.statistics.gov.uk/methods\\_quality/downloads/NSMAC8paper4.pdf](http://www.statistics.gov.uk/methods_quality/downloads/NSMAC8paper4.pdf))
- Ohlsson, E. (1995) Co-ordination of samples using permanent random numbers. In *Business Survey Methods* (ed. B. Cox, D. Binder, B. Chinnappa, A. Christianson, M. Colledge and P. Kott) pp. 153-169. New York: Wiley.

Figure 1: Proposed Units Model

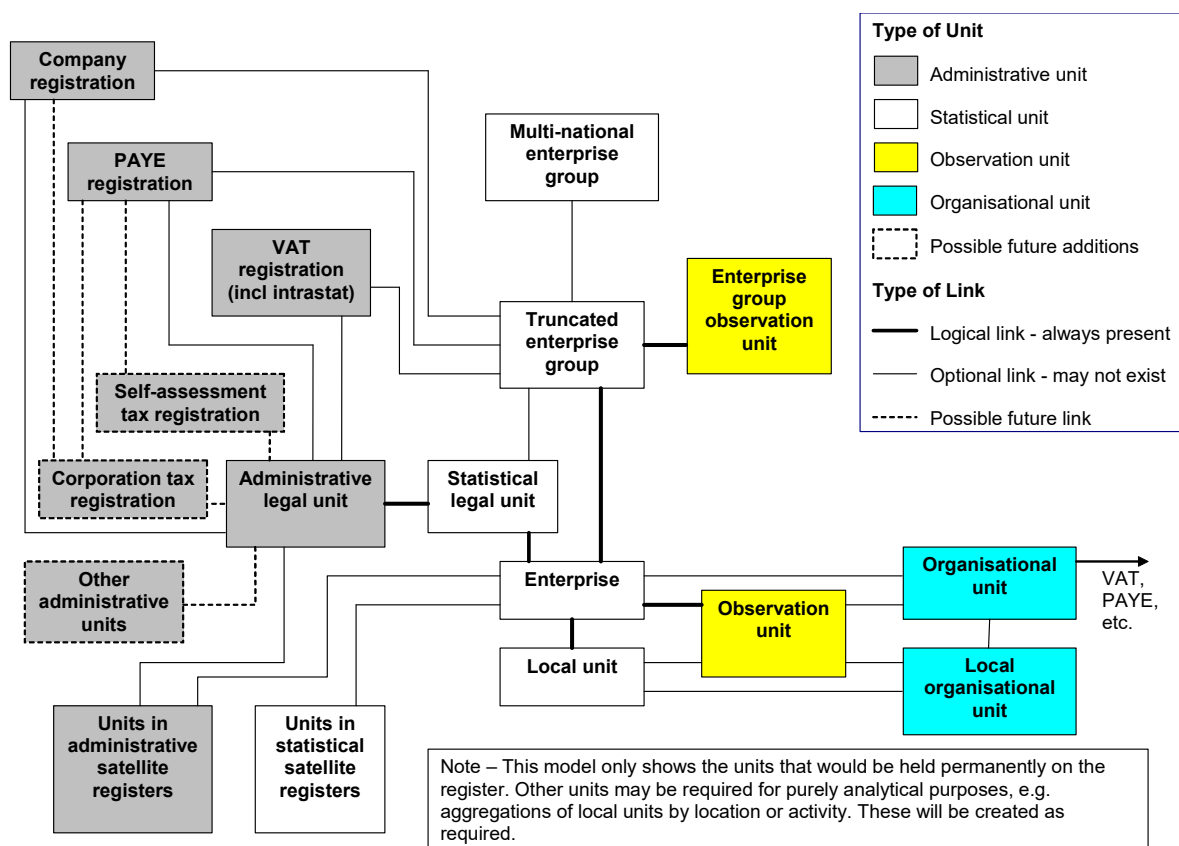


Figure 2: Current Units Model

