



Development of a Data Validation Unit for the Office for National Statistics (ONS) Business Surveys

PURPOSE OF THIS PAPER

Since 1994 the ONS has pursued a programme of change in the area of data collection and validation activity for business surveys. The objective was to bring together from its dispersed position all the activity in data collection and validation within subject areas to within a single command, creating a horizontal split between data collection and validation on the one hand, and the production and publication of results on the other. This involved a significant change in the way the Office operated.

This paper describes why ONS moved to its current structure for collecting and validating data, how the change was managed, the benefits of the new structure, and the lessons learnt from undergoing such a process.

WHY CHANGE TO A HORIZONTAL STRUCTURE?

There were two main drivers for change:

- The desire to develop a functional structure and through this, provide a separate focus on data collection and the production of results, and to allow better use to be made of specific expertise in these areas.
- The need to deliver efficiency savings in the collection and analysis of data often through standardisation and rationalisation, in order to provide funds for new initiatives, particularly in economic statistics.

MANAGING THE CHANGE

Prior to 1994, business surveys had been conducted as independent processes with the same staff working on the survey from the selection of the sample to the publication of results. The move from this vertical structure to a more horizontal approach was conducted in two stages.

Stage one - 1994-1998

The first stage was to separate data collection and validation activity away from analysis and publication for each Newport and Runcorn based survey. This brought the management of data collection activities for a number of surveys under common senior management. At this stage there were four Data Collection Units (DCUs), as they were then called, each providing the input data for a particular Results, Analysis and Publication unit (RAP), under the command of four Divisional Directors.

The most crucial element of this first stage was establishing the boundaries between the two units - DCUs and RAPs. The split between DCUs and RAPs is based on the

DCUs preparing data sets for use by the RAPs by clearing micro level issues with the data, often by recontacting the business. The RAPs look for macro (aggregate) level inconsistencies in the data, which may involve the DCUs carrying out further micro level investigations.

This first stage was achieved by process mapping the main business surveys, establishing clear boundaries between the two activities and negotiating Service Level Agreements (SLAs). SLAs between internal customers and suppliers was a relatively new concept in ONS at the time and posed quite a negotiating challenge. Today internal SLAs are in common use for business surveys and a standard template exists to ensure that all aspects are covered and in a consistent way. Once the data validation activity had been segregated it was possible to plot the peaks and troughs of various surveys, and by grouping compatible surveys together, provide staff with a constant flow of work. This arrangement was dependant on developing multi-skilled staff, competent in working on more than one survey.

Data capture improvements

This first stage also saw many improvements in data capture techniques. Shortly after the creation of the Data Collection Units, pilot projects were set up to test the feasibility of capturing data electronically. The following pilots were set up:

- **Document imaging** - Document imaging has been gradually applied to ONS business surveys since 1995 and has now been applied to 95% of business questionnaires received. The vast majority of business data returned to ONS is now computerised from the moment of entry into the Office, removing the need to handle and store hard copies of survey questionnaires and correspondence with staff working with the computerised images instead.
- **Telephone Data Entry** - Since 1995 data from forms with only small numbers of variables have increasingly been entered via Telephone Data Entry (Touchtone) technology. In 1995 the first pilot of Telephone Data Entry (TDE) was applied in the collection of data for a new Service Sector Price Index. The pilot proved successful and a programme to rollout this method to other suitable surveys such as the Producer Price Index, and the Retail Sales Inquiry, was introduced.
- **Use of Lotus Notes** - The organisational changes in data collection and the move towards early electronic capture of data and the use of document images and TDE gave rise to demands and opportunities for improvements to working procedures. Lotus Notes has been used as a common tool to reinforce the move away from single survey workgroups and to allow processes similar across surveys to be dealt with in a consistent way. Notes has been used to set up a shared database to record communication with respondents. This database allows all surveys to share the soft information arising from comments on questionnaires, conversations with respondents or desk research into business organisation. This has reduced the effort required to explain unusual data and eliminated duplicate telephone calls to a single respondent arising from data queries for different surveys. Notes has also been an excellent tool for work allocation, prioritisation and monitoring, which has eased the management of work areas with responsibility for a mix of short term and annual surveys.

Benefits of Stage One

The creation of the Data Collection Units (DCUs) involved significant change in the way the Office operated leading to the following benefits:

- Savings of between 40 and 50 per cent were achieved in the whole collection process;
- Improvements in data capture techniques - over the 5-year period during which pilot systems were fully implemented, annual costs were reduced by £1m (for a total investment of £800k over a 4 year period).;
- More effort was placed on the respondent relationship as well as on producing results;
- It became the responsibility of some staff to capture and validate data only and to view the operation from that perspective and this focus has led to improved operating practices often by building on what was working best for particular surveys.

Stage two - 1999-2002

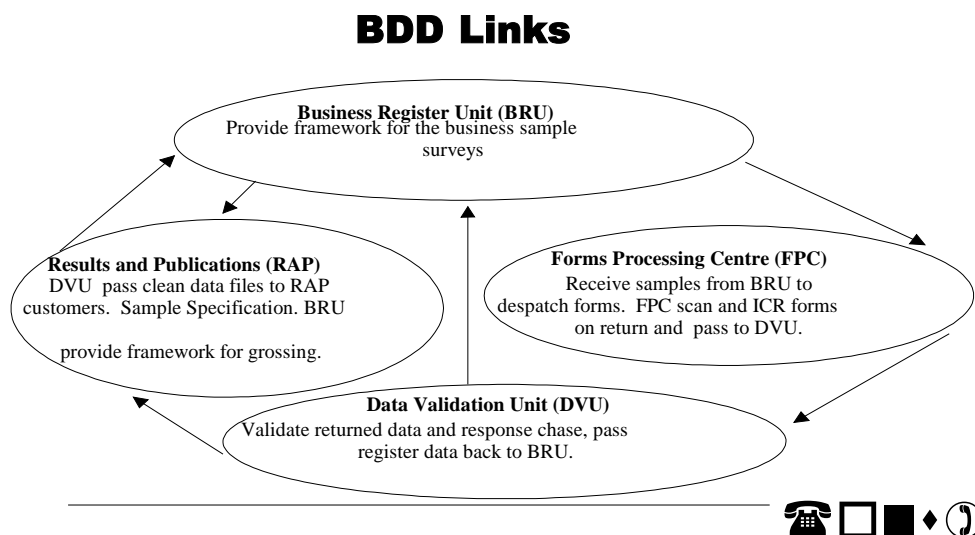
By the end of 1998 the first phase of data collection initiatives were completed and the DCUs had been operational for over four years. The next stage was to bring together the four DCUs into one division (Business Data Division) under one management command and to develop a new 5 year plan for Data Collection Initiatives. Since the creation of the DCUs in 1994, a number of financial, international and innovation surveys had been transferred to the Newport site from the London Office. It was agreed that these surveys should follow the same process Newport and Runcorn based surveys, beginning with the separation of functions whilst remaining outside the new Business Data Division.

Once again, the main driver for this next change was to provide further efficiency savings. An efficiency study conducted in 1998 set Prices and Business Group (which managed the collection and results work for business surveys) the objective of reducing its baseline by £1m in real terms over the period April 1999 to March 2002 through efficiency savings in its data collection activities. The aim was to deliver efficiency savings through re-engineering processes, extending the use of technologies and automating data processes - not through cuts that could lead to a reduction in data quality. An important part of the process was to ensure that the customers for the data collected were closely involved with the efficiency related initiatives. A project management structure was set up to ensure that all stakeholders were involved and that each initiative was properly evaluated to ensure that savings produced did not have an adverse impact on the quality of the Prices and Business Group outputs.

The first stage in the process was the creation of the Business Data Division (BDD) in April 1999. This new division took responsibility for all operational aspects of business surveys comprising of three operational units:

- **Data Validation Unit** - bringing together the four previous DCUs to be responsible for data validation and response chasing outstanding questionnaires under one management command, whilst still retaining four Data Validation Branches each responsible for providing input data for a number of surveys. The change of name from collection to validation was introduced to highlight that the work of the unit was not simply data capture.
- **Business Register Unit** - responsible for maintaining the Inter-Departmental Business Register, the sampling frame for the majority of business surveys.
- **Forms Processing Centre** - responsible for forms printing and despatch and the receipting of returned questionnaires.

The diagram below illustrates how the three BDD units linked with each other and with the Results, Analysis and Publication (RAP) customers.



Reviewing processes and the organisational structure

The creation of BDD provided the ideal opportunity to review processes and the organisation of work within the three main units. Centralising work and streamlining processes was a common theme. Centralising activities in one area has resulted in further common processes and standard working practices being introduced. Merging smaller teams helped to even out peaks and troughs and ensured essential back-up systems were in place. Economies of scale made possible through centralising areas such as the four Data Validation Branch's planning and support areas, merging the Complex Business Unit and the Business Profiling Team and creating a central contact area for handling complaints, produced quality improvements as well as efficiency savings. Process mapping was again used to streamline processes, such as the treatment of Business Register amendments and to help with a review of the Forms Processing Centre located within BDD. Again, quality improvements were achieved through these initiatives, not just efficiency savings.

Introducing new methodological approaches to data validation

This stage also explored the use of statistical methods to automate parts of the data processes. The project concentrated on short-term improvements to existing systems, rather than rewriting them. Its aim was to develop and apply new methodological approaches to improve the efficiency of data validation and editing without adversely affecting data quality. Two successful methods have been developed:

- **automating the correction of systematic errors** - Automatic editing takes place as part of batch data take-on. Where a total does not equal the sum of its component parts and the difference is within a set tolerance, the total will be automatically adjusted by the system. Financial data are collected in £'000s. If it is evident from previously supplied data that the business has provided actual data, rather than data rounded to £'000s, then the returned data will be automatically adjusted by the system. Automatic editing was piloted on one business survey and the impact on the quality of the outputs was shown to be negligible (or zero) using estimates of the variance and bias introduced into the estimates using the current estimators and current estimation methodology. Following the success of this pilot automatic editing was successfully rolled out to all appropriate business surveys.
- **selective editing** - or prioritising validation failures so that only those for which correction is expected to have a material impact on the survey outputs are followed up for corrective editing. This process also takes place as part of batch data take-on and follows any automatic editing. Data validation and editing is one of the most time consuming processes in the production of official statistics; for business surveys it is generally reckoned to account for up to 40% of survey costs. However, many validation failures result in no change; many editing changes have negligible effect on the survey estimates; and some are thought to introduce further errors.

Selective editing allows for individual businesses data that fails standard system validation checks to be assessed and scored according to its level of impact on the survey output. The possible application of selective editing to business surveys in ONS was investigated and analysed, using four months of data from one key ONS business survey, by the University of Southampton (UoS). The UoS methodological evaluation study identified two key variables that required scoring according to its level of impact on the survey output. Individual businesses' data with scores above a predetermined threshold were identified as requiring scrutiny and editing, and those below the threshold were accepted without further scrutiny. The pilot of selective editing on one ONS business survey resulted in a reduction in the number of data items requiring scrutiny of between 50 - 60% and the effects on estimates of level and change were negligible for all industries. Selective editing has now been rolled out to a number of appropriate surveys. Although the principle remained the same this did involve identifying specific key variables for the scoring of each survey's data. Studies are underway to both investigate the possible application of selective editing to the more complex surveys and

to see whether the thresholds can be raised allowing more data through without editing. This will take time as piloting has demonstrated because of survey specific functionality, the same method cannot be simply transferred to another survey without impacting on the quality of the survey estimates.

Improving data capture technology

Work also continued on improving data capture technology. The facility to use Telephone Data Entry (TDE) is offered on most short term surveys and for a small number of surveys, virtually all its data capture is via TDE. An improved TDE system has now been installed and is still being expanded. The development of an indexing system for incoming faxed forms has been introduced and also produced savings. Successful pilots have been launched to collect data over the Internet, but the roll-out of these systems has been delayed due to continuing discussions about security and registration issues.

Benefits of Stage Two

This second stage officially came to a close in March 2002 with the delivery of £0.9m of efficiency savings. These savings were achieved through a combination of centralisation, standardisation of working practices and reorganisation (often following process reviews) and the introduction of new data collection techniques such as automatic and selective editing techniques and the expansion of TDE. These real savings of £0.9m have been achieved while additionally absorbing over £1.2m of inflationary pressures during 1999-2001. All these savings have been achieved without adverse impact on the quality of the business inquiry data collection service provided to the data customers and in many cases they have been accompanied by quality improvements.

Stage two also saw the development of a Staff Development Programme (SDP), a tailor made training scheme for Data Analysts (A2s). It is a competence based scheme linked to the ONS existing competence framework where possible. It is specifically geared to the needs of the individual, job and business area, with a modular design that allows a flexible approach to training and ensures that both the needs of the individual and business are met. Accreditation by Oxford, Cambridge and Royal Society of Arts (OCR) has been gained resulting in a certificate of competence to business standards.

THE ORGANISATIONAL STRUCTURE OF THE DVU POST 2002

The first two stages of change in data collection organisation achieved the objective of creating specialised areas of expertise in data collection and in the production of results, and met the need to deliver efficiency savings in the collection and analysis of data. It also created a culture of continuous improvement in BDD, so the next natural step was to review the basic structure of the Data Validation Unit data validation branches and to plan for the move of the final group of financial, international and innovation surveys to BDD. The move to a horizontal structure had resulted in a

number of organisational changes, but little had been done to review the basic structure of the four Data Validation Branches (DVBs) that make up the DVU. These remained organised largely in the historical way in which they came into BDD into 1999. This resulted in each DVB still being organised and working in slightly different ways and some concerns about the balance of work in the four areas.

This next stage began by identifying what we wanted from a revised structure. From a customer and respondent perspective, the ideal structure would be one where a group of staff process all the surveys sent out within a particular industry sector of the economy. This would provide a single contact for the respondent and the business and would build up industry expertise. Whilst this has been achieved to some extent, the large number of surveys conducted in certain areas of the economy, in particular manufacturing, make it impractical to fully implement this approach. A balance has to be struck between building up industry expertise and creating even workloads across the DVBs, eliminating where possible, peaks and troughs. It is also important to maintain a structure that can respond quickly to changes in customer requirements. Besides needing to integrate the financial surveys into the existing structure, new surveys also needed to be incorporated. Since 1999, two substantial monthly surveys (Index of Services and Vacancy Survey) and two annual surveys (covering E-Commerce and Internet Service Providers) have been introduced into the DVU, as well as four expansions to existing surveys and a number of pilot surveys. It is apparent that the "ideal" structure would be difficult to implement and maintain, so a set of principles have been agreed which govern the current structure and will be applied to any future changes.

Ideally, the structure of the DVU should be one that:

- reflects our business objectives;
- is as optimal as possible;
- provides a consistent approach across the DVBs;
- ensures a balanced and even workload across all grades;
- is flexible to respond to our customer's changing requirements;
- is flexible to respond to DVU's changing requirements;
- provides a more efficient and effective interface with respondents to business surveys and other areas of BDD;
- allows the integration of data collection, validation and support activities for any other surveys that could transfer or be introduced into the DVU.

Revised DVU structure

By the end of 2002, the first stage of the reorganisation of the DVBs following these guidelines was completed. This first stage concentrated on evening out workloads across the DVBs and transferring work from the Runcorn office to Newport. The revised structure addressed some of the main concerns (balanced workloads, creating centres of expertise) whilst still leaving the Unit flexible to meet further changes, in particular the transfer of the financial surveys.

In December 2002 the structure of the DVU became:

Diagram 1

Data Validation Unit (DVU)

Data Validation Branch A	Data Validation Branch B	Data Validation Branch C	Data Validation Branch D
New Earning Survey <i>(operated April - August)</i>	Structural surveys (Annual Business Inquiry - financial and employment).	Price inquiries (Producer, Export, Import, Imported Capital Goods and Corporate Services Prices).	Energy and Mineral Inquiries (conducted on behalf of the DTI)
Annual Register Inquiry (ARI/1) <i>(operated September - March)</i>	Monthly employment and turnover surveys (Retail Sales Inquiry, Monthly Inquiry into Distribution and Services Sectors, Monthly Production Inquiry).	Product Sales surveys (quarterly and annual PRODCOM). <i>(operated January - July for annual survey using part-year data analysts)</i>	Enforcement and Response Chasing Teams
Monthly Wages and Salaries Survey	Quarterly Capital Expenditure Inquiry.	Annual E-commerce survey <i>(operated January - July/August)</i>	DVU Support Team
Structural surveys (Annual Business Inquiry (financial and employment))	Quarterly Inquiry into Inventories	Monthly Vacancy Rate Survey.	Compliance and Correspondence Team.
		Annual Business Spend on Capital Items Inquiry.	

Whilst not an optimum structure, it reduced peaks and troughs and improved congruency and coherence and followed the agreed principles for the ideal DVU structure. Where possible associated surveys were brigaded together, for example Price Inquiries under the same command as Prodcorn who provide the sampling frame for the Producer Price Indices, and the annual structural survey with the monthly turnover surveys. This has created mini centres of expertise and improved congruence between short-term and annual surveys and within the monthly turnover and employment surveys. In the case of the New Earnings Survey and the Annual Register Inquiry, these have been put together not because of a similar subject matter, but because the timing of the surveys means that they can be run sequentially, and consequently more efficiently.

Workloads at all grades are more evenly balanced and each DVB is organised in a slightly different way, responding to the specific needs of each of its surveys. In DVB B, staff process a number of surveys at the same time - typically, a monthly, quarterly and an annual, covering similar subject matter. This utilises the data analysts sector and respondent knowledge, but does require careful planning and organisation to ensure that one survey is not overlooked particularly in relation to meeting agreed targets. In some DVBS staff work mainly on one survey, for example Prodcorn and E-commerce, but where possible these surveys utilise part-year staff to iron out peaks and troughs. This revised structure recognised that new surveys, such as E-commerce can be demanding on the DVB Branch Head and that new work transferred in, such as that in the DVB A, need more senior management attention during the transition period. The success of the DVU structure depends largely on it being flexible to changing demands. So although staff may be responsible for a set of respondents within one or more surveys, they also assist any other areas that are experiencing short-term difficulties.

The DVU structure is continuing to evolve. In 2003, the data collection activity for the financial, international and innovation surveys transferred into the DVU. Again, the approach was to break down the process into stages, with the work being transferred initially into the DVU with the minimum of organisational change. After a "bedding in" period of six months, the organisation and placing of these new surveys into the structure of the four DVBS was reviewed by both the DVU and the customers of these surveys and it was agreed to locate half of the surveys in DVB D and the other half in DVB A. The main reason for splitting this set of surveys was to give them more senior management attention.

The ONS itself has also been undergoing a large Statistical Modernisation Programme to redesign our statistical systems and place them in a modern IT environment, supported by standard statistical tools and methods. In order to directly support the Statistical Modernisation Programme, ONS has reorganised into two central groups of activities engaged in the production of statistics: firstly the Surveys and Administrative Sources Directorate (Sources for short), and secondly two Analysis Directorates. The impact of this reorganisation has not had a great impact on BDD as it was already organised by function, but it did result in the move of the Business Register Unit from BDD to the Statistical Framework Division.

PERCEIVED BENEFITS OF THE RESTRUCTURE PROGRAMME

The main drivers for the move to a more horizontal structure, increased efficiency and development of specific expertise in data collection and in the production of results, have been achieved. However, there have been many other benefits:

- **Common processing systems and standardised working practices** - as the DVU is responsible for editing data from a number of surveys, it became essential to move towards common processing systems, standardised working practices, and common variables and definitions across surveys.
- **Enabled coherent common strategies to be taken forward:**
 - Development of a Staff Development Programme (SDP) - this is a tailor made training scheme for Data Analysts (A2s). It is a competence based scheme linked to the ONS existing competence framework where possible. It is specifically geared to the needs of the individual, job and business area, with a modular design that allows a flexible approach to training and ensures that both the needs of the individual and business are met. Accreditation by Oxford, Cambridge and Royal Society of Arts (OCR) has been gained resulting in a certificate of competence to business standards.
 - Creation of in-house training teams in each DVB. This has resulted in all new staff receiving appropriate induction training that has clearly defined learning objectives and assessment measures, ensuring a consistent approach to training across the DVU.
 - Introduction of a response chasing and enforcement strategy - next to validation, intensive follow-up action is the second most important and costly activity carried out by the DVU. All response chasing and enforcement activity is now managed via this strategy through a Project Board, ensuring that a systematic approach is used to secure consistent response from businesses.
 - Creation of a dedicated Correspondence Unit to improve the effectiveness and quality of communication with respondents. The Team deal with the more complex correspondence, monitor complaints, and provide advice and guidance to the DVU.
 - Improved congruence and coherence between surveys. As the DVU is responsible for editing data from a number of surveys, often within the same team, it has been easier to introduce congruency checks at validation of data stage and to raise the profile of the importance of coherence between inquiries.
- **Improved communication** - a central shared database was created to hold details of all contact with businesses, encouraging the sharing of data about businesses across surveys. This reduces the burden on business and facilitates congruency checking between variables across surveys.
- **Facilitates the introduction of new processes and developments** - concentrating effort on the process of collecting and editing data from results processing facilitates the introduction of new processes and developments. With the DVU's main focus on the effective and efficient collection and validation of data, it is easier to drive forward and manage new developments relating to the

input processes. This has certainly been the case for the exploration and implementation of alternative methods of data capture, such as Telephone Data Entry and the pilots of Internet collection, the development of new processing systems and the development and introduction of new methodological approaches to data validation.

- **Increased flexibility** - a key point of the current organisation is that having the vast majority of business surveys in one business area not only makes introducing change easier, but also allows the Unit to respond quickly to new demands and to reallocate resources to overcome problems. The system has provided far more control for managers. Not only are peaks removed (we used to staff up to capacity for annual surveys) by mixing short-term and annual surveys, managers can get some of their work done by another area if there are concerns about reaching targets.
- **Improved relationship with the respondent** - besides the benefits of a shared contributor database and the improvement of written communication with the respondent, concentrating effort and resources on the process of collecting and editing data has helped to identify problems experienced by respondents. When one survey is getting comments from respondents, or even one DVB, often there are not enough common themes to identify problems. Putting all 1.7 million forms and 400+ staff in one area means that significant improvements have been made as a result of a more holistic approach and reviewing data collection from a respondent point of view. A good example is - "why do I get these 10 forms ". This question was actually common but ONS never appreciated that certain size companies and SICs were hit more than others. It did in fact affect a small number (00s) of companies. Significant steps have been taken to help this small number of small companies hit by a large number of forms, largely through survey rationalisation and survey overlap control.
- **Increased opportunities for generalist staff** - the creation of the DVU and BDD means that survey processing staff can now aspire to much higher grades - there is no reason why the Director of BDD could not be someone who has risen through the survey processing administrative grades. This is a significant change from not many years ago where survey processing staff aspirations would stop middle management level (Band C).

LESSONS LEARNT

This programme of change brought about a significant change to the way the Office operated. During this process mistakes have been made, but these have been outweighed by the many successes. Here are some of the lessons learnt during the process:

- Significant change requires commitment from top down. You need to take staff with you. There was much scepticism about the change, particularly early on. A key point here was starting small and picking the best people to set up good systems and then sell it to the newcomers. It's important to pick some good leaders who believe in the new structure and send a strong message from the top of the office - some people will be very negative to the proposed change and will stay in denial for some time, even when the decision to restructure has been made.

- Have a clear strategy but be prepared to be flexible. Go for a staged approach and achieve some early success. Take it gradually and do not expect to get to the ideal in one go. Plan in 5 year stages and make assumptions about technology development for that 5 year period. Save the difficult areas (for example, financial surveys) to the end.
- Once the first stage has been successful, it is easier to move to more radical structures.
- Inform staff as early as possible that a change is being made and involve them, through project teams and focus groups, allowing ideas for change to flow up as well as down.
- Accept that people will be concerned about loss of data quality and control, in particular customers. Help to overcome this by:
 - Agreeing boundaries early on in the process. Clear boundaries between the organisational units is essential. Ensure the outputs of each are clearly specified as are the arrangements for liaison;
 - Service Level Agreements need to be drawn up thoroughly and agreed by all parties annually;
 - Standard guidelines and procedures need to be agreed and maintained.
- Procedures and guidelines need to be specified clearly through Service Level Agreements. These require data customers to set out their requirements and agree resources with the DVU. These mean 'ad hoc' checks done - previously there was a suspicion that checks were undertaken in sufficient quantity to fill the time. The DVU needs to work to fairly fixed timetables, which have been built up over time to achieve the response targets for each survey. This works well, with some adjustment if there are problems.
- Initially, staff are familiar with the work of both areas having worked on the process from start to finish. This diminishes over time and care needs to be taken that both sides understand each others work. This can be overcome by:
 - Staff movement between areas - although in practice this does not happen that often as staff tend to prefer one type of work over the other. Secondment has proved more successful;
 - Use of job-shadowing;
 - Training seminars to explain what happens to the data once it has left the DVU. Attendance at some Press Releases of key data outputs has proved extremely popular.
- Strong communication between the two areas is essential. This is maintained formally through pause and review sessions at the end of each processing cycle, annual End of Survey reports covering all aspects of performance, including the RAP awarding marks for performance against targets. And informally, by visiting each other and not relying on e-mail and processing systems.

- A good training structure is essential, particularly where multi-skilling is introduced. The Staff Development Programme and the structured DVB training teams has improved the expertise of the DVU.
- A good management information system is required to identify potential efficiency savings and to track all savings. It is very easy for savings to seep unchecked into quality.
- There was a misconception by some people that the work undertaken on the results and analysis side was more complex than the DVU and therefore required higher skilled staff. They require different skill sets, but are not less skilled. This problem still exists to some extent, with some people still thinking of the DVU as a factory environment populated by casual staff.
- Where staff work on more than one survey short term targets tend to take priority and where there are resource issues the annual work can be pushed back leading to pressure as the annual close down date approaches. This can be overcome by setting targets for the length of time an annual query can be left unanswered and by having "annual weeks" where staff concentrate almost exclusively on the annual surveys.
- Survey processing staff who work on more than one survey can switch between their different surveys easily through our workflow control procedure although they will be given a clear lead from their managers as to where the priorities lie at a point in time.

Pat Thomas
Business Data Division
16 February 2004