



National Statistics Quality Review Series

Report No. 2

---

# **Review of the Inter-Departmental Business Register**

Office for National Statistics

© Crown copyright 2001

Published with the permission of the Office for National Statistics on behalf of the Controller of Her Majesty's Stationery Office

ISBN 1 85774 438 1

Applications for reproduction:

Copyright Manager, Office for National Statistics

Room B1/09, 1 Drummond Gate,

LONDON SW1V 2QQ

fax: 020 7533 5685

e-mail: [copyright@ons.gov.uk](mailto:copyright@ons.gov.uk)

### **Contact points**

For enquiries about this publication, contact

Caroline Bustard

Tel: 020 7533 6208

e-mail: [ns.quality.reviews.team.ldn@ons.gov.uk](mailto:ns.quality.reviews.team.ldn@ons.gov.uk)

To order this publication, call National Statistics Direct on **01633 812078**. See also back cover.

For general enquiries, contact the National Statistics

Public Enquiry Service on **020 7533 5888**

(minicom: 01633 812399)

e-mail: [info@statistics.gov.uk](mailto:info@statistics.gov.uk)

fax: 01633 652747

Letters: Room DG/18, 1 Drummond Gate,

LONDON SW1V 2QQ

You can also find National Statistics on the Internet

at [www.statistics.gov.uk](http://www.statistics.gov.uk)

### **About the Office for National Statistics**

The Office for National Statistics (ONS) is the government agency responsible for compiling, analysing and disseminating many of the United Kingdom's economic, social and demographic statistics, including the retail prices index, trade figures and labour market data, as well as the periodic census of the population and health statistics. The Director of ONS is also the National Statistician and the Registrar General for England and Wales, and the agency that administers the statutory registration of births, marriages and deaths there.

### **A National Statistics Publication**

National Statistics are produced to high professional standards set out in the National Statistics Code of Practice. They undergo regular quality assurance reviews to ensure that they meet customer needs. They are produced free from any political interference.

## NATIONAL STATISTICS QUALITY REVIEWS

The White Paper *Building Trust in Statistics* set out the framework for quality assuring National Statistics. A key component of the framework is:

“a programme of thorough reviews of key outputs, at least every five years, with the involvement of methodologists and outside expertise, as appropriate.”

This programme of Quality Reviews is an important way of ensuring that National Statistics and other official statistical outputs are fit for purpose and that we are continuing to improve the quality and value of these outputs.

Further advice and information on the Review Programme is available from the Review Programme Management Team.

Tel. 020 7533 6208/6298 or e-mail [caroline.bustard@ons.gov.uk](mailto:caroline.bustard@ons.gov.uk)

Use BOOKMARKS or THUMBNAI LS from the adjacent Navigation Pane, or LINKS from the CONTENTS list or List of Annexes to navigate within this document.

Select the HAND tool and click on the title of the CONTENTS list or List of Annexes to enlarge the view.

If the hand tool displays an arrow on it, clicking advances the view down the page.

If the hand tool changes to a pointing finger, clicking on an entry in the list takes you to that item.

## CONTENTS

	<b>Page</b>
<b>Principal Findings and Recommendations</b>	<b>5</b>
<b>Chapter 1    Introductions</b>	<b>14</b>
<b>Chapter 2    User Consultation and User Requirements for Quality</b>	<b>24</b>
<b>Chapter 3    International Aspects</b>	<b>28</b>
<b>Chapter 4    Quality Measurement</b>	<b>32</b>
<b>Chapter 5    Administrative Sources</b>	<b>41</b>
<b>Chapter 6    Annual Register Inquiry</b>	<b>46</b>
<b>Chapter 7    Updating Statistical Information on the IDBR from Survey Sources</b>	<b>59</b>
<b>Chapter 8    Complex Businesses</b>	<b>67</b>
<b>Chapter 9    Coverage</b>	<b>74</b>
<b>Chapter 10   Online Processes and Validation of ONS Input Data</b>	<b>79</b>
<b>Chapter 11   Management and Communication with Users</b>	<b>81</b>
<b>Chapter 12   Implementation of Recommendations from Kokic/Brewer Study</b>	<b>86</b>
<b>Chapter 13   Concluding Remarks</b>	<b>89</b>
<b>Notes and References</b>	<b>90</b>
<b>List of Annexes</b>	<b>91</b>
<b>Glossary of Abbreviations</b>	<b>92</b>
<b>Glossary of Terms</b>	<b>94</b>
<b>Some Useful Websites</b>	<b>96</b>
<b>Annexes</b>	<b>97</b>

## **Principal Findings and Recommendations**

### **Summary**

The Inter-Departmental Business Register (IDBR) is a list of UK businesses that is maintained by ONS. It is used for selecting samples for surveys of businesses, to produce analyses of business activity and to provide lists of businesses. It is based on inputs from three administrative sources: traders registered for Value Added Tax (VAT) purposes with HM Customs and Excise (HMCE); employers operating a Pay As You Earn (PAYE) scheme, registered with the Inland Revenue (IR); and incorporated businesses registered at Companies House (CH). The ONS Annual Register Inquiry (ARI) and other surveys are used to identify and maintain the business structures necessary to produce industry and small area statistics.

The purposes of the Review were to establish what users need from the IDBR and to evaluate whether the IDBR's outputs are fit for purpose, taking into account the various attributes of quality. In addressing the second purpose, we considered how we could improve the quality of the IDBR's processes and the information about the IDBR's quality that is available to users.

This report outlines:

- the results of our consultation with users;
- the reviews of the provision of data from administrative sources, the Annual Register Inquiry, the mechanism for updating statistical information on the IDBR, complex businesses, coverage, and quality assurance and measurement.

In addition, the report covers issues relating to the management of the IDBR and communication with users, and looks at the extent to which a recent review of some aspects of register operations has been implemented.

Our overall assessment of is that the IDBR is of good quality. Improvements are needed in some areas to ensure that the IDBR better meets the needs of all users. Based on the limited range of quality measures currently available, the IDBR appears to be among the leading business registers in the world.

The main strengths of the IDBR are:

- a committed staff, many with years of experience in register operations, supported by a dedicated training unit;
- that it is updated frequently from administrative sources;
- the existence of systems that regularly update the structures of large businesses;
- its excellent coverage, which is due to the use of two comprehensive administrative sources;

- the fact that there is, for each business, a wide range of statistical information available for use in sampling and estimation;
- the unit dedicated to ensuring that the structures of the largest, most complex businesses are correct; and
- that extensive work is undertaken to reduce the level of duplication arising from the use of multiple administrative sources.

We have identified a range of areas where changes, building on the strengths identified, can help meet customer needs better. These are grouped logically within the broad areas of the IDBR that we reviewed.

The recommendations are ranked according to the priority that should be attached to them. Those with priority 1 are very important to undertake and should be addressed in the short-term. Those with priority 2 are also important, and those with priority 3 are still important, but have a lower priority than the other recommendations.

## **Quality Measurement**

Quality measurement is important for two reasons. As well as enabling users to appreciate a product's quality, measurement of quality helps people responsible for managing the output to devote resources in the most cost-effective way to improve quality where required. Measures of both the level of quality and the change over time are important.

ONS has begun to develop measures of register quality. In Chapter 4, we suggest an expansion to the range of measures available to include indicators in the following areas:

- coverage;
- accuracy of industrial classification, employment and turnover;
- consistency;
- currency of information;
- availability of contact information;
- accuracy of inputs and processes;
- level of service to customers; and
- customer satisfaction.

**Recommendation 1** *ONS should publish regularly a wide range of measures of both the level of the IDBR's quality and the change in quality over time. The existing user satisfaction questionnaire should be improved to include a wider range of questions relating to the IDBR's quality and that of the service provided by the Business Registers Unit. (Priority: 1)*

## **Administrative sources**

The IDBR is based on a comprehensive range of data from the administrative sources. This is updated frequently – weekly in the case of information on VAT traders from HMCE. We identified data items that are collected by the administrative departments that could be used by ONS to make improvements to the quality of the IDBR (section 5.3). In addition,

standards for the quality of data should be agreed by ONS with the administrative departments (section 5.10).

We endorse the continued involvement of ONS in work on the Comprehensive Business Directory (section 5.2). This work could benefit the business community by reducing the level of reporting to government, and may also benefit National Statistics by improving consistency and quality through the development of standards and wider access to data. A major benefit to the IDBR will be consistent, high quality business descriptions leading to more accurate industrial classification.

**Recommendation 2** *The development of the Comprehensive Business Directory should continue to be a priority within ONS, as a means of improving access to data from administrative sources, improving the quality of the IDBR through better matching of source data and of reducing the need to collect data directly from businesses. (Priority: 1)*

**Recommendation 3** *Building on the existing relationships with HM Customs and Excise (HMCE), Inland Revenue (IR) and Companies House (CH), ONS should:*

- *seek to extend the range of data supplied by the administrative departments to include, from HMCE, company numbers and previous VAT registration details and, from all administrative sources, comprehensive contact information (Priority: 1);*
- *seek to improve the method and frequency of transferring data from the administrative sources and seek more frequent transfer of data (Priority: 3);*
- *consider feeding back industrial classification data to the administrative departments within the scope of the prevailing legal framework (Priority: 2); and*
- *seek to improve the quality of industrial classification coming from administrative sources by using consistent data-collection and coding techniques, and by supporting those people who classify businesses. (Priority: 1)*

**Recommendation 4** *Quality standards for industrial classification, legal status, company number and address should be agreed and documented by ONS and the administrative departments. (Priority: 1)*

### **Annual Register Inquiry**

We reviewed the current practices used in the Annual Register Inquiry (ARI). The ARI contains many of the characteristics of the former Annual Employment Survey (AES), which it replaced in part. The ARI is now a tool for maintaining the IDBR, and has a different purpose from the AES. Its sole purpose is to update the structure of businesses on the IDBR, including the maintenance of auxiliary information – that is the employment, turnover and industrial classification information on the IDBR used to stratify survey samples, and to improve the precision of survey estimates. The recommendations made here reflect that change of emphasis and address the need to maximise the IDBR’s usefulness as a sampling frame, and as a tool for analysis and minimising respondent burden.

In particular, we found that the sample design and selection should be better focused on the requirements for updating the Register. Follow-up of non-response is particularly important for the ARI, and this needs to be more effective in order to ensure that up-to-date structural information is available (section 6.4.3). Previous work in ONS identified that some local units were being missed in responses to the AES questionnaire. Further development of the ARI questionnaire is needed to ensure complete coverage of local units on the IDBR (section 6.4.3.2). The electronic data-collection arrangements for a small number of the largest enterprises is a positive step, and there is potential for improved quality and reduced respondent burden by extending these arrangements – ultimately resulting in Internet-based data collection (section 6.4.3.1).

**Recommendation 5** *ONS should revise the Annual Register Inquiry sample design and selection so that it:*

- *completely enumerates enterprises with employment of 50 or more;*
- *includes a wider range of enterprises with conflicting auxiliary information;*
- *includes, every four years, enterprises with employment of 10 or more;*
- *includes enterprises with employment of less than 10 only to the extent that is necessary for adequate quality measurement; and*
- *can be used to update the IDBR in an unbiased way. (Priority 1)*

**Recommendation 6** *In the Annual Register Inquiry*

- *The response rate should be improved by following up non-response more effectively. The response rate for the largest enterprises should be 100 per cent. (Priority: 2)*
- *The general procedures used by ONS to ensure compliance with the Statistics of Trade Act 1947 should be rigorously applied to the Annual Register Inquiry. (Priority: 2)*

**Recommendation 7** *Questionnaire design for the Annual Register Inquiry should be given priority within ONS's proposed business-survey questionnaire design project. Among the particular topics that need to be addressed are omission of local units; delineation of business units; feedback of known industrial classification information; and efficient collection of business description information. (Priority: 1)*

**Recommendation 8** *ONS should continue to extend the present electronic data-collection arrangements for the Annual Register Inquiry. (Priority: 2)*

## **Updating Industrial Classification on the IDBR**

The demands of a wide range of users have resulted in a mechanism for updating industrial classification on the IDBR that, while ensuring a limited range of reporting units have a more up-to-date and, possibly, more accurate classification, is biased as a whole and leads to inconsistency. Our review found that the overall quality of industrial classifications can be improved by reducing the range of sources used to update those variables. Improvements to the quality of codes from both the ARI and the administrative sources will enable those



sources to provide high quality industrial classification information for the IDBR. Less reliance would be placed on the other sources, which would be used only where the ARI or the administrative sources do not give enough information to allow the accurate coding of individual reporting units (section 7.6).

ONS uses an electronic coding tool, called the Precision Data Coder, to provide better quality coding of business descriptions. Rules governing its use need to be strengthened to ensure its universal use and continued development. In addition, other products should be evaluated to ensure that ONS is using the best available tool (section 7.7).

**Recommendation 9** *The Annual Register Inquiry, Business Profiling Team and the administrative sources should have highest priority for updating IDBR auxiliary information. In the case of industrial classification, this depends on improvements to the current level of quality of information from those sources. Other sources should be used by the Business Registers Unit to resolve uncertainty and ambiguity. (Priority: 1)*

**Recommendation 10** *The agreed electronic tool should be used to aid classifying business descriptions. All staff members should feed back information on deficiencies to those controlling the system to improve the quality of the coding tool. The use of expertise from outside ONS (for example, industry groups) should be considered as a way of improving coding quality. ONS should review the Precision Data Coder alongside other similar products to ensure that the most effective tool for the job is used. (Priority: 2)*

**Recommendation 11** *The rules for updating the industrial classification and employment variables, set up to maintain stability of sample membership for short-period inquiries, should be reviewed to see if they could better accommodate register maintenance requirements and should be applied rigorously. (Priority: 3)*

## **Complex Businesses**

The most complex enterprises contribute to the economy to an extent that is disproportionate to the number of such businesses. Therefore, it is important to ensure that the structures of these businesses and associated auxiliary information are up to date and correct. ONS has identified this priority and established a unit that is responsible for maintaining the structure (business profiling) of such businesses. In order that the profiling work is as effective as possible, the range of enterprises included and a programme of both desk work and visits need to be agreed (section 8.2).

Communication should be improved between the Business Profiling Team and those responsible for running surveys to ensure that updates to business structures are reported in advance. Opportunities to co-ordinate the collection of data relating to business structures should be pursued with other government departments, within the requirements for maintaining data confidentiality (section 8.2.5).

**Recommendation 12** *A business profiling strategy should be agreed based on two objective measures: the level of secondary activity; and the impact on survey estimates and analyses. A solution should be developed to bring profiling and ARI operations together for the largest businesses. A trigger mechanism should be established for sub-annual profiling of the most complex enterprises. (Priority: 1)*

**Recommendation 13** *The Business Profiling Team (BPT) should initially establish the correct structures of the businesses at the desk. For complex businesses, an agreed programme of profiling visits should be put in place, and this should be made available to survey managers in advance. The BPT should ensure that data collection and survey results managers are informed, in advance, of its updates to business structures on the IDBR. (Priority: 3)*

**Recommendation 14** *To reduce the burden on business and improve coherence across National Statistics, ONS should explore with other government departments opportunities, within the requirements for maintaining data confidentiality, to co-ordinate the collection of data relating to the structure of complex enterprises. (Priority: 3)*

For some surveys, non-standard arrangements have been set up to facilitate the reporting of business data (section 8.3). The maintenance of reporting structures relevant to only a single survey places undue pressure on the resources of the Business Registers Unit (BRU). It may also result in lower quality outputs through the possible omission and duplication of economic activity. We recommend improvements in this area.

**Recommendation 15** *A detailed review of the requirements for survey-specific reporting arrangements should be carried out with the aim of standardising reporting arrangements for all surveys. Where non-standard arrangements are necessary, the procedures for managing them should be simplified. (Priority: 2)*

## Coverage

Through the use of two main administrative sources, the IDBR has coverage estimated at around 99 per cent of economic activity. There is a well-used method of measuring the extent of the undercoverage (section 9.2.1), although work needs to be extended to improve the precision and industrial detail of the estimates.

**Recommendation 16** *ONS should work with the Department of Trade and Industry to improve the method of estimating the undercoverage that arises from businesses not required to register with HM Customs and Excise and Inland Revenue. (Priority: 3)*

Duplication (arising through the dual sourcing of the IDBR) may cause problems. Much work is undertaken to link records from the two main administrative sources (section 9.3). This has resulted in the linking of many records, and improvements to matching software. At present,

business records that are thought to be duplicates are removed from survey populations (section 9.3.2). The bias introduced by this treatment should be measured, and procedures changed as appropriate.

**Recommendation 17** *ONS should measure any bias that may arise because of the treatment of potential duplicate businesses, and revise the treatment of reporting units in surveys as appropriate. (Priority: 2)*

### **Online Processes and Validating ONS Input Data**

Overall, the level of quality assurance of clerical processes is high. Supervisors check much of their staff's work, particularly that of new members. In addition, an independent audit is carried out every two months to establish whether procedures are correct and have been followed properly (section 10.2). This audit also results in changes to procedural guidelines. The audit is generally good and we make observations in this report that will ensure the audit process is as effective as possible.

Data from administrative sources are routinely vetted to ensure that erroneous data are not added to the IDBR. However, data from ONS sources are not checked by those responsible for updating the IDBR, and may not be appropriate for updating the Register without further validation (section 10.3).

**Recommendation 18** *The audit of online amendments should be more timely and frequent. The sample size of the audit should be reviewed and the audit should be focused on those amendments that can have material impact on the quality of the IDBR. Recommendations should be developed and agreed in partnership between the audit team and work areas. (Priority: 2)*

**Recommendation 19** *Business Registers Unit should extend the range of validation checks on inputs to ensure that statistical data present on the IDBR are of adequate quality. (Priority: 3)*

### **Management and Communication**

Overall, provision has been made to ensure that members of staff receive the training appropriate to the job (section 11.2). The Business Registers Unit (BRU) has created its own training unit that supports this. Members of staff should be encouraged to work in a variety of areas of business statistics to ensure a breadth, as well as a depth, of expertise. This could be considered as part of the development of a recruitment strategy for Register staff. Managers are supported by a wide and increasing range of management information (section 11.4). This can be improved by its being more focused on the impact of quality. A monthly report is produced, which presents analyses of the distribution of businesses and identifies the effect of recent updates on the IDBR. Further development of this will increase its usefulness to recipients.

Documentation is good for those responsible for maintaining the IDBR, but needs to be made more comprehensive for Register users (section 11.3). We also identified that, although training is provided for regular users, occasional users weren't usually well-versed in the principles of maintaining a business register.

Communication needs to be improved between BRU and Northern Ireland's Department of Enterprise, Trade and Investment, which is responsible for maintaining NI businesses on the IDBR. In addition, the potential for overcoming the present legal restrictions in place between the two departments should be explored.

Three committees share responsibility for some aspects of the IDBR's management. Notably, no single ONS or National Statistics group has taken responsibility for ensuring that recommendations arising from past methodological studies were evaluated and implemented (section 11.6). There is a risk that lack of clarity regarding these groups' responsibilities could adversely affect the quality of the IDBR. In addition to ensuring the implementation of recommendations, an impact assessment is needed for proposed changes to the IDBR.

**Recommendation 20** *Business Registers Unit (BRU) should develop a specific recruitment policy and training programme that reflects its specialist function. This policy should include arrangements to make sure that members of staff are encouraged to move within BRU and across Prices and Business Group to ensure a breadth, as well as a depth, of expertise. (Priority: 2)*

**Recommendation 21** *The Business Registers Unit should improve electronic communication with the Department of Enterprise, Trade and Investment Northern Ireland to ensure consistent practice is adopted in the two departments. This should include consideration of how present legal constraints can be overcome in future. (Priority: 2)*

**Recommendation 22** *ONS should strengthen staff training with regard to the principles of maintaining and using a business register. The use of the Data Analyst and Statistical Analyst schemes for this purpose is endorsed. (Priority: 3)*

**Recommendation 23** *Additional documentation should be produced for users of the Register. Part of this documentation should take the form of a frequently asked questions document. (Priority: 1)*

**Recommendation 24** *The existing quarterly and monthly management reports should be focused on measuring the impact on quality of changes to the Register. Business Registers Unit should continue to develop its monthly report of changes to the size and structure of the Register in consultation with users. (Priority: 2)*

**Recommendation 25** *An improved online interface should be developed to make it easier for authorised users to access and update the Register. (Priority: 2)*

**Recommendation 26** *ONS should ensure that the impact on outputs is assessed for any proposed changes to the IDBR. (Priority: 1)*

**Recommendation 27** *ONS's Register strategy group should take responsibility for overseeing the implementation of recommendations from projects relating to the improvement of the IDBR. The membership of the group should be reviewed and include a non-ONS member. (Priority: 1)*

### **Review of the Implementation of Recommendations from a Previous Study**

In 1996, Phil Kokic and Ken Brewer, then of the University of Southampton, were commissioned to review certain aspects of IDBR's operation (Chapter 12). As part of this review we evaluated the extent to which recommendations of that study had been implemented.

The Kokic/Brewer report identifies a range of topics relating to sample design that ONS is proposing to take forward in a separate project. In addition, there are two recommendations relating directly to the IDBR that should be addressed along with the rest of the recommendations in this review.

**Recommendation 28** *ONS should change sample selection procedures to ensure stability of sample membership for businesses newly included in samples. (Priority: 2)*

**Recommendation 29** *ONS should set aside resources to study further the impact of births and deaths of enterprises on statistical outputs. (Priority: 1)*

### **Conclusion**

The report contains 29 recommendations that will develop the IDBR as a key element of the ONS infrastructure, delivering a service that responds to increasing user demands, while minimising the burden on business. This will be achieved by working more closely with the administrative departments, improved use of technology and taking account of the best practices of other statistical offices. By implementing these changes, the IDBR will continue to deliver a high quality service to its users.

## **Chapter 1 Introductions**

### **1.1 This Review**

A series of reviews has been established as part of the Quality Strategy for National Statistics.<sup>1</sup> The Review of the Inter-Departmental Business Register (IDBR) is the second such review, and the first in the Commerce, Energy and Industry theme group. This review was undertaken by a project team comprising:

Mark Pont, from ONS's Methodology Group (MG), as project leader;  
John Perry, head of ONS's Business Registers Unit (BRU); and  
Ruth Studley from MG.

Gareth Jones from ONS's Prices and Business Group (PBG) was part of the team representing customer interests. Professor Ray Chambers from the University of Southampton provided some quality assurance.

The project was managed under PRINCE guidelines. A Project Board oversaw the work, and reported to a steering group, which included members from other government departments (OGDs) and from a trade association.

### **1.2 Scope of the Review**

The scope of the Review was to examine the main IDBR processes (those relating to updating the IDBR from external and internal sources, selecting survey samples, and quality control) to establish the effect they have on the Register's quality. The Review also aimed to define quality measures that can be used to monitor the IDBR's quality. It was not the purpose of the Review to examine whether customers make appropriate use of the IDBR.

Specifically excluded from the Review were:

- development of alternative geographic referencing systems that are being taken forward by the ONS Geographic Referencing Strategy; and
- industrial coverage, particularly relating to agriculture, where there are long-standing plans to merge information from the agricultural census conducted by the Ministry of Agriculture, Fisheries and Food (MAFF) with the IDBR.

A copy of the agreed scope of the Review is attached at Annex 1.

### **1.3 This Report**

This report outlines our findings, and makes recommendations for ONS to consider. Some recommendations need to be addressed in conjunction with other government departments, particularly when considering the administrative data sources. The recommendations are ranked according to the priority that should be attached to them. Those with priority 1 are very important to undertake and should be addressed in the short-term. Those with priority 2

are also important, and those with priority 3 are still important, but have a lower priority than the other recommendations.

After introducing the IDBR and some of its concepts, we report on the information that can be used to compare the IDBR's quality with that of other countries' business registers. We recommend a range of indicators that should be produced regularly to monitor the quality of the IDBR. Following an account of the results of our consultation with users, we assess the quality of various register processes and outputs, and make recommendations for improvements (it wasn't within the scope of the Review to consider the use of additional data sources). The report is structured to correspond with the broad areas of the Review.

## **1.4 The IDBR**

The IDBR<sup>2</sup> was developed in the mid-1990s to replace two separate registers: the VAT-based register of the former Central Statistical Office and the PAYE-based register of the Employment Department. The IDBR plays an important strategic role in official statistics, and a good quality Register is essential for underpinning the production of National Accounts.

The main drivers for developing the IDBR were the need to improve coverage and to improve the coherence of estimates produced from surveys that had previously used the two registers as separate frames. In particular, the derivation of productivity and unit wage costs would be possible using statistics generated on a more consistent basis.

One perceived benefit was that the Register would be more complete if it used two distinct administrative data sources. The IDBR is based on data from two main sources: HM Customs and Excise (HMCE), which registers traders for Value Added Tax (VAT) purposes; and Inland Revenue (IR), which registers employers who operate Pay As You Earn (PAYE) income tax schemes. In addition, the IDBR is updated from register quality surveys, statistical surveys and other sources.

Information from HMCE is transferred to ONS under the VAT Act 1994, and information from IR, under the Finance Act 1969. Both Acts, and the Statistics of Trade Act 1947 (SoTA), which covers data collected directly by ONS, restrict the release of data outside ONS. The table at Annex 2 summarises the release arrangements under the different Acts.

The IDBR covers all businesses registered for VAT or PAYE. Businesses in the agricultural sector are recorded on the IDBR, although the Ministry of Agriculture, Fisheries and Food (MAFF) is responsible for collecting agricultural statistics. Construction businesses are also on the IDBR, and both ONS and the Department of the Environment, Transport and the Regions (DETR) publish data relating to the construction sector.

Third, the IDBR covers all UK businesses. However, separate legislation covers the British and Northern Irish parts of the IDBR. ONS is responsible for maintaining the British part of the Register, while the Department of Enterprise, Trade and Investment Northern Ireland (DETINI) deals with the Northern Irish part. Consequently, some procedures operated by DETINI are different from those in place at ONS. Moreover, while ONS runs an Annual Register Inquiry, DETINI runs a two-yearly Census of Employment.

This report and its recommendations are written for ONS. However, the principles underpinning the recommendations apply equally to both the GB and NI parts of the Register. ONS should consult with DETINI throughout the process of implementing the recommendations. Where recommendations have been made that relate to GB-only processes, such as the ARI, ONS should consider with DETINI the applicability of the recommendations for DETINI systems. Consistency is important, particularly for direct production of UK statistics and for the production of statistics for regional comparison. Therefore, the recommendations from this review must be implemented in such a way that ensures consistency between the NI and GB parts of the IDBR.

### **1.4.1 IDBR Units**

IDBR units are classified into three types: administrative units, statistical units and reporting units.

#### *Administrative Units*

The administrative units are the VAT traders and PAYE employers.

#### *Statistical Units*

The statistical units are the enterprise, enterprise group, local unit and kind of activity unit (KAU), all of which are defined precisely in the EU Regulation on Statistical Units<sup>3</sup> as follows.

#### *Enterprise*

‘The enterprise is the smallest combination of legal units that is an organizational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources. An enterprise carries out one or more activities at one or more locations. An enterprise may be a sole legal unit.’

#### *Enterprise Group*

‘An enterprise group is an association of enterprises bound together by legal and/or financial links. A group of enterprises can have more than one decision-making centre, especially for policy on production, sales and profits. It may centralize certain aspects of financial management and taxation. It constitutes an economic entity which is empowered to make choices, particularly concerning the units which it comprises.’

#### *Local Unit*

‘The local unit is an enterprise or part thereof (eg a workshop, factory, warehouse, office, mine or depot) situated in a geographically identified place. At or from this place economic



activity is carried out for which – save for certain exceptions – one or more persons work (even if only part-time) for one and the same enterprise.’

### *Kind of Activity Unit*

‘The kind of activity unit (KAU) groups all the parts of an enterprise contributing to the performance of an activity at class level (four digits) of NACE Rev 1 and corresponds to one or more operational subdivisions of the enterprise. The enterprise’s information system must be capable of indicating or calculating for each KAU at least the value of production, intermediate consumption, manpower costs, the operating surplus and employment and gross fixed capital formation.’

### *Reporting Unit*

The reporting unit 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 that correspond to KAUs, and are defined by specifying the appropriate local units from within an enterprise.

## **1.4.2 Administrative Sources**

### *VAT Traders*

VAT registration is compulsory for traders with an annual turnover above the ‘VAT threshold’ (£52,000 in 2000/01) except in a small number of exempt industries, such as health and education. Businesses register with HMCE, which in turn provides data to ONS under the VAT Act 1994. HMCE supplies important variables to the IDBR, specifically, name, address, postcode, industrial classification (SIC(92)), turnover and legal status. Data are provided from HMCE to ONS either weekly, monthly or quarterly.

The weekly data file includes details of all new registrations, VAT traders that have closed their account with HMCE and changes to existing VAT traders. Changes to industrial classification, turnover, legal status, address and postcode for existing traders are also notified.

The monthly file includes members of VAT groups. In some cases, HMCE allows a group of businesses under common ownership to register together for VAT payment purposes. One member of the group is nominated to liaise with HMCE, paying tax due. This member is called the representative member. The other members of the group are called non-representative members. The monthly update relates to the non-representatives (representatives are included in the weekly data).

Every quarter, ONS is provided with details of the most recent year’s turnover for all VAT-registered businesses except non-representatives.

### *PAYE Employers*

PAYE registration with IR is compulsory for employers operating PAYE schemes. Every quarter, IR supplies ONS with data under the Finance Act 1969, which relates to all PAYE schemes. Changes, including births and deaths of employer schemes, are detected by comparing the files provided by IR with the information currently on the IDBR. IR supplies a similar range of data to that provided by HMCE, the main difference being that PAYE gives details of the number of employees within each PAYE scheme, and not the turnover. The industrial classification used by IR is the Trade Classification Number (TCN), which is not aligned to SIC(92).

### *Companies House*

Every quarter, ONS uses Companies House (CH) information to help link corporate VAT and PAYE records. CH provides information on name, industrial classification, activity status and company number. The classification is aligned to SIC(92) at the class level. ONS matches the name against existing corporate records where the company number is not known by ONS. In addition, ONS has online access through the Companies House Direct service, which is used to provide current information to resolve queries.

### *Dun and Bradstreet*

Dun and Bradstreet (D&B) provides data annually to ONS, so that information on UK companies, their foreign subsidiaries and parents, can be updated at the enterprise group level on IDBR. This is from D&B's *Worldbase* system. In addition, the ONS uses quarterly updates on CD-ROM and, occasionally, the D&B online service.

## **1.4.3 Register Maintenance Sources**

### *Annual Register Inquiry*

Apart from the administrative sources, the Annual Register Inquiry (ARI) is the main vehicle for updating the IDBR. The ARI is designed, together with the Annual Business Inquiry (ABI), to meet user requirements for the provision of statistics for small geographical areas.

The objectives of the ARI are to:

- provide employment data at the local unit level for ABI to calculate disaggregated employment estimates;
- maintain the quality of the IDBR;
- meet the EU Regulation that requires local units to be updated every four years; and
- assist IDBR's proving work.

The survey comprises two parts: ARI/1, which is an annual check of existing enterprises; and ARI/2, the confirmation of details about the larger new administrative units, which is carried out every quarter.

The purpose of ARI/1 is to maintain information on business structures, primarily for those enterprises that operate at more than one location.

ARI/2 is designed to validate the information that has been sent to ONS from the administrative sources, and to establish the local unit structure for new enterprises. This process is known as ‘proving’. Proven enterprises on the IDBR are identified by the existence of explicitly recorded local units on the Register. Unproven enterprises are assumed to have just a single site.

Both ARI/1 and ARI/2 are for GB businesses only. DETINI carries out continuous proving for Northern Ireland enterprises. DETINI carries out a census of employment every two years, covering all employees (apart from in some parts of the agricultural sector) in Northern Ireland.

### *Complex Businesses*

Some complex businesses are maintained by a team that agrees the most appropriate structures with the businesses concerned.

## **1.4.4 Other Sources**

### *Statistical Surveys*

Statistical surveys conducted by ONS and other government departments can update the IDBR in certain situations. In some circumstances, updating is carried out immediately: the death of a business, name and address changes, and business restructuring. Updating of industrial classification is annual in batch, but can be more frequent for individual units. In the case of industrial classification, those sources thought to provide the best quality data are given higher priority.

### *Management of the IDBR*

The Business Registers Unit (BRU), which is part of ONS’s Prices and Business Group (PBG), is responsible for managing the IDBR. A staff of around 130 is responsible for maintaining the Register, including the ARI’s operation, and provision of analysis for users. In addition, a dedicated IS programmer resource is available for developing and maintaining the online and batch-processing systems.

Three groups with some responsibility for IDBR policy have been established: an inter-departmental management group, an ONS group addressing quality issues, and a working-level ONS group.

## **1.4.5 Uses of the IDBR**

The IDBR is used mainly as a statistical tool, but it also has some limited administrative uses.

### *Statistical Uses*

At present, IDBR has two statistical uses: as a frame for selecting samples and producing survey estimates, and as a direct analysis tool.

### *Conducting Surveys*

The IDBR is used as a sampling frame and mailing list for official business surveys. Its specific roles are:

- to select businesses to be included in the survey samples;
- as a way of contacting businesses, primarily by mailing questionnaires, and to provide a receipting and reminder service;
- to provide information to specify domains of analysis;
- to provide auxiliary information to enable estimates to be made for businesses that don't respond to surveys (imputation), or weren't asked to participate in a particular survey (estimation); and
- to facilitate the monitoring of burden on business and control of overlap between survey samples.

At present, most surveys run from the IDBR are undertaken for ONS or for DETINI. Surveys are also carried out on behalf of OGDs. Most surveys are repeated regularly and are run to satisfy the demands of economic statistics. One-off surveys are also conducted, primarily for policy-making. Data are usually collected using mail-back paper questionnaires that are sent out to selected businesses.

### *Direct Analysis*

In addition to providing samples and population information for surveys, the IDBR provides an invaluable source of direct analysis. Direct analyses have been carried out from the UK's central business register since the early 1980s. These are annual snapshots, regular business demographic analyses, or ad hoc analyses. The IDBR is invaluable because it has both the detail and the immediacy that are lacking in survey sources. On the other hand, the IDBR does not have the same wide range of variables as surveys, and it can be less consistent over time than survey sources.

ONS has for many years published a regular annual snapshot of the distribution of businesses in its *Size Analysis of Businesses* volume.<sup>4</sup> In addition, IDBR data are also used in other ONS publications such as the *Focus On* series, *Regional Trends* and *Sector Reviews*. Data are increasingly being released on the National Statistics website.

IDBR data on enterprise demography,<sup>5</sup> and business start-ups and closures<sup>6</sup> are also published annually by the Small Business Service of the Department of Trade and Industry (DTI). There is a particular interest in this type of analysis within Europe.

Increasingly, analyses are being requested by a wide range of users from within ONS, OGDs, the research community, the wider public and the Statistical Office of the European Community (Eurostat). Often, such analyses are used to support the evaluation of changes to

government policy. In addition to analysis, individual data are made available for analysis by OGDs.

### *Administrative Uses*

The IDBR also has some administrative purposes: to allow DTI, National Assembly for Wales (NAW) and Scottish Executive (SE) to provide a service to industry; for HMCE to update Standard Industrial Classification (rev 1992) (SIC(92))<sup>7</sup> codes; and to assist local authorities in planning.

### *Future Use*

Two new developments require a high quality IDBR. These developments will drive the way the IDBR is set up and maintained in future.

The first is as a move towards the better use of administrative data, which the IDBR can facilitate. There are potentially significant benefits, particularly in terms of cost savings and the reduction of burden on business, of using administrative data better. Administrative data also provide a more complete coverage of the economy than surveys. Better use of data from the administrative sources gives opportunities to eliminate, or substantially reduce, the need to run statistical surveys. Instead, resources can be used to ensure that the administrative data can be used as the basis of statistics published by ONS.

The second development is the increased requirement for small area statistics. This requirement comes from both central government and local authorities. Good quality small area statistics rely on detailed information about the local unit structure of businesses. This information is not available from any administrative source, and can only be obtained from the IDBR.

## **1.4.6 User Requirements for the IDBR**

In this section we outline the characteristics that the IDBR will need in order to meet the range of user requirements identified in section 1.4.5. We distinguish between the two present uses – conducting surveys and direct analysis – and the future developments that will allow better use of administrative data and increase the extent of small area statistics.

### *For conducting surveys*

- An effective and timely sample selection mechanism.
- Complete, accurate and up-to-date contact information.
- Complete, accurate, up-to-date and consistent auxiliary information.
- Up-to-date information relating to business structures.
- Complete coverage.

*For direct analysis*

- Complete, accurate, up-to-date and consistent auxiliary information.
- Up-to-date information relating to business structures.
- Complete coverage.

*For enabling better use of administrative data*

- Timely updating of information from administrative sources.
- Ability to link data from different sources.

*For improving small area statistics*

- Complete, accurate and up-to-date information about local units.

In Chapter 2, we describe some of the main comments made by users relating to the quality of these aspects of the IDBR. Chapter 3 compares these aspects of the IDBR with other business registers, and in Chapter 4 we propose a range of indicators that will measure the extent to which the IDBR meets users' needs for quality.

## 1.5 Summary Statistics for the IDBR

In this section we report the IDBR's size and composition. Tables 1 to 3 provide counts of reporting units within the UK and within the standard survey populations.

**Table 1** IDBR Reporting Units by Industrial Sector

<b>Industrial Sector</b>	<b>Number of Units</b>
Agriculture	153,313
Production	169,355
Construction	190,283
Distribution	521,210
Transport	79,849
Finance	27,513
Govt & other services	774,550
<b>Total</b>	<b>1,916,073</b>

*Source: IDBR, October 2000*

**Table 2 IDBR Reporting Units by Structure**

<b>Unit Type</b>	<b>Number of Units</b>
Proven	655,975
<i>of which: one local unit</i>	588,505
<i>two local units</i>	40,860
<i>more than two local units</i>	26,610
Unproven	1,260,098
<i>of which: VAT and PAYE</i>	328,720
<i>PAYE only</i>	150,398
<i>VAT only</i>	780,980
Total	1,916,073

*Source: IDBR, October 2000*

**Table 3 IDBR Reporting Units by Legal Status**

<b>Legal Status</b>	<b>Number of Units</b>
Company	822,811
Sole proprietor	647,328
Partnership	390,668
Public corporation	524
Central Government	499
Local authority	2,526
Non-profit-making body or association	51,717
Total	1,916,073

*Source: IDBR, October 2000*

## **Chapter 2**

### **User Consultation and User Requirements for Quality**

#### **2.1 Introduction**

In this chapter we examine users' perceptions and needs regarding the IDBR's quality. In addition, we report the quality improvements suggested by users to ensure the continued and increased use of the IDBR.

#### **2.2 Description of the Consultation**

A selection of users and potential users of the IDBR was consulted. Users were selected from within ONS, OGDs and the wider business, research and public community. We identified potential customers as those who had requested samples or analysis from the IDBR, but, for one reason or another, could not have their request met. Thirty-nine responses were received. Follow-up conversations were held where more information was required. Copies of the questionnaires and a list of respondents are included at Annex 3.

In this chapter we report the comments that were made most frequently about the IDBR. In some cases, users raised specific requirements that hadn't previously been raised with BRU, and they are included here to provide a checklist for both users and BRU to follow up.

#### **2.3 General**

The main general comments made by respondents were that:

- on the whole, IDBR staff provided a great deal of good quality information, and were quick to respond to requests for analysis and other queries;
- some respondents would have liked faster responses to their requests;
- many users wanted more comprehensive documentation, containing less technical jargon, on the methods and practices of the IDBR; and
- the range of information available on the IDBR, particularly that relating to business structures, employment and turnover, was useful, although not always accurate. Respondents also thought that the facility to link with other sources was useful.

We refer to employment, turnover and industrial classification information on the IDBR as auxiliary information, or auxiliary variables. These data are available for stratifying survey samples, and for improving the precision of survey estimates.

#### **2.4 Coverage**

Undercoverage restricted some respondents' use of the IDBR. Particular areas mentioned were the agriculture, forestry and fishing industries. Work is going on separately to link MAFF's farm register with the IDBR. Users who survey specific domains often have more specialised knowledge about those particular industries and felt that the IDBR's coverage was inaccurate. Such users commented that there is no formal method for updating the IDBR with



their more specialised data. Undercoverage of sole traders and small businesses is a recognised problem.

Issues relating to IDBR's coverage are addressed in Chapter 9.

## **2.5 Business Structures**

The ability to view the coverage and hierarchy of businesses is a useful facility. Users benefited from the availability of local unit data, and from reporting units that were set up below the level of the enterprise. This is particularly important for NI where UK-wide enterprises are split into separate reporting units for GB and NI. There was some concern that employment data for local units are sometimes inaccurate and unreliable. Incorrect employment data cause problems, particularly for people responsible for collecting and validating survey data. In many cases, these problems have arisen where the local unit employment data were out of date.

The maintenance of local unit information from the ARI is discussed in detail in Chapter 6.

## **2.6 Industrial Classification**

Non-ONS users were generally impressed with the quality of industrial classification. Some ONS survey managers appreciated the fact that industrial classification codes from surveys could be used to correct existing codes. Others were concerned about the bias this could introduce, particularly with rotating samples. Also, different surveys have different methods of classifying businesses and this may cause inconsistency. ONS users suggested that the ARI should have priority over other surveys in determining the broad sector of a unit, but that other sources could be used to determine the specific SIC(92) within the broad sector already determined by ARI.

Some users felt that while the SIC(92) codes aren't perfect they are improving. However, there is concern that the amount of incorrect industrial classification on the IDBR limits the Register's use.

Industrial classification issues are addressed in the chapters relating to the administrative sources (Chapter 5), the ARI (Chapter 6) and Chapter 7, which deals with updating industrial classification.

## **2.7 Speed and Frequency of Updates of the IDBR**

The speed and frequency with which the IDBR is updated makes it particularly useful. However, it was felt that more validation of the automatic updates would be helpful, particularly responses from the ARI. Local unit employment information is often out of date. ONS users were concerned about the timing of updates to the structures of complex businesses.

Up-to-date registration and deregistration information is helpful for users, although some regarded the lag between mergers and de-mergers in the real world and the relevant structural changes appearing on the IDBR as being too slow.

A constraint on the use of the IDBR is the frequency of reclassification. Opinion varied: some felt that updates were too frequent, and some too infrequent. Users within ONS were happy with the timing of classification updates.

Survey managers would like to monitor the movement of local units between enterprises to ensure that when an enterprise restructures, the scope of the surveys remains constant. This is particularly relevant for surveys that depend on local units for small area estimation.

A frequent criticism was that the death date of a reporting unit on the IDBR reflects the time the data were updated and not the actual date the business ceased trading. This treatment is correct, as the date stored on the IDBR refers to the date the reporting unit became unavailable for selection in samples, not the date the business ceased trading. Information about the actual date of death is held for every enterprise.

## **2.8 Contact Details**

Storing telephone numbers for all businesses would allow users to conduct telephone surveys. Telephone interviewing is carried out as part of many research projects and surveys drawn from the IDBR. While phone numbers aren't widely available on the IDBR, the Register won't meet researchers' needs. In addition, local unit contact details need to be correct because sampling at this level is increasing.

Geographical information is required for both regional and small area statistics, and therefore needs to be up to date and accurate. The inclusion of the Geographic Referencing System code for all units on the IDBR would be particularly helpful to certain users, as this gives a more precise geographical code, and improves regional data. BRU is already committed to the ONS Geographic Referencing Strategy.

Recommendation 3 in Chapter 5 addresses the issue of the provision of contact information.

## **2.9 Communication and Access**

Legal restrictions on using business data hinder analysis by some users. Many users do not have access to disaggregate data and data relating to individual IDBR units. Certain research organisations argued that if they had access to the IDBR, they could check its quality. Legal restrictions and disclosure rules prevent OGDs from exploiting the IDBR fully. Users from OGDs would prefer that ONS rely more heavily on data-exchange contracts to ensure confidentiality is maintained. Users within ONS want freer access to the data, including the opportunity to obtain data and produce analyses themselves.

The infrastructure of the IDBR was the subject of many comments. These covered, in particular:

- the search facility, which didn't satisfy all users' requirements;

- the provision of a simple navigation system with an easy-to-use interface;
- the fact that it's very helpful for survey sections to see which businesses have been included in samples; and
- the facility that allows users to analyse businesses over time.

Issues relating to communication are discussed in Chapter 11.

## **2.10 Sampling**

Users made a range of observations relating to how the IDBR selects samples. All the points raised are covered in Chapter 12 of this report, which assesses the implementation of recommendations made by a previous review of some aspects of the IDBR's operation. These points will be addressed in the sampling and stratification review proposed by ONS.

## **2.11 Other Suggested Developments**

The consultation exercise produced a range of suggested improvements to the IDBR. Many of these are addressed under the main themes of the Review. In addition, we identify here a range of other issues that ONS should consider pursuing with the relevant users. These are:

- having a menu-driven system for establishing users' needs and a web-based data-ordering system, rather than one that requires customers to request data from IDBR staff;
- constructing a reference table for IDBR reporting units to hold the history of relationships with local units;
- providing OGDs with an easier method of logging-on to the system;
- a quicker update of annual employment;
- sending to BRU for further investigation details of the smaller businesses that repeatedly don't respond to ONS surveys;
- producing extra variables, such as business-to-business turnover and business-to-customer turnover;
- working with Dun and Bradstreet to improve the quality of its data; and
- allowing public access to some register information, for example name, address and SIC(92) code.

## **Chapter 3**

### **International Aspects**

#### **3.1 Introduction**

In this chapter we consider international best practice, and examine the work to measure and improve quality that has been carried out within Europe and the rest of the world, although we note that quality measures are currently rather few and far between. This scarcity has been identified by Eurostat, which is proposing to expand the range of information it collects about quality from 2001 onwards. Eurostat's keenness on quality is reflected in the resources it has given ONS for developing quality measures.

#### **3.2 International Comparisons**

##### **3.2.1 Europe**

The UK complies with the requirements of the 1993 EU Regulation on Harmonisation of Business Registers for Statistical Purposes.<sup>8</sup> In summary, the Regulation requires:

- the existence of enterprises, legal units and local units and links between them, with the exception of very small enterprises (equivalent to employment of less than 0.5 of a person) and enterprises classified to agriculture or public administration;
- the existence of specific characteristics of business units, mainly name, address, industrial classification and size information;
- annual updating of entries to and removals from the register; and
- an annual extract of the register.

This Regulation applies to all member states, of which only Germany has derogation. A published manual of recommendations, which includes a chapter on business register quality, supports the Regulation.

The Regulation is managed by a committee that meets at least once a year. Every year since 1995, member states have completed a questionnaire to determine their convergence to the Regulation's requirements. In 1999, Eurostat produced through the Statistical Programme Committee (SPC) a report to the Council demonstrating progress and recommending a future work programme. The report included a section on business register quality, which is included at Annex 4. The annual questionnaire doesn't yet collect information about quality.

The UK meets the requirements of the Regulation fully. There has been some discussion of coverage, where the UK has requested an opinion on the sufficiency of VAT and PAYE as the basis for the IDBR. The response from Eurostat was that UK coverage complied with the Regulation. The table below shows how the UK compares with other member states (including Norway) on meeting the Regulation's requirements. The material contained in the following table is a summary of information published in the SPC report mentioned above.

**Table 4 Compliance with Main Aspects of the European Regulations**

	Units		Coverage				All compulsory variables held for	
	legal unit differs from enterprise?	local unit held?	Agriculture (optional)	public admin (optional)	finance	education	enterprise	local unit
Austria				N			N	N
Belgium		N						N
Denmark	N							
Finland	N						N	N
France							N	N
Germany	N		N		N	N	N	N
Greece	N		N	N			N	N
Ireland	N		N	N				
Italy	N		N	N			N	N
Luxembourg			N	N			N	N
Netherlands							N	N
Norway	N						N	N
Portugal	N						N	N
Spain			N	N				
Sweden							N	N
UK								

*Source: Statistical Program Committee Report on Register Quality*

The table shows that the UK compares favourably with other European countries. The UK is the only country that meets the Regulation fully.

It is important to note, however, that while the UK meets the EU Regulation, the Regulation itself addresses only a narrow range of quality issues, namely frequency of updating and coverage. The Regulation does not require any particular standards for other aspects of quality, such as accuracy or consistency. Therefore, compliance with the Regulation is only one indicator of a register's quality. Measures of other aspects of quality are hard to establish. However, Eurostat has realised the importance of measuring quality, and will pilot a quality questionnaire from 2001, and include quality measurement as a substantive item on the agenda for the Business Registers Committee in 2001.

Quality measures are specified for one of the main outputs of the Business Register: the Structural Business Statistics Regulation. Eurostat is collecting data and is also exploring quality in relation to the Short-Term Statistics Regulation. A key issue is the impact of statistical units on the quality of the statistical outputs. The UK and the Netherlands are considered by Eurostat to be the leaders in this field.

### 3.2.2 International Roundtable on Business Survey Frames

Some information on aspects of quality and quality measurement is available through the papers presented at the International Roundtable on Business Survey Frames. (The Roundtable is an informal meeting, generally held annually, of countries with an interest in developing their business registers. Membership varies and there is no formal constitution.) Because most papers describe quality initiatives in only one country, it is difficult to use them for making comparisons of quality between countries.

The meeting in 1995 attempted to present some comparative data on quality. Several countries completed a questionnaire covering a wide range of issues:

- statistical units and plans for change;
- data sources and plans for change;
- management of quality measurement;
- counts of enterprises and local units;
- coverage (included and excluded units);
- quality management (undercoverage and overcoverage) and plans for change;
- industrial classification, its quality measurement and plans for change;
- burden on business;
- geographical classification;
- country of ownership;
- sector classification;
- size measures; and
- marketing, and measurement of customer satisfaction.

The response to the questionnaire provided some descriptions of the type of activity that was being undertaken in the various countries. However, the questionnaire did not provide a good basis for the comparison of quality across countries. As within Europe, issues relating to the quality of variables were not covered. Based on the information that can be gleaned from the report, the IDBR appears to be one of the better business registers.

Quality has also been an important item at the last two Roundtables, one in Paris in 1999 and the other in New Zealand in 2000.

In Paris, seven papers were presented covering the following four themes:

- quality problems regarding sources for business registers;
- overcoverage because of falsely active units;
- measurements of the quality of business registers; and
- means to improve the quality of business registers.

A whole day and ten papers at the New Zealand Roundtable were devoted to the subject of quality. The papers concentrated on:

- defining the concept and scope of quality measurement;
- measures to improve the quality of business registers;
- measurement of quality of business registers; and
- methodological approaches to improving quality (on employees and classification changes).

Discussions during the last two Roundtables have showed a broad agreement on the main quality issues to be considered. The papers on quality in New Zealand (although varied) promote a common theme and deal with an extensive list of issues that should be addressed to ensure business registers are of high quality. The discussions in New Zealand concluded that brief and aggregate quality indicators are required. There is such diverse use of business registers that to provide quality assessments to meet each user's needs would not be practical.

However, a limited set of disaggregate measures is also seen as being important in delivering a good set of quality indicators for a business register.

Having agreed a common set of quality indicators, it is important to be able to assess how the measurements against these are perceived by users, and how they compare with other National Statistical Institutes (NSIs). To this end, the Roundtable has established projects to enable NSIs to consider how and why other NSIs are performing better in certain areas. The projects will:

- consider the quality issues that should be most important to every NSI business register, and produce a list of those that can be measured periodically; and
- develop a questionnaire so that each NSI can provide a series of details about its business register for every Roundtable meeting. The details would include size of register (number of units, employment), sources, main users and quality indicator measures (to be agreed by the first project group).

Quality will continue to be an important topic for the Roundtable, particularly with regard to the questionnaire that will be developed.

### **3.2.3 United Nations Economic Commission for Europe**

The United Nations Economic Commission for Europe (UNECE) is responsible for the dissemination of good practice throughout Europe. Every two years, UNECE organises a meeting to examine this subject, which it co-ordinates with Eurostat's Business Register Committee.

One action was to create an information sheet, which would provide a summary of the business registers in member countries. Information was collected on:

- legislation,
- units of the Regulation,
- other units,
- sources,
- variables,
- distribution of enterprises and legal units by NACE Rev 1 sections (NACE Rev 1 is equivalent to SIC(92) down to class level), and
- counts of total and registered enterprises.

This information has not been kept up to date due to lack of resources within UNECE.

## **Chapter 4**

### **Quality Measurement**

#### **4.1 Introduction**

Quality measurement is important for two reasons. First, it allows users to make an assessment of a product's quality that they might have been unable to do independently. This is especially so in the case of a business register where some concepts may not be easily understood or appreciated. Quality covers many attributes, including accuracy, timeliness, accessibility, interpretability, consistency, compatibility and completeness.

Second, measurement of quality helps people who are responsible for managing an output to determine the quality levels required, and to devote appropriate resources in the most cost effective way. Across statistical programmes there are differing objectives, priorities, constraints and opportunities, and thus differing quality characteristics. One of the challenges of maintaining the IDBR is to strike an appropriate balance between the evolving needs of clients, costs and the burden on respondents, and the various dimensions of quality.

In this chapter we review the work ONS has done (some with funding from Eurostat) to develop measures of register quality and then recommend an expanded range of indicators that should be used to measure the quality of the IDBR. This chapter is not about quality assurance, which is covered elsewhere in the report.

#### **4.2 Review of Register Quality Measurement in the UK**

Various quality initiatives have been undertaken during the late 1990s aimed at measuring the quality of the Register. Some of the most important have been to:

- conduct a Register Quality Survey;
- conduct a pilot survey to elicit users' views on quality; and
- contribute to a Europe-wide project on developing model quality reports.

##### **4.2.1 Register Quality Survey**

###### *4.2.1.1 Direct Quality Measurement from the Register Quality Survey*

In 1998, BRU conducted a Register Quality Survey (RQS) of the IDBR,<sup>9</sup> with funding from Eurostat. The topics that were addressed as part of the survey were dictated by the funding available for the project and the need to address the range of register quality issues that Eurostat was interested in. As such, the analysis of the results of the survey were based on these and aimed at measuring the following aspects of quality that are important to Eurostat:

- overcoverage due to duplication and falsely active enterprises;
- undercoverage due to matching problems between VAT and PAYE administrative data;
- errors of industrial classification;
- errors of size measures – employment and turnover; and
- delineation of business structures.



The study found that the quality of the IDBR was good in most respects, but highlighted areas where further improvement was required. Overcoverage and undercoverage only occur to a small extent. Register employment was shown to be of good quality, with the exception of enterprises based only on VAT data (where employment is imputed from turnover). The main quality problem lies with industrial classification. Some delineation problems were highlighted, but these did not have an adverse effect on the Register.

Overcoverage on the IDBR is due to enterprises that have ceased trading, but have not been removed from the Register, and duplication of enterprises. Overcoverage is more prevalent within the smaller sizebands. The survey indicated that overall, overcoverage due to falsely active enterprises accounted for about 0.8 per cent of the IDBR population, and the level of duplication was the equivalent of 0.5 per cent of the population.

Two sources of undercoverage measured were: unmatched PAYE units that have not been included on the IDBR; and reporting units that have been removed from the Register because they were identified as dead by a statistical survey, but are currently trading. The RQS found that a third of PAYE-only units examined were genuine, and should have been included in survey populations. Applying this rate to all the unmatched PAYE-only units on the IDBR indicates a level of undercoverage equivalent to about 0.9 per cent of the Register population. Only a few cases of reporting units thought dead by surveys were actually trading; these had an immaterial impact on the survey population.

Total employment was estimated to be in error by about 0.5 per cent. There were large differences for relatively few enterprises and these were found to be related to structural problems that would have no impact on the overall estimates of employment. Of most concern were enterprises where the employment value is derived from VAT turnover.

The major concern identified by the survey is the quality of industrial classification. About a third of small enterprises (fewer than 20 people employed) were incorrectly coded at the SIC(92) subclass level. About a fifth of all classifications was wrong at the division level, the majority of which were from data provided by HMCE. Examined by SIC(92) code, most of the problems are from enterprises coded to the 'other' and 'not elsewhere classified' categories particularly in the service sector. Further work on the administrative inputs to the IDBR and classification guidelines was identified, and this is discussed in the next chapter.

#### *4.2.1.2 Combined Effect on the Survey Population*

In addition to direct measures of quality from the RQS, ONS has begun to consider measuring the overall effect on the survey population. It is important for survey users that the IDBR shows the right distribution across sizebands, regions and industries. In general, some errors will cancel each other out, resulting in a smaller overall effect than that suggested by a raw measure of error. We endorse the continuing development of this work.

### **4.2.2 Register Quality Matrix**

Following the RQS, the link between the five quality criteria and their impact on users was examined. A matrix was developed that reported, for several aspects of quality, both their importance to users, and users' perceptions of each of these aspects. The aspects of quality

considered were those that Eurostat identified as being important – industrial classification, employment, undercoverage, overcoverage and business structures. As a pilot exercise, some of the main IDBR users (National Accounts, ONS surveys and users in OGDs) were contacted. The questionnaire that was used is included as Annex 5. The results are summarised in the table below.

**Table 5 Quality Rating and Impact by Quality Criterion and User**

	ONS's Economic Statistics Directorate	DTI		DfEE	DETR	DETINI	MAFF	Scottish Executive
		Small Firms	Other					
SIC(92)	3 / 1	3 / 2	3 / 1	4 / 1	2 / 3	1 / 1	4 / 2	3 / 1
Employment	3 / 1	2 / 1	3 / 2	4 / 1	2 / 2	1 / 1	4 / 3	2 / 1
Undercoverage	3 / 2	-	3 / 2	- / 1	2 / 3	1 / 1	4 / 3	2 / 2
Overcoverage	3 / 2	3 / 2	3 / 2	4 / 1	3 / 2	3 / 1	3 / 3	2 / 1
Structures	3 / 1	4 / 2	2 / 2	4 / 1	3 / 4	3 / 1	4 / 2	2 / 3

*Source: IDBR Management Committee paper 04/00*

Note: the 'mark' for each box is  $x / y$ , where  $x$  is the rating between 1 and 5 for the quality rating (1 = excellent to 5 = poor) and  $y$  is the rating for the impact of the criterion on the user (1 = high and 5 = low).

These results show that the perception of the IDBR's quality was average overall. Industrial classification and business structures were seen as needing most attention. The development of the matrix elicited a range of responses from customer areas that were largely similar to those we obtained from our user consultation described in Chapter 2.

### 4.2.3 Model Quality Reports

ONS contributed to an international project – *Model quality report in business statistics*<sup>10</sup> – part-funded by Eurostat to develop a quality measurement framework and model quality reports for business statistics. Within that, a section was devoted to the theory and methods for assessing frame errors.

### 4.3 Proposals for Register Quality Measurement in ONS

In the remainder of this section, we propose a range of indicators that should be used to measure the quality of the most important aspects of the IDBR. The proposals take into account users' views, previous work in ONS and published work from around the world. The measures can be used to report the quality of the IDBR directly, or for survey sections to measure the impact of changes to the IDBR on their outputs. We propose only quality measures, not to determine what constitutes high quality – ONS should determine this once measurement of quality is established. In particular, there are many aspects of quality that have different priorities for different users.

The indicators we propose can be measured relatively straightforwardly from the IDBR and the ARI. Some also need data from the administrative sources, a customer satisfaction questionnaire, or management information relating to IDBR processes. In general, we propose that quality measures be published annually, although in some cases, more frequent measurement will be needed for diagnostic purposes. We note that some measures are already produced, and are recorded here for completeness.

**Recommendation 1** *ONS should publish regularly a wide range of measures of both the level of the IDBR's quality and the change in quality over time. The existing user satisfaction questionnaire should be improved to include a wider range of questions relating to the IDBR's quality and that of the service provided by the Business Registers Unit. (Priority: 1)*

In the rest of this section we outline the range of areas for quality measurement.

### **4.3.1 Indicators of Coverage**

It is important to assess the coverage of the IDBR, both in terms of businesses that are potentially missing (undercoverage) and businesses that are erroneously recorded on the Register (overcoverage).

Undercoverage and overcoverage due to delays in recording births and deaths of businesses can be estimated from administrative data. A method for measuring undercoverage and overcoverage due to the lags in recording births and deaths was developed and presented in ONS's study into births and deaths of businesses.<sup>11</sup> We recommend that this be used. A summary of the method is attached as Annex 6.

A more timely indicator of the potential for coverage deficiencies is the average lag or distribution of lags between births and deaths and their update on the IDBR. In principle, these measures could be calculated as the updates from HMCE and IR respectively are taken onto the IDBR. The lag gives an indication of the extent of undercoverage.

For measuring undercoverage below the registration threshold, the method used by DTI (described in Chapter 9) is the presently accepted method. We have proposed elsewhere in this report that that method be updated.

Overcoverage due to duplication is an important measure of quality. First, it gives important information about the quality of our processes to link VAT and PAYE units – if the level of duplication is rising, for example, this is a clear indication that there may be some problem with the matching routines. Second, it provides information for survey users to consider whether they need to incorporate some mechanism into their results process to account for the level (or changing level) of duplication.

The presence of non-standard reporting units (see section 8.3) can affect consistency of outputs as populations may be defined differently, and double-counting may occur. The number of inquiry-specific reporting units on the IDBR should be measured in this context.

### **4.3.2 Quality of Auxiliary Information**

Some measures of the quality of auxiliary information have already been developed by ONS. In addition to the present list of industrial classification, employment, and region, the quality of turnover and legal status codes should also be measured.

We consider there should be two measures of quality. First, there is the absolute level of correctness, that is simply the number or proportion of reporting units that have the correct code. For SIC(92), this could be measured at any of the levels. The RQS measured quality of coding at division and subclass level, and this seems reasonable. Likewise for legal status, it may be worthwhile not only to measure the quality of coding at the detailed level, but also to measure the extent of inaccuracy at the public/private level.

The second indicator of quality is the effect of these errors on the sampling population. This is based on work described in section 4.2.1.2 to assess the overall impact on the IDBR of known inaccuracies.

The proportion of reporting units that revise their industrial classification to a code that they had recently held should be measured. This will give an indication of the innate volatility in the coding of industrial classification.

In addition, the total employment as recorded by the IDBR should be compared against other similar published measures of employment. It is impossible to tell from any difference which of the two measures is more accurate, but changes in the difference could be a useful diagnostic tool.

### **4.3.3 Consistency**

In addition to the accuracy of information, it is important for many users that data held on the IDBR can produce consistent outputs. This is particularly the case for National Accounts and others who are trying to draw together information relating to the whole economy. It is important that the data used to update the IDBR from a variety of sources are consistently defined, specifically for the main auxiliary variables: SIC(92), employment and turnover. Consistency applies to both the data at a single point in time, and to the data over a range of time periods.

It is easy to use the IDBR to measure the number or proportion of reporting units, employment or turnover that have data derived from the various sources. Since the definition of the variable used by each of these sources is known or easily determined, it is easy to report the extent to which information on the IDBR is defined consistently.

### **4.3.4 Currency of Information**

Another important quality measure is how up to date the Register is. This review has highlighted several areas where out-of-date information has created difficulties for users. It is important for the Register to be up to date if it is to reflect the current state of the economy. Having up-to-date information on business structures and auxiliary information is particularly important.

The IDBR itself can be used as the tool for measuring currency of information as the date to which information relates is stored alongside the data themselves. The information could be presented in a form similar to that which we used in Table 10 in section 6.4.2. ONS should at least aim to ensure that the requirements of the EU Regulations are met and that no information is more than four years out of date.

It is also important to make sure that business profiles remain up to date, particularly as the structures of the larger, more complex enterprises can change often and rapidly. This can be measured by reference to the latest date that the business profile (rather than details of individual local units) was checked.

#### **4.3.5 Availability of Contact Information**

The availability of contact details should also be measured. This is particularly important for survey users where access to businesses is essential. Therefore, we propose the measurement of the number (or proportion) of reporting units for which phone numbers, postal and e-mail addresses are available. This does not address whether the contact information is correct, and more research is needed to determine how best to calculate such a measure.

#### **4.3.6 Process Quality**

The Register's quality can also be assessed by measuring the quality of the processes associated with it. There are close links between the two as high quality processes contribute to high quality outputs.

The regular audit of online amendments discussed in section 10.2 should be used to measure process quality. At present, some limited management information is produced which reports the number of amendments made during given time periods. The results of the audit are classified into a range of outcomes, such as 'comments', 'minor clerical error', 'significant clerical error' and so on. The number of amendments in these categories should be published on a regular basis. Given the present size of the audit (60 amendments per audit), information that covers several audit periods may need to be collated to ensure a meaningful indicator.

#### **4.3.7 Level of Register Service**

An important measure of the IDBR's usefulness is the speed and accuracy with which users have access to the Register.

For online users, a guaranteed service is important. At present, the proportion of time that the IDBR is available for interrogation (viewing) and updating is reported. Survey users require samples to be selected accurately and at the correct time, and ONS should report the number or proportion of samples that have been selected to specification and on time. Likewise, users of register analysis need accurate analyses, produced quickly. Here, we propose that the time delay between requests for analysis and the particular analyses being produced be measured.

#### **4.3.8 User Satisfaction**

ONS has already trialled a questionnaire to elicit users' requirements for, and perceptions of, quality. In addition to more objective and quantifiable measures, it is important to monitor regularly that users' requirements are being met. Our user consultation suggested that it is difficult for users to be able to pin down exactly what level of quality they need from the IDBR, and a specific questionnaire can help them to do that.

In conducting a user survey it is possible to ask users both what quality they require and what quality they currently perceive against a range of register characteristics. In doing this, we can avoid the need for users having to specify exactly what quality level they require by asking for responses on a scale from, say, very high to very low. From this, it is possible to measure whether aspects of the IDBR meet users' needs without being concerned about the absolute level of quality that is attained. Statistics Netherlands has recently used a customer satisfaction questionnaire that collects information mostly relating to the quality of service provided by their Business Register division. Questions are asked about the level of quality that is required, and the level of quality that is perceived, both on a seven-point scale. This is then used to measure quality and improvements to quality. The subject matter of the questions is attached as Annex 7 to this report.

ONS should expand the range of questions in its questionnaire beyond the minimum range of quality indicators in which Eurostat has been interested. This should include questions relating to the service provided by BRU. Good practice from the Dutch questionnaire should be adopted.

#### **4.3.9 Local Unit Quality Indicators**

The quality measures discussed in this section relate to reporting units. Business structures, and local unit information in particular, are important to an increasing number of users, as the demand for statistics on small areas increases. Of particular importance is the existence or not of a local unit on the IDBR. Therefore, it is important that quality indicators for local units are developed in addition to those relating to reporting units.

### **4.4 Producing Quality Measures from the Annual Register Inquiry**

Although the ARI is clearly a potentially useful tool in measuring the IDBR's quality, using this source to measure quality isn't necessarily an easy task, and requires some care. In particular, production of estimates of quality measures from the ARI relies on the ability to use the ARI as a survey, producing estimates with sampling weights. Our work in trying to assess the impact of classification and employment errors on the IDBR reported in section 4.2.1.2 has highlighted some difficult problems that will need to be faced. In trying to measure quality, it is essential that the appropriate planning goes into the design of the ARI sample and processes.

## 4.5 Summary of Proposed Quality Measures

In this section, we summarise the range of specific quality indicators that should be considered, in conjunction with users, for publication. The range here is not comprehensive.

### Coverage

- Level of undercoverage due to birth lags.
- Level of overcoverage due to death lags.
- Level of undercoverage arising from businesses below the registration thresholds.
- Average lag and distribution of lags between businesses starting to trade and being added to IDBR.
- Average lag and distribution of lags between businesses ceasing to trade and being removed from IDBR.
- Extent of overcoverage due to non-matching of duplicate VAT and PAYE units.
- Extent of undercoverage due to omission of PAYE-only units from survey populations.
- Number of inquiry-specific reporting units on the IDBR.

### Quality of Auxiliary Information

- Proportion of reporting units/employment/turnover with:
  - incorrect SIC(92) code at division level;
  - incorrect SIC(92) code at subclass level;
  - incorrect legal status code;
  - incorrect legal status code at public/private sector level;
  - incorrect employment;
  - incorrect turnover; and
  - SIC(92) code being changed to a code held in the previous two years.
- Difference between IDBR employment and employment figures published elsewhere.

### Consistency

- Number of reporting units/employment/turnover with consistently defined:
  - SIC(92) codes;
  - employment; and
  - turnover.

### **Currency of Information**

- Distribution of reporting units/employment/turnover by currency of update of:  
    employment;  
    turnover; and  
    SIC(92).
- Distribution of complex businesses by currency of profiling.

### **Availability of Contact Details**

- Number of reporting units/employment/turnover with missing/incorrect:  
    name;  
    address;  
    postcode;  
    phone number;  
    fax number; and  
    e-mail address.

### **Process Quality**

- Distribution of outcomes from the audit of online amendments.

### **Level of Service**

- Distribution of the timeliness of register analysis.
- Distribution of the timeliness of sample selection.
- Percentage availability of the IDBR for interrogation and updating.

### **User Satisfaction**

- Comparisons of user requirements against user perceptions for a range of aspects of frame quality.



## **Chapter 5**

### **Administrative Sources**

#### **5.1 Introduction and Summary of Main Quality Issues**

In this chapter we describe the processes by which ONS obtains data from administrative sources, principally those concerning the birth of new businesses, and the closure of existing ones. The major inputs are from HMCE and IR, which together provide the coverage and range of variables necessary to support the enterprise level within the IDBR. The main part of this chapter focuses on the HMCE processes because they are more extensive and complex than those relating to IR. Companies House (CH) provides supporting information relating only to incorporated businesses.

Data on births from HMCE are received weekly, soon after HMCE itself is notified of changes. In addition, annual turnover figures are of high quality, as these are required by law to be provided by traders registered for VAT. Likewise, the employee figures received from IR are also of high quality.

Industrial classification data are of lower quality. In HMCE, data are collected from businesses using a question that has no explanatory notes, and coding is done without the use of expert coders or electronic coding tools. Within IR, the industrial classification system is not aligned with SIC(92). Turnover figures supplied by HMCE for new businesses are of lower quality than the main quarterly return, as they are estimated in advance by the businesses concerned.

#### **5.2 Comprehensive Business Directory**

The Comprehensive Business Directory (CBD) project is a recognised strategic project that has been researched and developed with 'Invest to Save'<sup>12</sup> money and additional departmental funding. Work on the CBD has produced a technology-based model that can be used to provide a government-wide service for locating, matching and presenting data to operational services in relation to business data. It is also undertaking extensive legal research to ensure that the project's current partners (HMCE, IR, CH and ONS) are informed of the legal and policy changes required to allow implementation of the CBD. The legal work also acts as a signpost for work to bring on new partners.

The CBD brings together data sets from both the public and private sector thereby providing considerable opportunities for more responsive and effective business-with-government services. Greater accessibility will improve the information available within government which, in turn, will lead to better policy targeting. The CBD will be a business information locator; that is, it will hold information on what, and where, information relating to the business community is kept. In addition, the facility will provide a framework for authorising and facilitating access to elements of data where this can be provided. However, where data come together from different sources it is important to establish with foolproof accuracy that the data items refer exactly to the same unit of business. Matching can be direct (based on common reference numbers), or contextual (based on, for example, business function or type of work).

For the ONS specifically, the CBD will improve access to administrative data and reduce the need to collect data directly from businesses. The IDBR will be improved by the better linking of source data.

The emergence of the CBD has created interest at operational levels and within policy units, including HMT where the formal response to the Cruikshank Banking Review<sup>13</sup> refers to a CBD. This demonstrates the greater application of business registers as tools for co-ordination.

A recent management report on progress in this project is attached at Annex 8.

**Recommendation 2** *The development of the Comprehensive Business Directory should continue to be a priority within ONS as a means of improving access to data from administrative sources, improving the quality of the IDBR through better matching of source data and of reducing the need to collect data directly from businesses. (Priority: 1)*

### **5.3 Registration**

Businesses that enquire about VAT registration are sent a registration form. This is returned to a Regional Registration Centre (RRC) where the information provided is vetted and, once a form with clean data is available, the business is registered. Some data from the registration form is entered onto the computer system. (Ultimately, it is information from this system that is fed, by way of HMCE at Southend, to ONS via magnetic tapes for maintaining the IDBR.)

Notably, for registered companies, the company number is entered on the registration form, but is only used to check the validity of the registration at CH. The company number is not entered onto the HMCE computer system, and consequently is not transferred to ONS. Resources are used in ONS Newport to establish the company number of registered companies by matching with a CH source. Therefore, there is an opportunity to reduce the efforts required in ONS by modifying the procedures in HMCE. In addition, quality should improve as there is less risk of missing a match. There is a small additional cost for HMCE staff, but we could expect net savings across government as a result.

In the case of group registrations, the representative member of the group is entered onto the main HMCE computer system in the same way as standard registrations. Each non-representative member is recorded on a separate form. These forms are sent from the RRCs to HMCE in Liverpool, where they are collated and sent to ONS Newport. It is understood that these cannot be included in the main computer system at HMCE because of the way that VAT registration numbers are allocated. It is the responsibility of clerks in ONS to create new IDBR records for non-representatives from the paper forms received from HMCE.

### **5.4 Deregistration**

A deregistration pack is sent to businesses that ring up to cancel their registration. Once HMCE is sure that the cancellation is appropriate, the registration is cancelled. The account remains open until all tax due is paid, a process that can take many months or years to complete.

## **5.5 Registration and Deregistration Dates**

The effective date of registration is taken to be the earliest of the many dates that can be entered on the VAT registration form. The deregistration date is listed as the day after the last 'taxable supply' is made. When a trader ceases to be liable for VAT purposes, the trader deregisters. At this stage, ONS receives the deregistration date, and HMCE processes the information and labels the trader as a redundant trader. A file of redundant traders is sent to ONS twice a year to confirm the deregistrations. In general, there is only a small lag between the RRC processing the registration and ONS being sent the information, because new computer records are sent daily from the RRC to HMCE's Southend office.

## **5.6 Industrial Classification**

A main business activity description is collected on the VAT registration form. The question is simply 'Describe your main business activity'. There are no notes to clarify how businesses should determine their main activity. It is expected that businesses will list the activity that contributes most financially. However, this doesn't always happen. HMCE staff highlighted cases where businesses had reported what they considered was their major activity, perhaps in terms of time spent, investment, or their major business activity at the time they were established. Within 18 months of registration, an Assurance Officer makes a visit to each registered trader. It was reported that in some cases either the business completed the registration form incorrectly, or that their activity had changed by the time the visit was made.

Clerks at the desk convert the business description into a SIC(92) code. (There are some differences between the version of SIC(92) used by HMCE and that used by ONS.) Clerks receive no training in the classification of businesses, and don't appear to have the level of understanding of the codes' use required by ONS. The use of expert (electronic) coding systems and improvements to the descriptions of activity collected from businesses will improve the quality of industrial classification coding. These issues are being addressed as part of the CLAMOUR project, details of which are attached at Annex 9.

At present, a Ministerial Direction permits ONS to feed back industrial classification information to HMCE. Such feedback may benefit the business community by reducing the burden on business by the single collection of information about businesses' activity. Feeding back could also improve consistency because it would require that industrial classification standards be adopted throughout government.

## **5.7 Turnover**

At the time of registration, businesses are asked to estimate their turnover for the next 12 months. HMCE makes use of this to distinguish between compulsory, voluntary and intending registrations. The precise accuracy of the figures is unimportant to HMCE.

However, the turnover figure is important for new registrations in ONS, as it is used directly for analysis and as an auxiliary variable in estimation. It is also used indirectly to determine stratification variables, whether a business should be included in ONS survey populations and

to determine whether particular enterprises are small enough to be given survey holidays. ONS needs to bear in mind the quality of such information when using the figures.

Turnover data that are supplied to ONS quarterly from HMCE are of high quality as businesses are required by law to submit accurate and timely figures to HMCE.

## 5.8 Inland Revenue

The discussion so far has focused on the processes for registering VAT traders at HMCE. We have not carried out such an in-depth investigation of the processes at IR, because the processes there are less complex.

Every quarter, ONS receives IR data relating to employer PAYE schemes. The data include counts of employees who pay tax, which are used to update employment for those enterprises that have not been included in the ARI. In addition, ONS depends on PAYE data for industrial classification for the relatively small proportion of businesses not registered for VAT. However, IR uses its own industrial classification system, the TCN, which is not aligned with SIC(92) and is of low priority within the IR system. The discussion above regarding the improvement of industrial activity coding applies equally to IR.

In terms of extra information not already accessed by ONS, the amount available from the registration process and ongoing maintenance of PAYE records is not on the same scale as that from HMCE. The main improvement to IDBR quality can be made by the more frequent receipt of information, and receiving information electronically rather than on magnetic tapes. More frequent transfer of data improves the IDBR's coverage by reducing the lags between real-world events, and recording those events on the Register. More frequent receipt of employer data also opens the possibility of investigating the direct statistical use of those data.

## 5.9 Companies House

Data from Companies House are used mainly to identify live businesses and their legal names, which helps link corporate VAT and PAYE records. Industrial classification and addresses for incorporated businesses are also used by ONS in maintaining the IDBR.

**Recommendation 3** *Building on the existing relationships with HM Customs and Excise (HMCE), Inland Revenue (IR) and Companies House (CH), ONS should:*

- *seek to extend the range of data supplied by the administrative departments to include, from HMCE, company numbers and previous VAT registration details and, from all administrative sources, comprehensive contact information (Priority: 1);*
- *seek to improve the method of transferring data from the administrative sources and seek more frequent transfer of data (Priority: 3);*
- *consider feeding back industrial classification data to the administrative departments within the scope of the prevailing legal framework (Priority: 2); and*

- *seek to improve the quality of industrial classification coming from administrative sources by using consistent data-collection and coding techniques, and by supporting those people who classify businesses. (Priority: 1)*

### **5.10 Setting and Monitoring Quality Standards**

At present, service-level agreements exist between ONS and supplier departments. These agreements set the standards for variables to be delivered, format and timeliness, but don't mention quality. It is important that quality standards are specified for all data supplied for updating the IDBR and, while ONS receives data from administrative sources, quality standards should be specified.

**Recommendation 4** *Quality standards for industrial classification, legal status, company number and address should be agreed and documented by ONS and the administrative departments. (Priority: 1)*

## **Chapter 6**

### **Annual Register Inquiry**

#### **6.1 Introduction and Summary of Main Quality Issues**

In this chapter, we document the current practice of the Annual Register Inquiry (ARI). Then, we describe the improvements that will ensure the ARI is the most effective register-maintenance tool possible.

At present, the ARI is essentially the same as the previous Annual Employment Survey (AES), which was used to produce estimates of employment. The ARI has different objectives, being solely a tool for register maintenance. The ARI includes all enterprises with at least 100 people employed, and a small range of those that are thought to have inconsistent employee data. It includes in its sample about 70 per cent of employment recorded on the IDBR. The ARI collects details from individual local units, information that is not available from any other source. The ARI has also developed an efficient system of collecting data electronically from the largest businesses.

The ARI sample is selected in such a way that feeding information back to the IDBR from the sampled strata can lead to bias in the Register. With regard to sample design, we consider that the sample of small enterprises is too large for the ARI's current purpose, and not enough enterprises showing inconsistent data are being included in the ARI sample. There has been a considerable degree of overlap between the ARI and the work of the profiling unit, which has led to duplication of effort in the past. We are pleased to note that new organisational arrangements for 2000 have reduced this.

There are some problems with the ARI questionnaire. These are partly due to attempts to adapt the questionnaire used for the AES, and integrate it into the operational systems for ARI. Weaknesses in the questionnaire have been identified by this review and in two separate pieces of recent work within ONS. Finally, follow-up of non-respondents needs to be improved to reduce the quality risks associated with repeated non-response to the ARI.

#### **6.2 Purpose of the Annual Register Inquiry**

The ARI was established as a result of the development of the Annual Business Inquiry (ABI) to provide high quality information on business structures. The ARI, which was first run in 1999, was developed from the AES. The AES had two purposes: to produce national and small area employment estimates, and to maintain information relating to business structures. The ARI is somewhat different from the AES in that the ARI is simply intended to maintain structural information about enterprises on the IDBR, particularly the largest enterprises where local unit information is not available from any other source. This need is emphasised in the context of the work on neighbourhood statistics, which requires local unit information to be present. Consequently, its design must reflect that.

The ARI is run in two parts. Part 1 (ARI/1) is an annual survey of approximately 68,000 enterprises, which collects information from businesses about each of their sites. In addition, VAT and PAYE references are collected to facilitate the matching of businesses registered under the two administrative schemes. Part 2 of ARI (ARI/2) is a simpler, rolling survey

which is conducted on a smaller scale and is undertaken every quarter. ARI/2 includes new births onto the IDBR. Its purpose is to ratify the information received from the administrative sources, check whether the business is genuinely new or is connected with an existing IDBR enterprise, and to establish the sites operated by the enterprise. For the rest of this review, we are concerned only with the more complex ARI/1, and will use ARI to mean ARI/1. However, many of the principles relating to ARI/1 also apply to ARI/2.

In section 6.3 we outline some of the main characteristics of the ARI and, in section 6.4, discuss changes that should be made to ensure the ARI fulfils its aims.

## 6.3 Current Practice

### 6.3.1 Sample Design

The ARI presently has a sample allocation of about 68,000 enterprises. The following table describes the breakdown of the sample into the strata used for ARI.

**Table 6 Population and Sample Sizes of ARI**

Category of Enterprise	Sample size 1999	Sample size 2000	Population size 2000
Enterprises with employment of 100 or more (reporting for the whole enterprise)	20,578	20,580	20,603
Enterprises with employment of 100 or more (reporting for parts of the enterprise separately)	811	971	971
Enterprises with unusual characteristics (see below)	1,415	1,110	1,110
Enterprises with 20–99 people employed	20,000	19,999	73,900
VAT-based enterprises, 10–19 people employed	6,000	6,000	101,044
VAT-based enterprises, 0–9 people employed	14,000	14,001	1,292,652
PAYE-only enterprises, 10–19 people employed	1,600	1,600	20,229
PAYE-only enterprises, 0–9 people employed	3,400	3,400	358,413
Enterprises with special (disk) reporting arrangements	165	167	167
Enterprises with special (paper) reporting arrangements	43	12	12
Total	68,012	67,840	1,869,101

*Source: ARI, July 1999 & July 2000*

The ARI sample is selected from the whole of the IDBR with the following exclusions:

- certain SIC(92) divisions, namely, division 01 – Agriculture, hunting and related service activities, division 95 – Private households with employed persons, and division 99 – Extra-territorial organisations and bodies; and
- enterprises awaiting proving.

Note that because of its special function of updating the IDBR, the ARI selects unmatched PAYE units, which are usually excluded from ONS business survey populations because of potential duplication.

### 6.3.2 Enterprises with Unusual Characteristics

The ARI includes enterprises for which there is a large discrepancy between the employee figures shown on the IDBR and by the PAYE source. Specifically, enterprises are included if the average of the four quarters' PAYE jobs figures, and the number of employees shown by the local units, do not satisfy the following inequalities:

$$\frac{PAYE + 20}{3} \leq IDBR \text{ Employees} + 20 \leq 3(PAYE + 20)$$

### 6.3.3 Sample Rotation

The sample is selected according to the standard ONS method of Permanent Random Numbers (PRNs) described in Annex 10. There is no year-on-year overlap in the sampled strata, that is, all enterprises in the sampled sizebands are rotated out each year.

### 6.3.4 Coverage

The ARI covers around 70 per cent of employment on the IDBR. Enterprises with 100 or more employment, which account for almost 65 per cent of the total employment on the IDBR, are completely enumerated each year. The following table shows the distribution of employment by ARI sizeband.

**Table 7**      **Distribution of Employment by Sizeband**

<b>Employment band</b>	<b>Employment (m)</b>	<b>Percentage of total</b>
0–9	4.0	16.0
10–19	1.7	6.8
20–49 single site	1.4	5.8
20–49 multi-site	0.4	1.5
50–99	1.2	5.0
100 or more	16.1	64.9
Total	24.8	100.0

*Source: IDBR, June 2000*

### 6.3.5 Data Collection and Questionnaire Design

In most cases, data are collected using paper self-completion questionnaires that are mailed to businesses, and returned to ONS on completion. The following information is collected, or checked on the questionnaire.



From the enterprise:

- information about whether the enterprise is currently in operation;
- VAT and PAYE reference numbers;
- number of working proprietors; and
- number of workplaces.

For local units that are stored on the IDBR:

- information about whether the local unit is in operation;
- description of its business activity;
- company number (if operated by a holding company);
- number of employees (split into males and females, full and part-time); and
- reasons for any large changes in the number of employees.

For new local units, the same information (with the exception of reasons for any large changes in employment) is collected, as well as the unit's name and address.

For 200 enterprises covering 25 per cent of the local units in the sample, ONS has made special data-collection arrangements. In general, this has been done for the largest businesses, where respondent burden has been reduced by ONS's accepting data in a format that is easy for the business to supply. Data for some of the cases for which special arrangements apply have been collected on paper, and some electronically. For the 2000 survey, there were substantially fewer cases where data were collected on paper. Those providing data electronically did so by means of spreadsheets.

## **6.4 Proposals for Change**

### **6.4.1 Sample Design**

The structure of the current sample design is similar to that for the preceding AES, and largely reflects the needs of that survey – to produce estimates of employment. The largest enterprises are completely enumerated, and medium-sized businesses are included on a four-yearly basis to meet the requirements of the EU Regulation. In addition, the IDBR is updated from a sample of the smallest enterprises.

It is important that the ARI can facilitate the regular updating of local units with high quality information. Therefore, the focus must be on enterprises that are known to have, or are likely to have, more than one site. In addition, the IDBR's quality is improved by ensuring that enterprises undergoing structural change, or those where there is a potential problem with the structure (illustrated for example by disparate employment and turnover figures), are reviewed regularly.

### 6.4.1.1 Multi-site Enterprises

In order to maintain the structures of multi-site enterprises, ARI should approach those enterprises following Eurostat guidance. Structural information is available only for enterprises that have previously been sampled. Even then, the structure of a business may have changed since it was last included in AES or ARI. With regard to this point, we examine the distribution of multi-site enterprises by sizeband. The following table is based on responses to the ARI 1999.

**Table 8 Percentage Distribution of Local Units by Enterprise Sizeband**

Employment band	Number of local units		
	1	2	3 or more
0–9	95	5	1
10–19	82	13	5
20–29	76	14	10
30–39	69	15	16
40–49	65	17	18
50–59	59	18	23
60–69	54	19	26
70–79	53	20	27
80–89	52	18	30
90–99	52	17	32
100 or more	36	14	49

*Source: ARI, 1999*

As we would expect, the table shows that the proportion of multi-site enterprises decreases with the size of enterprises. For enterprises with 100 or more people employed, almost two-thirds have more than one local unit, and half have three or more. For employment between 50 and 99, almost half the enterprises reported more than one site to the ARI. On the basis of the figures in the this table, we identify that the IDBR’s coverage would benefit from the annual complete enumeration of businesses with at least 50 people employed.

The table suggests that the threshold for complete enumeration over four years should be changed from 20 to 10 people employed as one-fifth of businesses in the 10–19 sizeband have more than site. For simplicity’s sake all enterprises with 10–49 people employed would be grouped together, and completely enumerated every four years. This has a small additional compliance cost, the result of an increase of about 5 per cent in the number of local units included in the ARI sample, but offers more complete coverage of local units.

Only one in twenty of the smallest businesses (fewer than 10 people employed) have more than one site. We question the additional benefit of collecting site information for such businesses, and in section 6.4.1.3 we identify that the sample size of enterprises with fewer than 10 people employed could be reduced.

These sizebandings reflect the international definitions for business size used by Eurostat. These are shown in the table below.

**Table 9 Eurostat Descriptions of Enterprise Size**

<b>Size (employment)</b>	<b>Description of size</b>
0–9	Micro
10–49	Small
50–249	Medium
250 or more	Large

*Source: EU Recommendation 96/280/EC*

#### *6.4.1.2 Enterprises with Unusual Characteristics*

We discussed in section 6.3.2 the inclusion of enterprises that appear to have discrepant PAYE and survey employment. About 1,000 such enterprises have been included in the ARI for each of the two years of its existence. Note that many businesses meeting those requirements for inclusion have employment of 100 or more and are thus included in the completely enumerated stratum of large businesses.

In addition to these, further categories for complete enumeration in ARI should be considered. In this report, we offer two suggestions for enterprises that could be included.

#### *Enterprises Changing Size During the Course of the Year*

It is important to include all enterprises that appear to be undergoing a period of restructuring (in this context, restructuring is taken to mean the expansion and contraction in the number of local units belonging to the enterprise). This can be done by selecting enterprises that have shown a large change in the quarterly PAYE job figures throughout the year.

#### *Enterprises with Inconsistent Employment and Turnover*

Second, ONS business surveys collecting data on output variables, such as turnover, use turnover as the auxiliary variable in ratio estimation. Employment is used as the variable for size stratification. In some cases, the ratio of turnover to employment for individual enterprises is extreme, and this can cause instability in survey estimates. Some turnover and employment values are genuine, but some are known to arise where the structure of the enterprise on the IDBR is incorrect.

As part of this review, we investigated some enterprises with turnover-per-head figures that were extremely different from the average turnover per head for the particular industry to which the enterprise belonged. The research identified several reasons for such disparate figures. These included old local units that had not been updated since the 1993 Census of Employment, enterprises with no VAT unit, and enterprises with unusual structures.

Given the existence of disparate turnover and employment figures, the worst cases should be examined annually by complete enumeration in the ARI. It might be beneficial to include an additional rotating sample of some of the next worst cases.

#### 6.4.1.3 Small Enterprises

We have just discussed the need to ensure that the sample is correctly focused on the largest enterprises that are operating at more than one site. At present, 25,000 enterprises are selected from the 0–19 sizebands, approximately 17,000 of which are in the 0–9 sizeband. Since the ARI is not used to produce estimates of employment for the whole economy, little value is added to the IDBR through the inclusion of so many of the smallest enterprises – they account for a relatively insignificant proportion of businesses on the Register. Updating from these enterprises does not materially affect the quality of the Register. Within the requirements of the IDBR, and allied with the recommendations we have made to improve the quality of the administrative data, the sample size of the small enterprises should be set to provide adequate quality measures and control.

#### 6.4.1.4 Sample Selection

In order to facilitate the updating mechanism specified, it is important that the ARI sample be selected in such a way that it is statistically independent of other ONS surveys. At present, this is not the case, and needs to be changed. ONS will need to ensure that the mechanism complies with its guarantee that small enterprises will be included in only one survey sample at once.

**Recommendation 5** *ONS should revise the Annual Register Inquiry sample design and selection so that it:*

- *completely enumerates enterprises with employment of 50 or more;*
- *includes a wider range of enterprises with conflicting auxiliary information;*
- *includes, every four years, enterprises with employment of 10 or more;*
- *includes enterprises with employment of less than 10 only to the extent that is necessary for adequate quality measurement; and*
- *can be used to update the IDBR in an unbiased way. (Priority 1)*

### 6.4.2 Ensuring Response to the Annual Register Inquiry

The following table shows the distribution of IDBR employment by the date that the local unit employment data were last updated, and the size of enterprise to which the local units belong.

**Table 10 Percentage Distribution of Employment by Date and Enterprise Size**

Year of update	Enterprise size (employment)				All enterprises
	0-9	10-19	20-99	100 or more	
1991	0.3	0.6	0.1	0.0	0.1
1992	0.0	0.1	0.0	0.0	0.0
1993	28.7	40.2	9.3	0.2	8.5
1994	0.2	0.4	0.1	0.0	0.1
1995	1.9	3.8	2.7	0.4	1.1
1996	1.6	4.7	4.7	1.0	1.8
1997	3.8	8.4	12.2	5.5	6.2
1998	3.0	8.2	32.6	12.7	13.2
1999	3.2	10.8	34.8	47.8	36.6
2000	0.4	1.1	3.4	32.4	21.7
Unproven enterprises	56.9	21.8	0.0	0.0	10.7
Total	100.0	100.0	100.0	100.0	100.0

*Source: IDBR, June 2000*

The table shows that much of the employment recorded on the IDBR had not been updated for at least four years. In particular, one-twelfth of employment has not been updated since the Census of Employment in 1993. This has arisen because some of those enterprises included in the Census have not subsequently been included in the AES or ARI. In some cases, enterprises have been included in the ARI or AES (especially those in the completely enumerated sizebands, that is enterprises with employment of 100 or more), but own local units that have not been updated since 1993.

That this needs addressing has already been identified in ONS. It is proposed that all enterprises with information more than four years old should rely on the administrative sources, and have the local unit structures removed. This seems a reasonable solution as most of the enterprises with old data are the smaller single-site businesses. Implementation of the change as stated would introduce some instability into repeated samples of local units, so consideration needs to be given to the best way to address this issue.

In conducting a survey that aims to maintain the quality of the Register, it is important to ensure, as far as possible, that responses are received from all the most important businesses. Follow-up should be focused on ensuring that data are collected from the enterprises that have been updated least recently, and whose contribution to the economy is most important. The importance of having a response from the largest businesses cannot be stressed enough, and ONS should ensure compliance with the Statistics of Trade Act 1947 (SoTA) for these businesses.

## **Recommendation 6** *In the Annual Register Inquiry*

- *The response rate should be improved by following up non-response more effectively. The response rate for the largest enterprises should be 100 per cent. (Priority: 2)*
- *The general procedures used by ONS to ensure compliance with the Statistics of Trade Act 1947 should be rigorously applied to the Annual Register Inquiry. (Priority: 2)*

### **6.4.3 Data Collection and Questionnaire Design**

#### *6.4.3.1 Electronic Data Collection*

ONS is presently working to increase the number of special arrangements and to collect data via the Internet (ideally over the web, but possibly also by e-mail). In the short term, it is likely that electronic data will continue to be collected via spreadsheets. Ultimately, it is envisaged that web-based data collection will be available to a wider range of businesses. This will use improved data collection instruments that could include online validation and coding of industrial classification.

#### *6.4.3.2 Paper Questionnaire*

The ARI questionnaire is an important tool in the cost-effective maintenance of the IDBR. The questionnaire allows for the accurate collection of relevant information, while imposing the smallest possible burden on responding businesses.

The questionnaire for ARI was based on that used for the AES as the surveys were somewhat similar, and the ARI had to be consistent with the AES until employment estimates from the ABI were proved to be acceptable. The questionnaires for both AES and ARI have been reviewed recently, the AES questionnaire as part of the work on reconciling employment estimates from ABI and AES, and the ARI questionnaire as part of the pilot study into the expert review of business survey questionnaires. Here, we pick up some of the main findings of those two reviews, and make other observations that have arisen throughout the course of this review.

#### *Background to the Purpose of ARI*

First, not enough information is given about the ARI's purpose and the importance of the ARI and the IDBR to the whole range of economic statistics. An important step in maintaining a high quality register is to convince businesses of the importance of the ARI.

#### *Undercoverage of Local Units*

A major reason for the incorrect recording of data in the ARI is the omission of new local units. Work to reconcile the differences between employment estimates in AES and ABI uncovered a good deal of under-reporting, leading to under-estimates of total employment of up to 500,000 employees, which accounts for between 1 and 2 per cent of total employment. The ARI questionnaire requires respondents to complete a sheet for each local unit. Details of

existing local units held on the IDBR at the time of sample selection are sent to businesses for updating and return. Businesses are asked to complete a blank sheet for each new local unit. The work to reconcile employment estimates from AES and ABI found that some businesses failed to report new local units. This is an important issue for coverage of both local units and the employment auxiliary variable, and represents a real risk to the quality of the IDBR and economic statistics. Therefore, it must be addressed as soon as possible.

A related problem is that the questionnaire asks businesses to report whether they have sold any of their sites to other enterprises. If such a sale is reported, the corresponding local unit is simply closed on the IDBR. Unless the purchasing enterprise has responded to the ARI, the local unit will be missing from the IDBR. It may be possible to counteract this by revising the ARI questionnaire.

### *Delineation of Business Units*

Clearer delineation of the business unit to which the questionnaire is being sent is required. This helps to prevent both under-reporting and double counting. This is particularly necessary for the more complex enterprises and for businesses that have recently changed their structure.

### *Business Activity*

Another major area of concern is the collection of information describing businesses' activities. Several areas of our review highlighted the fact that improvements in the way business activity is collected, both in ARI and more generally, are necessary and possible. We emphasise the need to ensure that such data are collected in a consistent way, and that, wherever possible, collection is not duplicated.

The explanatory notes to the ARI questionnaire relating to the business description question don't specify the kind of information that should be supplied, or in what depth the information is required. In particular, there is no indication of what ONS means by 'main' business activity.

To reduce respondent burden, ONS feeds back previously supplied business descriptions, asking businesses to correct the description only where it's wrong. This reduces the need for businesses to enter the same text in subsequent years. For local units where ONS does not have a business description, the generic heading of the SIC(92) subclass to which the local unit has been classified is fed back. It is suspected that such feedback of information causes under-reporting of the number of changes to IDBR industrial classifications, as businesses would be less likely to correct a classification that looked reasonable but wasn't right.

The procedure for feeding back previously recorded information to businesses should be evaluated to consider whether it can be improved. In particular, it may be beneficial to feed back both business description and the SIC(92) subclass heading to which ONS has coded an enterprise. If information is available from another source, it may be possible to integrate this into the feedback mechanism. It may be possible and beneficial to vary the style of feedback according to the confidence with which a particular code has been allocated.

Some businesses have hundreds or thousands of local units. In many cases, we have found that all local units within an enterprise have the same SIC(92) classification at subclass level. About 2,000 enterprises with employment of 100 or more, and with at least five local units have the same SIC(92) subclass for each local unit. In total about 40,000 out of the 65,000 multi-site enterprises have the same SIC(92) subclass for each local unit. ONS is aware that in some cases the presence of the same code does not mean that all local units do indeed have the same activity. Further work should be undertaken to evaluate whether better quality business description data can be collected more effectively for those businesses that appear to have the same industrial activity at all locations.

### *Other Design Features*

Finally, there are other features of the design and wording that could be improved. Most particularly, the questionnaire is crowded – it is generally accepted that more white space in documents makes navigation easier. Redesigning in this way should make the questionnaire easier to complete. We note that some constraints are applied to the design of the questionnaire because of the software ONS uses for printing and scanning questionnaires. However, such constraints must not be allowed to affect the quality of the IDBR.

It is important therefore, that the questionnaire for ARI is evaluated to determine in practice where its design is causing loss of quality. The questionnaire should be redesigned (and re-evaluated) in the light of those findings.

**Recommendation 7** *Questionnaire design for the Annual Register Inquiry should be given priority within ONS's proposed business-survey questionnaire design project. Among the particular topics that need to be addressed are omission of local units; delineation of business units; feedback of known industrial classification information; and efficient collection of business description information. (Priority: 1)*

**Recommendation 8** *ONS should continue to extend the present electronic data-collection arrangements for the Annual Register Inquiry. (Priority: 2)*

## **6.5 Updating from the Annual Register Inquiry**

### **6.5.1 Small Enterprises**

A particular problem identified by users relates to updates of the small (and micro) enterprises. When enterprises respond to the ARI, a local unit is created on the IDBR system with the information supplied by the business to the ARI. Once a local unit has been created, the administrative data play no further part in the updating of employment details on that reporting unit. That means that once the local unit is created, the enterprise will be inert unless included in a future ARI, and will become out of date. Inclusion in ARI is unlikely given the small sampling fractions in the strata containing the smallest businesses. This updating procedure has caused problems in cases where the turnover of an enterprise has increased markedly, and consequently the turnover and employment values are out of line.



This can lead to increases in variance in survey estimates that rely on the use of turnover as the auxiliary variable in ratio or regression estimation.

The method used for updating small enterprises needs to be reconsidered. On the one hand, if the ARI can improve on the information stored on the IDBR, it could be argued that such information ought to be used. However, we would want to consider when other data sources – HMCE for industrial classification information and IR for employment information – are allowed to take precedence over the ARI source.

The requirements of survey users differ from those of analytical users. For the former it is acceptable to have the Register correct on a global level – the correct distribution of businesses is the important factor. For analytical users, it is more important that the details are accurate on a finer level, which means in practice that the emphasis is on individual enterprises being correct. On balance, it is therefore sensible to update from businesses included in the ARI. After a pre-determined length of time (say two years to tie in with the NACE inertia rule on classifications) updating should be allowed from administrative sources. Once the quality of industrial classification is similar from all sources, it may not be necessary to conduct the ARI for small enterprises.

## **6.5.2 Medium and Large Enterprises**

For medium and large enterprises, the IDBR is heavily dependent on the ARI. Complete enumeration of the largest enterprises, in theory at least, gives a safeguard that structures, local unit employment and industrial classification can never be too out of date. However, the analysis reported above shows that some enterprises are updated infrequently. Given the important role of the ARI in maintaining the IDBR, and supporting the structure of the national accounts and other analyses, it is imperative that information relating to the largest businesses is kept up to date.

We discussed in section 6.4.2 the need to improve follow-up of non-respondents. If it is impossible to elicit a response from particular businesses, alternative updating mechanisms will be required. Development work should be undertaken to determine the best way to fill gaps arising because of non-response. This could require intervention by the Business Profiling Team, or reference to other sources.

## **6.6 Overlap with Business Profiling**

One of the clearest messages that came out of our research relates to the overlap between the work of the Business Profiling Team (BPT), and those responsible for running the ARI. Chapter 8 discusses in more detail issues relating to complex businesses. We are pleased to note that for the 2000 survey, a separate team was set up within the ARI processing team with sole responsibility for dealing with the BPT. Improved communication between the ARI team and the BPT has reduced the amount of correction work required, compared with the previous year.

Profiling and conducting the ARI for large, complex businesses are so closely linked that the same people should be responsible for the ARI as for profiling. In many cases, collection of ARI data is equivalent to updating the profile of an enterprise. This measure will help to

address some of the problems relating to the identification of the correct business unit on the questionnaire.

## **6.7 Census of Businesses**

The possibility of a census of businesses has been considered. The purpose of such a census (possibly operating on a rolling decennial basis to even out the workload) would be to improve the quality of the IDBR, particularly of local unit details for smaller enterprises. This would have consequences for small area statistics, particularly those of employment, which could be improved through the direct collection of data in the census.

Such a census would be a large-scale operation. The number of small enterprises not covered in the ARI's rolling four-year programme is around 1.5 million, that is about one fifth of total employment in the economy. To survey even one-tenth of these available enterprises annually would represent a three- to fourfold increase in the sample size of the present ARI. The extra sample would cover approximately 2 per cent of employment.

It is unclear what the quality improvement of such a census would be. First, approximately 250,000 business units cease trading and are removed from the IDBR each year. Most of these come from the smaller sizebands. The effect of this would be that, assuming a random rotating sample was selected, approximately half of small enterprises on the IDBR would be too new to have been included in a census of businesses. Second, we must consider data collection. The response rate for small businesses to official surveys is low relative to other sizebands, so it would be unlikely that we would get a complete response. Furthermore, ONS is committed to ensuring that small businesses are included in only one survey at once, with a guaranteed exemption from other surveys for a period of time afterwards. Businesses may also resent providing ONS with much of the information they have already supplied to the administrative sources. Moreover, Recommendation 3, which relates to the provision of data from the administrative departments, should ensure that the quality of the smallest enterprises on the IDBR improves.

## **Chapter 7**

### **Updating Statistical Information on the IDBR from Survey Sources**

#### **7.1 Introduction and Summary of Main Quality Issues**

In this chapter, we draw together a range of topics relating to the principles and processes involved in updating statistical information on the IDBR. We make specific reference to the updating of industrial classification information, but the general principles apply to other variables too.

The need to meet a wide range of users' demands has resulted in a mechanism for updating the IDBR that, while ensuring a limited range of reporting units have a more up-to-date and possibly more accurate classification, is biased as a whole and leads to inconsistency. A set of updating rules was agreed with survey statisticians to improve sample stability for short-period surveys, and these rules should be enforced more rigorously.

#### **7.2 Updating the IDBR with Information from Statistical Surveys**

In this section, we document the procedures for feeding back information from sources that aren't statistically independent of ONS surveys. By statistically independent, we mean that the probability of a reporting unit being updated does not depend on its inclusion in a particular sample. Historically, some ONS sources have been considered of higher quality than the administrative sources, and the Register has been updated from such sources on the premise that its quality will be improved.

The main source used to feedback information to the IDBR is ONS's PRODCOM (PROducts of the European COMmunity) survey. PRODCOM collects sales information relating to a fine breakdown of products, and this information can be used to determine industrial classification. In addition, there is updating from the short-period surveys that collect turnover and employment information when businesses tell ONS that its classification is wrong. There is also some limited updating from retail commodity breakdowns in the ABI. Other ONS surveys contribute rather fewer updates. Updates to the IDBR are either ongoing, or made in batch every year.

#### **7.3 Effects of Feedback from Statistical Surveys**

Feedback from statistically-dependent sources biases the Register and introduces inconsistency even though the update to a particular business might provide correct information for that business. Sample estimators are biased if auxiliary information associated with members of the sample and the survey population as a whole aren't comparable. That bias should be recognised when selecting samples, producing estimates and analysing the IDBR directly. Inconsistency arises through the use of data that have been collected or produced using different methods. Improving the quality of individual reporting units does not necessarily improve the quality of the Register overall.

There are two cases of feedback from surveys to consider: surveys that cover the whole population, and those that cover only specific domains within a population.

### **7.3.1 Feedback from Surveys Covering the Whole Population**

#### *7.3.1.1 Bias*

Feeding back from surveys covering the whole population can bias the Register and this can be exacerbated by the use of a rotational sampling scheme such as that used by ONS.

To see clearly the effect of feedback to the Register, consider the example of deaths. Suppose that some businesses recorded as being live are, in fact, dead. There is a delay before ONS is notified of the deaths. Now, suppose a survey updates the reporting units on the IDBR that it finds are dead. A subsequent sample containing the same, or largely the same, sample members will contain fewer such deaths than the population as a whole, so population totals will be overestimated.

Similarly, the feedback of industrial classification or size information used in sampling and estimation can cause bias, particularly in those areas where there is systematic movement from one industry or sizeband to another.

#### *7.3.1.2 Inconsistency*

Feeding back from surveys introduces inconsistency into the Business Register. The NACE Rev 1 Regulation specifies that industrial activity should be determined according to the value added by the various activities within the business. If value added is unavailable, the industrial classification can be determined using an alternative input or output variable. Inconsistency arises in ONS because different methods and data items are used by the different ONS sources that determine industrial classification. In addition to a choice of variable, the level at which the classification is deduced also varies: for example, the ARI uses employees at individual local units, PRODCOM uses the value of sales of products.

### **7.3.2 Feedback from Surveys Covering Only Part of a Population**

The main concern with feeding back industrial classification information from such surveys is that businesses can only move one way – from within the particular survey population to outside it. There is no mechanism for movement the opposite way. This means that the survey population under such an updating scheme can only get smaller, and this may lead to underestimation of parameter estimates.

We examined the update of industrial classification from the PRODCOM sample that took place in early 2000, and found that about 3 per cent of the reporting units in that sample were reclassified from a production industry to a non-production industry. Table 11 gives more details.

**Table 11 IDBR Reclassifications**

	<b>Number of reporting units</b>	<b>Number changing</b>	<b>Proportion changing</b>
PRODCOM sample	21,827	1,670	7.7
<i>Moving from production to non-production</i>		659	3.0
PRODCOM population	137,047	519	0.4
Rest of IDBR population	1,489,074	11,391	0.8

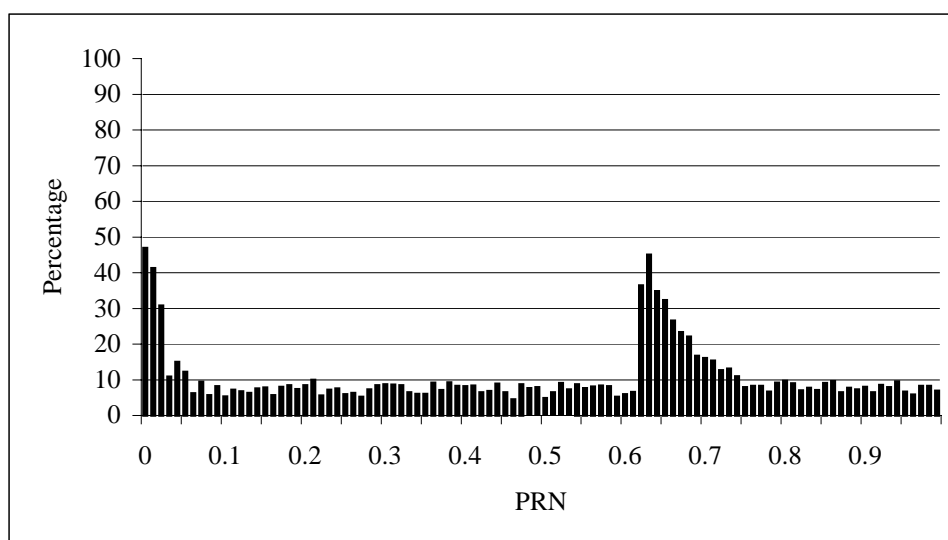
*Source: IDBR, January and July 2000*

As we have already discussed, the effect of such an updating pattern is to gradually reduce the size of the production sector as shown by the IDBR. The effect of one year’s update is to reduce the number of reporting units in the production sector by about 0.5 per cent. Most of the reclassifications come from businesses with fewer than 20 people employed, so the effect will be to reduce the amount of activity shown by the production sector by less than the 0.5 per cent that is measured by the number of reporting units.

#### 7.4 Impact of Updating from Statistical Sources

We found that the extent of feedback to the Register varies enormously along the PRN line. We examined the 33 strata of reporting units with 0–9 people employed at the SIC(92) class level, as each of these had more than 10,000 reporting units in the stratum and were therefore large enough to support our analysis. In 26 of the strata, the proportion of survey updates across most of the PRN line was around 10 per cent, with a peak or peaks in areas of the PRN line reflecting typically 30 to 60 per cent updating from surveys. A typical graph for SIC(92) class 52.48 is shown below.

**Graph 1 Survey-sourced IDBR Updates, SIC(92) class 52.48**



*Source: IDBR, June 2000*

The peaks are located on the PRN line in the place where samples have been selected in recent years. The low percentage of updates across most of the line represents reporting units that have been updated from other strata sampling in different parts of the line. This happens more in the strata of larger businesses, where sampling fractions are higher.

The graph illustrates that the incidence of survey feedback varies greatly according to the PRN range. We therefore expect surveys selecting samples from these strata to be at risk of producing biased estimates, since a sample selected from a part of the PRN line with few survey updates will have different characteristics from a sample selected from a part of the PRN line with many updates. In either case, the auxiliary information for the sample won't be comparable with that relating to the survey population as a whole (see section 7.3).

## 7.5 Quality of Updates from Statistical Sources

In this section, we describe our attempts to evaluate the quality of the various sources used for updating the IDBR. This is difficult as we don't have an absolute measure of correctness and, as discussed earlier, there can be more than one version of the correct answer.

We found that the quality of industrial classification from most of the main sources was similar. The following table sets out the proportion of industrial classification codes that were found to be different when the named source was compared with ARI. The data were taken from the ARI for 1999.

**Table 12** Number of Industrial Classification Updates to the IDBR by Source

Order of priority	Source	Number of updates	Number different	Proportion different (%)
1	CBU	393	53	13
	BST	110	32	29
2	PRODCOM	7,911	886	11
	RSI	764	82	11
	Financial Inquiries	33	11	33
3	AES	102	13	13
4	MIDSS/QIDSS	6,150	668	11
	ABI/ACOP/DSI	3,823	331	9
5	STES	1,489	147	10
	Other ONS surveys	428	105	25
6	DETR	719	135	19
7	Proving	969	107	11
	IDBR Maintenance	2,137	379	18
8	HMCE	28,849	4,920	17
9	IR	3,342	1,215	36

*Source: ARI 1999*

The table shows that for most surveys between 10 and 15 per cent of classifications differ from those on the ARI. It should be noted that there is some subjectivity in coding business descriptions – given the standard of business descriptions typically available to ONS,

previous work has shown that two expert coders would classify the same way in about 85 per cent of cases. This suggests that it would be hard to improve on a 10 to 15 per cent error rate unless a different mechanism were to be introduced for collecting business description data. The table also reflects the findings from the RQS that the classification data from the administrative sources, particularly the PAYE data from IR, are of relatively low quality. In the case of PAYE, over one-third of reporting units have a different classification at the subclass level.

The reasons for differences between ARI classifications and those from the Complex Business Unit (CBU) and the Business Structures Team (BST) require further investigation. In the case of the latter, almost one-third of classifications (albeit from a relatively small sample) were found to differ between the BST classification and that deduced by the ARI.

## **7.6 Discussion on Updating from Statistical Surveys**

The main points of the discussions in this chapter so far have been that:

- updating the Register from statistically-dependent sources, while possibly improving the quality of individual units, can lead to bias and inconsistency;
- updating from surveys covering only part of a population reduces the size of that sub-population over time;
- different parts of the PRN line have different levels of updating from statistical surveys;
- in theory, PRODCOM has the best method of classifying enterprises because it collects a fine breakdown of data; and that
- there is no clear evidence that any particular source is better than any other.

Evidence from ONS confirms the theoretical principle that updates should not be made from statistically dependent sources. That would suggest that updates are made solely from the ARI, profiling and the administrative sources, which are independent of our survey samples.

However, good practice used in PRODCOM and other sources needs to be considered. Such good practice should be adopted in ARI and the other main sources to ensure that the benefits are felt across the range of IDBR units, not just the relatively small number that happens to be selected for the individual survey.

We have found that the quality of information provided on ARI forms is sometimes quite low, often because business descriptions are too vague to be classified accurately. In such cases, additional information that is available from other sources may be able to inform the classification process. In one example, we saw the Precision Data Coder (PDC), ONS's electronic coding software, suggest three equally likely alternatives based on an ARI business description. Without additional information, the online coder had to make what was essentially a random choice of code, whereas information available from PRODCOM showed which of the three codes dominated product sales. Use of this information could have resulted in a correct code. We endorse the use of PRODCOM and other statistical data in such situations to resolve conflicts between the main sources, and to add an extra dimension where the information from the main source is too vague. Such an approach could be adopted on a wide scale if data from various sources were linked by way of each reporting unit's unique reference number. This approach would entail the feedback of PRODCOM classifications on ARI forms, which we discussed in section 6.4.3.2.

**Recommendation 9** *The Annual Register Inquiry, Business Profiling Team and the administrative sources should have highest priority for updating IDBR auxiliary information. In the case of industrial classification, this depends on improvements to the current level of quality of information from these sources. Other sources should be used by the Business Registers Unit to resolve uncertainty and ambiguity. (Priority: 1)*

## **7.7 Determining Industrial Classification from Business Descriptions**

Classifying business descriptions into specific SIC(92) codes is fundamental to the IDBR and ONS's economic statistics, and underpins the discussion in the previous section.

Unfortunately, coding business descriptions is difficult to do accurately. Tests have shown that two expert coders classifying the same businesses would only get on average about 85 per cent agreement of codes given the quality of data that is typical to ONS at present. Part of the reason for the differences between codes allocated by human coders is the subjective interpretation of the activity description. The use of electronic coding systems removes that subjectivity and is a more consistent solution to problems encountered in activity coding.

The standard package used by ONS since the mid-1990s to code business descriptions has been the Precision Data Coder (PDC). This is used for both activity coding, and for product coding in PRODCOM. The package can be run in either batch mode or interactively. In general, batch mode is used for bulk processing, such as the take-on of data from ARI. Interactive mode is used when coding small numbers of business units at the desk, and to resolve queries generated by batch processing.

In both cases, business descriptions are entered into PDC. PDC offers a range of possible SIC(92) codes, along with an indicator of 'confidence' in the accuracy of each suggested code. In batch mode for the ARI, the code is automatically created where a confidence level threshold is reached, and the PDC does not offer any viable alternatives. If there is serious doubt about the correct code, the PDC will refer the description for clerical intervention. In interactive mode, users choose what they consider the most appropriate code from a list of possible alternatives (ranked in order of confidence). Note that a user may choose a code not suggested by PDC.

This review found that some members of staff do not use PDC. The reason cited is that it cannot be trusted as it sometimes offers codes that are clearly wrong. Here there is potential for error and inconsistency on the IDBR. In trials, the PDC was shown to be at least as good as human coders. In addition, the PDC is upgraded annually to take account of any anomalies that have arisen during the year, and any changes to both the classification system and the business world. Given that codes suggested by the PDC can be rejected by the user, there is no reason not to use the PDC, as at the very least it provides a wider range of information to help classify a business.

The PDC was developed by Inference Group for the 1993 Census of Employment to allow automated coding of business descriptions captured from the census forms and for the IDBR development project. The product has changed little since then and, while it was the only product of its kind when it was first came on the market, there are now other products. It needs an upgrade from 16-bit to 32-bit architecture to comply with ONS operating standards. Subject to setting up contracts, the supplier has recently confirmed that this change can be



implemented, which will extend its life. ONS is currently exploring through contact with other NSIs and through the CLAMOUR project whether a longer-term replacement is desirable.

**Recommendation 10** *The agreed electronic tool should be used to aid classifying business descriptions. All staff members should feed back information on deficiencies to those controlling the system to improve the quality of the coding tool. The use of expertise from outside ONS (for example, industry groups) should be considered as a way of improving coding quality. ONS should review the Precision Data Coder alongside other similar products to ensure that the most effective tool for the job is used. (Priority: 2)*

## **7.8 Frozen and Current Classification**

### **7.8.1 Background**

A recent development in ONS, in response to the Kokic and Brewer review (see Chapter 12), was to introduce frozen variables for SIC(92) and employment to the IDBR. Before that, industrial classification codes were always updated immediately. However, this resulted in instability of sample membership – reporting units would move in and out of samples because of reclassification, rather than because of sample rotation. To address this, two variables were created. In addition to the existing ‘current’ data which would continue to be updated dynamically, 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. 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 to permit questionnaires appropriate to the new industry to be sent.

We mentioned above that in exceptional circumstances the frozen variables can be updated during the year. An update to a current field is also updated on the equivalent frozen field where the reporting unit’s employment (before or after the update) is 1,000 or more, and the unit has changed SIC(92) division. In addition, single-sourced unproven enterprises with employment of 20 or more can also have the frozen field updated. (Note that such enterprises are not included in survey populations, so there is no problem of sample instability as a result of the change.) In addition, the frozen variables are allowed to change where there is a restructuring of the business.

### **7.8.2 Extent to which the Rules are Being Followed**

We found that between January and December 1999, about 5,300 reporting units had one or more frozen variables updated when the rules for so doing were not met. Only 118 reporting units were updated according to the rules.

We found that most of the changes arose from BRU where a monthly exercise is run to improve the quality of industrial classification for new businesses in certain industries that are difficult to classify. These industries include public houses, utility industries, steel makers and finance companies. It’s often possible to tell from a business’ name whether it’s been

classified correctly. Where the industrial classification for a new enterprise appears to be incorrect, both the current and frozen SIC(92) variables will be updated. In addition, the frozen field is also updated where a change has occurred to a current field from another source. Larger businesses – those with employment of at least 2,000 – are usually classified correctly. Any changes arising from this exercise are dealt with by staff within the BRU and are completed quickly.

In conclusion, we have found that few of the updates to the frozen variables from survey sources do not meet the criteria for updating. This suggests the present rules are working well, but could be more rigorously applied. However, many updates are made from BRU and provision should be made in the rules for such updates.

**Recommendation 11** *The rules for updating the industrial classification and employment variables, set up to maintain stability of sample membership for short-period inquiries, should be reviewed to see if they could better accommodate register maintenance requirements and should be applied rigorously. (Priority: 3)*

## **Chapter 8**

### **Complex Businesses**

#### **8.1 Introduction and Summary of Main Quality Issues**

In this chapter, we consider complex businesses. We distinguish between two types of business. First, we look at large or complex businesses whose structure must be clearly understood to facilitate data collection. Then we address those businesses where an alternative reporting structure has been set up for use in individual surveys.

ONS has identified the importance of ensuring good quality structural information for the largest, most complex businesses, and has set up an operational unit to profile those businesses. In addition, the unit has taken over responsibility for reactive profiling of smaller businesses. The unit responsible for profiling has recently come under BRU's management.

The plan for a profiling unit was not implemented fully in the way that was described in CSO's Population Units report, which recommended the unit's foundation. There have also been several changes in the level of resources available to the unit. Consequently, several years on, many of the largest businesses have yet to be profiled, and ONS is still some way from being able to establish a programme of repeat profiling. As a result, ONS now lags behind Canada, New Zealand, Australia and the Netherlands, each of which has classed profiling as a priority. In addition, the database recording the results of profiling doesn't communicate Register changes to survey managers effectively.

The maintenance of reporting structures relevant to only a single survey places an undue burden on the BRU's resources, and runs a risk of lower quality outputs through the potential omission and duplication of economic activity.

#### **8.2 Complex Businesses**

A few businesses, comprising a disproportionate amount of economic activity, are complex. It is important for ONS to ensure that it delineates these businesses accurately, and collects survey data in the most helpful way for statistical purposes while minimising the burden on businesses. It is also important to ensure that coverage is complete with no double counting, and that the activities of the various parts of the business are allocated to the appropriate domain of analysis. Particular concerns about the collection of data relating to significant secondary activities within businesses provided the impetus behind a 1994 CSO review of the treatment of complex businesses – entitled *Population Units in the Major Business Surveys*.<sup>14</sup>

##### **8.2.1 Population Units Study and History of the Complex Business Unit**

The Population Units study identified the crucial importance of profiling large, complex businesses. The study recommended the foundation of a unit to agree the decomposition of enterprises into Kind of Activity Units (KAUs) based on the contribution of the secondary activities to the particular industries. The detailed recommendations of the report are:

‘We recommend that these enterprises are examined, and a complete and non-overlapping set of KAUs (for both production and services) agreed for the whole enterprise.

We recommend that an enterprise should be split into kind of activity (KAU) statistical units if its secondary activity would make a significant contribution to the total for that industry, or if it distorts the primary industry data.

In order to carry through the above priorities, we strongly recommend setting up a dedicated unit – the Large Business Unit.’

The Complex Business Unit (CBU) was set up to address the recommendations in the report mentioned above. The CBU carried out proactive profiling (that is, profiling initiated by ONS) by visiting businesses to establish the correct structure and best reporting arrangements. In addition to the proactive profiling, a smaller unit – the Business Structures Team (BST) – was set up within BRU to respond to structural queries arising from survey sections and elsewhere. This is referred to as reactive profiling, and usually takes place where there is a complex or large VAT group structure. Reactive profiling mainly takes place at the desk, by telephone or e-mail.

Initially, the CBU was separate from BRU, but in 2000 both CBU and BST were brought together within BRU under the name Business Profiling Team (BPT).

### **8.2.2 Responsibilities of Business Profiling Team**

At present, the BPT’s responsibilities are to profile businesses, either by visiting them or, for the smaller businesses, by desk profiling, which may involve contacting businesses by phone, letter or e-mail. BPT is also responsible for updating and maintaining complex business structures on the IDBR. For reactive profiling, BPT retains responsibility for maintaining enterprises for only a short time after the profile is established.

### **8.2.3 Remit of Business Profiling Team**

The remit of the CBU has changed considerably since it was set up. The requirements for the CBU were clearly spelt out in the Population Units report. However, during the late 1990s when budgets were tight, profiling was seen as a non-essential activity. Consequently, the resources available to the CBU fluctuated according to staffing requirements elsewhere.

BST was established to ‘mop up’ problems that were beyond the CBU’s scope. Problems typically arose from BRU updates such as those involving large or complex VAT unit structures, and enterprises where the PAYE and employment values appeared to be inconsistent. Profiling is reactive at present, but it is intended to move towards the profiling of categories of enterprises that can be identified by their complexity, VAT or PAYE structure.

We discussed in section 6.6 the overlap between the work of BST and the ARI, and suggest that profiling and ARI work for complex businesses be brought together within the same operational unit.

Profiling is important for several reasons. First, it establishes contacts within the largest businesses. This helps to ensure a better response rate to surveys and a better quality response. Second, profiling businesses helps us to ensure that any questionnaires sent out are directed to the right part of the business. This should reduce the risk of omission or double counting. Large and complex businesses make up a large proportion of economic activity in relation to their number, and the importance of ensuring the correct structure both for IDBR maintenance and survey operations cannot be stressed enough.

Most of our observations relate to the scope of enterprises available for proactive profiling. The Population Units report suggested that the scope of enterprises would be those with ‘100 employees in a secondary activity, but this could be tightened in particular industries to meet the specific needs of users, or relaxed in others for reasons of economy. Further work is needed to establish what the ‘significant contribution’ would be in each industry.’

Currently, enterprises with employment of 2,000 or more are profiled, irrespective of secondary activity or complexity. However, at the time of our analysis, there were over 800 enterprises on the IDBR with employment of 2,000 or more that had not been profiled.

The table below shows those enterprises with at least 2,000 people employed that haven’t been profiled and are reporting as an enterprise. Here, we define the primary activity of each enterprise as the SIC(92) division that has the most employment. All activity in other SIC(92) divisions is considered secondary activity.

**Table 13      Number of Enterprises by Proportion of Secondary Activity and Size**

Secondary activity percentage	Enterprise size (employment)				Total
	2000–4999	5000–9999	10000–24999	25000 or more	
0–9	454	102	26	11	593
10–19	16	6	1	0	23
20–29	10	5	3	3	21
30–39	10	12	12	4	38
40–49	21	22	17	2	62
50–59	24	28	19	3	74
60–69	6	25	8	0	39
70–79	2	1	2	1	6
Total	543	201	88	24	856

*Source: IDBR, October 2000*

The 856 enterprises account for employment of 5.4 million. Of this, 4.3 million is covered by the primary activities of these enterprises, while 1.1 million is due to the secondary activity and thus wrongly classified. This represents one-fifth of the total employment of these enterprises and approximately 4 per cent of employment on the whole IDBR. This is activity that is misclassified by being included in a large reporting unit with a different industrial classification. This analysis stresses the need to profile the largest businesses that have significant secondary activity.

Repeat profiling is something that ONS does not yet do. It is important to ensure that the structures of the largest businesses are kept up-to-date as they are likely to evolve or change dramatically from the time of first profiling. Part of BPT's resource should be allocated to this.

**Recommendation 12** *A business profiling strategy should be agreed based on two objective measures: the level of secondary activity; and the impact on survey estimates and analyses. A solution should be developed to bring profiling and ARI operations together for the largest businesses. A trigger mechanism should be established for sub-annual profiling of the most complex enterprises. (Priority: 1)*

Profiling can involve visiting businesses in order to establish their structures, and to establish the best way for businesses to respond to our surveys. Initially, visits were made by two members of BPT. More recently, however, following an efficiency review of the CBU, visits are now being made by only one person. We identify this as a threat to the quality of the IDBR and consequently also to economic statistics. Meetings with businesses involve discussing their structure and making notes for future reference. It is difficult for one person to do both these jobs, and it was reported to us that the profilers felt under increasing pressure, made sparser notes and were more likely to miss points from the discussion. It may be possible to counteract this risk by ensuring that visits to the largest or most complex businesses are undertaken by more than one member of ONS's staff. Conversely, we also recognise that in some simple cases, it may not be necessary to visit the business at all – the profiling can be done with equal quality electronically, by post or by phone.

In addition, the range of information relating to business structures that's available on the Internet should be evaluated in relation to its use in profiling. In particular, the Net offers scope for easy identification and maintenance of business structures and potential for data collection for profiling and register maintenance.

Users commented that they felt inadequately consulted or informed about BPT operations. A particular concern was that survey sections are not aware of the details of the profiling programme, nor the timetable for updates. They have often found changes to their samples and results because of restructuring. Identifying this in advance would make planning easier for IDBR users. The information is available in an ONS database, whose existence should be better publicised. In addition, DETINI should be consulted during the process of profiling enterprises with activity in both Britain and Northern Ireland.

**Recommendation 13** *The Business Profiling Team (BPT) should initially establish the correct structures of the businesses at the desk. For complex businesses, an agreed programme of profiling visits should be put in place, and this should be made available to survey managers in advance. The BPT should ensure that data collection and survey results managers are informed, in advance, of its updates to business structures on the IDBR. (Priority: 3)*

## 8.2.4 International Comparisons

During the early to mid-1990s, other NSIs also began to appreciate the importance of ensuring the correct structures of complex businesses, and the impact this would have on both the quality of published statistics and respondent burden. Consequently, profiling units were also set up in other countries. The units now established in Statistics Canada, Statistics New Zealand, Statistics Netherlands and the Australian Bureau of Statistics are considered to be among the best in the world, and we look to them for international benchmarks of best practice.

Profiling of the largest and most complex businesses is common practice everywhere. The rules for determining which businesses are in scope vary, but are typically based on some combination of size, complexity and contribution to overall statistics. In many cases, the most complex businesses are re-profiled every year, and in some cases the very largest and most complex enterprises are profiled more often. In Statistics Netherlands, for example, some of the most complex and volatile businesses are visited up to ten times a year.

The attachments in Annex 11 are extracted from the progress reports presented to the recent Roundtable by Statistics Canada and the ABS. In addition, several other countries reported that they are about to set up profiling units, or to expand their current profiling operations.

## 8.2.5 Other Government Departments

The Business Competitiveness Group (BCG) of DTI is interested in understanding the structure of the businesses it deals with. In certain well-defined circumstances, the IDBR may be of assistance to them. Conversely, information held by BCG about the businesses it has contacted could possibly supplement the BPT's knowledge.

**Recommendation 14** *To reduce the burden on business and improve coherence across National Statistics, ONS should explore with other government departments opportunities, within the requirements for maintaining data confidentiality, to co-ordinate the collection of data relating to the structure of complex enterprises. (Priority: 3)*

## 8.3 Survey Specific Reporting Arrangements

In some cases, surveys have non-standard reporting requirements. In such cases, Inquiry Specific Reporting Units (ISRUs) are set up for individual surveys.

First, some enterprises have activity in both Britain and Northern Ireland. Legal restrictions mean that data from NI and GB must be collected separately, and ISRUs are set up for NI activity contained in UK-wide enterprises. These are maintained automatically by the IDBR.

Second, ISRUs are created for use by individual surveys when the standard reporting unit does not provide the correct coverage or contact details for collecting data in a particular survey. ISRUs have been widely used for creating units that report for only part of an enterprise.

We found that in June 2000, almost 5,000 ISRUs existed on the IDBR, although this was fewer than in previous years. We have split the ISRUs into those belonging to ONS surveys, and those belonging to OGDs' surveys that are run by ONS. The table below shows a breakdown by survey ownership.

**Table 14**      **Distribution of ISRUs by Survey**

<b>Ownership of survey</b>	<b>Number of ISRUs</b>
ONS	1,853
Non-ONS	2,916

*Source: IDBR, June 2000*

We found that most surveys have only a few ISRUs. However, there are a few surveys with many ISRUs present. These are shown in Table 15.

**Table 15**      **Distribution of ISRUs for ONS Surveys**

<b>ONS Survey</b>	<b>Number</b>
Producer Price Index (PPI)	451
Monthly Wages and Salaries Survey (MWSS)	206
Export Prices Index	132
Corporate Services Price Index	59
Import Price Index	42
All other ONS surveys	963
<b>Total</b>	<b>1,853</b>

*Source: IDBR, June 2000*

The most prolific users of ISRUs within the ONS are the PPI and MWSS. The PPI sample is drawn from the PRODCOM sample. PPI uses ISRUs because in some cases it is necessary to send two or three separate questionnaires to different contact points within the same reporting unit. Over half of PPI's ISRUs are included in its sample.

In June 2000, the MWSS sampled approximately 8,300 reporting units from a population of 96,000. Of the 206 ISRUs for the MWSS, 47 were in the sample and 144 were in the population, with the remaining 15 not eligible for inclusion in the population.

Many non-ONS surveys also have ISRUs. Most are set up for the Energy and Mineral surveys that ONS conducts on behalf of DETR. In these cases, it is understood that the relevant businesses cannot report according to the standard business structure on the IDBR.

In general, most of the ISRUs for the non-ONS surveys were included in the relevant samples. Some however, were not, and in some cases we found that many of the ISRUs set up for particular surveys did not even feature in survey populations. This could be because the changing structure of an enterprise has caused the ISRU to become out of scope of the



survey population. In any case, it appears to be an inappropriate use of resources to create reporting units that are not being used. Survey sections themselves are responsible for weeding redundant ISRUs, but this does not appear to happen in all cases.

We found that survey sections (particularly Data Validation Branches) didn't fully appreciate the need for, and operation of, ISRUs. A particular problem is that survey sections receive certain information relating to the enterprise (such as new contact details), but only update the information on the ISRUs and associated local units. This can lead to confusion and has led to extra work in BRU to rectify problems. In some cases, it isn't clear whether ISRUs are the answer to the reporting problem. A better solution might be to change the information that is stored on the IDBR alongside primary reporting units, or alter the way that information relating to the scope of the return is printed on questionnaires. In some cases, ISRUs have been created with exactly the same structure as an existing primary reporting unit.

Finally, as ISRUs are set up and maintained by survey sections independently of each other, there could be incongruity between surveys. Undercoverage and duplication could also result.

Overall, it was clear to us that the setting up of ISRUs was not always the correct answer to a particular problem. We found that survey sections were not well educated in the use of non-standard reporting arrangements or how to maintain ISRUs. Consequently, a great deal of BRU's resources were spent sorting out problems that should have been kept within the confines of surveys. The impact on register quality is that BRU resources are being spent maintaining structures that benefit only single surveys. These resources would be better spent improving the quality of a part of the Register that is used across the board. Unfortunately, there wasn't enough time during this review to examine in detail whether all ISRUs are justified, nor to propose an alternative mechanism for dealing with them.

**Recommendation 15** *A detailed review of the requirements for survey-specific reporting arrangements should be carried out with the aim of standardising reporting arrangements for all surveys. Where non-standard arrangements are necessary, the procedures for managing them should be simplified. (Priority: 2)*

## **Chapter 9 Coverage**

### **9.1 Introduction and Summary of Main Quality Issues**

In this chapter, we look at some aspects of coverage that haven't been addressed elsewhere in the report. We address two problems: undercoverage due to non-registration of the smallest businesses; and potential duplication arising from the use of two independent sources to maintain the Register.

The IDBR's coverage is very good, representing around 99 per cent of economic activity. This rivals any business register in the world. There is a well-used method of measuring the extent of overcoverage, although that could benefit from being updated. Duplication (arising through the dual sourcing of the IDBR) is a potential problem. Possible duplicates are excluded from the IDBR as a precautionary measure. Following extensive matching work, the level of potential duplication has been reduced, and the matching procedures have been improved.

### **9.2 Undercoverage due to Non-Registration of Small Businesses**

At present, the IDBR covers approximately 99 per cent of all economic activity in the UK. Both the Business Registers Regulation and the European System of Accounts (ESA) require complete coverage. The UK is in the process of presenting to Eurostat its methods and sources for the National Accounts on an ESA basis. One of the many aspects that Eurostat will review is the exhaustive coverage of the Gross National Income estimates (including adjustments for the hidden economy, the illegal economy and the incompleteness of the IDBR). If the UK doesn't meet all the criteria then Eurostat will impose a Reservation on the UK's accounts. This will in turn mean further work to address the issues raised. As far as the Business Registers Regulation is concerned, omission of the smallest businesses has been discussed fully by the Eurostat Business Registers Committee, which agreed that the coverage was consistent with the Regulation on business registers. Furthermore, the Committee agreed that coverage was considered sufficient if all enterprises where the equivalent of at least 0.5 of a person was employed were included.

The IDBR is built up from traders who are registered for VAT, and employers who operate PAYE Income Tax schemes. Therefore, the businesses legitimately not covered are those that neither run a PAYE scheme for employees, nor have turnover that is above the threshold required to register for VAT. These will be primarily self-employed proprietors and partnerships. Some non-VAT, non-PAYE units are included for specific surveys, usually in the financial services industries. Such enterprises are corporate bodies registered at Companies House. It would be possible to include on IDBR all organisations of this type but most are dormant and would not add to the quality of the Register. Other data sources are available to cover self-employed individuals. Legal restrictions are in place which prevent the transfer of individual data from IR to ONS, but it is possible to receive aggregate data.

### 9.2.1 Estimating Undercoverage

DTI's Small Business Service estimates of the number of businesses and level of activity that are missing from the IDBR.<sup>15</sup> Here, we present a summary of the method.

The first step is to count the number of working proprietors on the IDBR. This is compared with the self-employed stock estimated from the Labour Force Survey (LFS) and cross-checked with data from IR's Survey of Personal Incomes. It is then possible to estimate the number of unregistered working proprietors.

Then, the number of unregistered working proprietors is converted into an estimate of the number of businesses using information about the proportion of unregistered businesses that are either sole proprietors or partnerships. In doing so, it is assumed that all unregistered businesses have no employees. Turnover is estimated using scaled down turnover-per-head figures stored on the IDBR.

Further adjustment is needed in some areas, for example where the IDBR is showing more working proprietors than the self-employed stock from the LFS. It is reported that problems are getting more common, possibly because self-employment is not such a clear-cut and easily obtainable tax status as it once was.

### 9.2.2 Results

It is estimated that up to 2 million self-employed people are not recorded on the IDBR. This represents approximately half of all business units, and about 7 per cent of total employment. Overall, these people are estimated to contribute about 1 per cent of economic activity. The detailed breakdown of this analysis by SIC(92) division is included in Annex 12.

It is difficult to see how the costs of receiving, updating and maintaining the data would be outweighed by any benefits of having what might be seen as full coverage. However, certain industrial sectors contain a higher proportion of small businesses than others. Construction, forestry, fishing and certain service sector industries are cases in point. In each of these, unregistered businesses contribute at least 5 per cent of activity to the appropriate industry totals. It would clearly be more beneficial to these sectors if they were to have more comprehensive coverage, although it wouldn't necessarily be cost effective to do so. The sectors concerned generally make small contributions to the total economy.

Instead, it seems more prudent to put resources into developing the method presently used to estimate the extent of this undercoverage on the IDBR. The present method, although it produces plausible figures, is becoming less usable as working practices within the economy change, particularly given less clarity regarding the various employment statuses. Improvement of the method is less important for measuring the quality of the IDBR than for other uses such as measuring the extent of undercoverage in the National Accounts.

**Recommendation 16** *ONS should work with the Department of Trade and Industry to improve the method of estimating the undercoverage that arises from businesses not required to register with HM Customs and Excise and Inland Revenue. (Priority: 3)*

### **9.3 Duplication and Record Linking**

One of the major difficulties in maintaining the IDBR is ensuring minimal duplication arises through the use of the two main administrative data sources. At present, the two administrative sources operate independently, and there is no way of cross-referencing the two sources. Therefore, matching must take place at ONS to minimise the chance of overstating the size of the survey population. In addition, it is important for the IDBR to be able to identify the extent of duplication. This section deals with those two aspects.

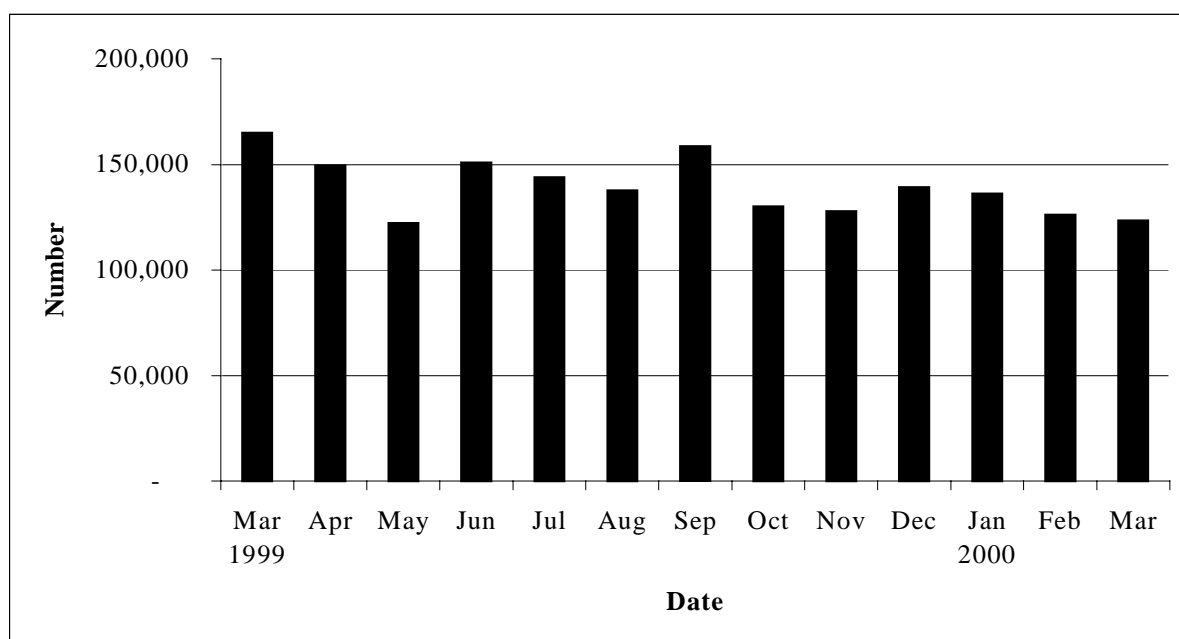
#### **9.3.1 Identifying Duplication**

Matching is undertaken using software written by Search Software America, called SSANAME3 and SSA-EXTENSIONS. The matching process involves creating 'namekey' codes from the names supplied by the administrative departments, and those already stored on the IDBR. Input names, addresses and postcodes are matched against existing IDBR records, resulting in a percentage score representing the quality of the match. For high scores, the match is considered definite, and the VAT and PAYE units are linked together automatically. Intermediate scores are considered possible matches, and are sent for further clerical investigation in order to decide whether there is a match. A low score indicates a non-match.

Since the IDBR's inception, it has been impossible to match all new registrations. The process of determining whether possible matches are genuine or not is labour intensive. As a precaution, unmatched PAYE units have been stored away from the main IDBR because they are expected to represent duplicates of VAT units already present on the Register.

Work has been undertaken in ONS since 1999 to improve the matching procedures, and to reduce the number of unmatched PAYE units. At the start of the year's work in March 1999, there were 160,000 unmatched PAYE units on the IDBR. At the end of the year, about 120,000 remained. The number of enterprises matched during the year was about 100,000 with 60,000 new unmatched PAYE units arising through the year from the quarterly PAYE update. The graph below shows the monthly progress throughout the year.

**Graph 2 Unmatched PAYE-only Units on the IDBR**



*Source: Pilot Study on the Improvement of Business Registers for Statistical Purposes, ONS 2000*

Further investigative work has also been done based on 1,500 previously unmatched enterprises that were included in the ARI for 1999. The purpose of this work was to try to match VAT and PAYE units wherever possible using the information returned to ARI. ONS managed to match almost half the survey responses where a match could be expected.

During the course of the various elements of the work described above, a range of possible improvements to the matching process were identified, including:

- development of processes for determining namekeys;
- use of trading names in matching;
- use of a unique address identifier; and
- use of address-cleaning software to improve address matching.

### **9.3.2 Treatment of Duplication**

Since plans to integrate the AES and the structural surveys of the production, construction and services sectors into the ABI, much consideration has been given to ensuring that the surveys are measuring parameters relating to the same survey population. At first, there were differences in population definition as the AES and other employer surveys used a PAYE-based register for sample selection and estimation, and the other structural surveys first used the VAT-based register and then the IDBR. Therefore, PAYE-only-based enterprises, but not enterprises based only on VAT, were included in the employer surveys. Structural surveys included VAT-only-based enterprises, but not those sourced solely from PAYE units. The populations were thus different.

In ONS, multiplicative adjustment factors have been proposed. The factors represented the probability of a reporting unit being correctly unmatched. The adjustment factors were based on the RQS. They were calculated by sizeband and broad industry banding. The factors were

then applied to estimates derived from the PAYE-only and VAT-only strata. The inclusion of the factors in the estimation procedure results in the removal of that part of the estimate relating to business units that were believed to be duplicates.

Recently, ONS has run a project to reconcile differences between employment estimates from the ABI and the AES. Work as part of this project showed that the application of the present adjustment factors to the population used by AES gave an equivalent (to within 0.01 per cent) population to that being used by ABI.

The use of adjustment factors in general should be seen as a stopgap measure. The quality of adjustment factors is uncertain, and their application is based on assumptions that may not be warranted. They are often developed with little research. In this report, we do not recommend the wholesale abandonment of factors where they are known to improve quality. However, it is usually better to tackle the problems at source. We therefore recommend that matching work continue.

**Recommendation 17** *ONS should measure any bias that may arise because of the treatment of potential duplicate businesses, and revise the treatment of reporting units in surveys as appropriate. (Priority: 2)*

## **Chapter 10**

### **Online Processes and Validation of ONS Input Data**

#### **10.1 Introduction and Summary of Main Quality Issues**

In this chapter, we describe the checks that are in place to ensure the quality of online processes, and the validation of data used for updating the IDBR.

Overall, the level of quality assurance is high. Supervisors check a lot of their staff's work, particularly that of new staff members. In addition, an independent audit is carried out every two months to establish whether procedures are correct and have been followed properly. The audit is generally good and we make a few observations that will ensure it is as effective as possible.

Data from the administrative sources are routinely vetted for quality, although ONS data are not, because they are assumed to be of good quality. Consequently, some data loaded onto the IDBR weren't of sufficient quality.

#### **10.2 Online Amendments**

We are pleased to report that a good deal of quality assurance of online amendments takes place. In particular, supervisors play an important role in checking their staff's work. Most notably, extensive checking is carried out on amendments made by new members of staff.

In addition to checking work within teams, an independent audit is carried out by BRU staff. This involves the selection of about 60 pieces of work, covering a range of different amendment activities and involving different members of BRU's staff in Newport. The selection is made every two months.

The audit team obtains documentation relating to each update from the team concerned and then checks that the update was made correctly. At the end of the period of investigation, a report is written, containing recommendations for changing practices and improving documentation. At the time of writing this report, the audit process was being extended to include amendments made by staff in Runcorn. Amendments made by DETINI's staff are not included in the audit, although their inclusion is being considered.

The audit is a good quality-assurance mechanism because it identifies procedures that need to be clarified. It was reported that about one-third of amendments checked result in some form of recommendation or advice. The report of each audit is disseminated to the more senior managers within BRU, who then cascade the relevant information to their staff. Overall, the process of releasing information within teams appeared to be effective, with each team concentrating on those topics that are relevant to its responsibilities. Other possibilities could be considered, such as BRU-wide seminars on important topics.

We identify some additional changes to the audit process that could maximise its value. First, the audit needs to be timelier. In some cases, many weeks will pass between a piece of work being completed and the audit being carried out. It isn't always easy to reconstruct actions

and thought processes with such a delay. It may be possible to make the audit more timely by increasing its frequency.

Second, the audit needs to be more focused on those amendments that can have a material effect on the quality of survey estimates or other analyses. Some users of the audit felt that too many of the amendments checked were very simple, and the benefit of correcting any error would be so small it wouldn't be worth it.

Third, it was reported to us that it is sometimes difficult to determine precisely what recommendations should follow particular audit findings, as there is often a degree of subjectivity in how best to treat some amendments. It would benefit the audit process if a small group was set up to consider the impact of each set of findings, and agree recommendations.

Fourth, we found that at the time the reports were being written there wasn't enough consultation between the people conducting the audit and the people being audited. It would be better if audit reports were written in consultation with the BRU's administrative managers. Consulting members of staff who are responsible for systems and have detailed knowledge of updating the IDBR will help to ensure that good recommendations are made and that the audit is as useful as possible.

**Recommendation 18** *The audit of online amendments should be more timely and frequent. The sample size of the audit should be reviewed and the audit should be focused on those amendments that can have material impact on the quality of the IDBR. Recommendations should be developed and agreed in partnership between the audit team and work areas. (Priority: 2)*

### **10.3 Validation of Input Data**

Validation of data used to update the IDBR is important. At present, data from the administrative sources are vetted. Prints are produced for clerical intervention where incoming data fail to meet a wide range of quality standards.

Conversely, ONS data used to update the Register are not vetted. In such cases, the quality of ONS sources is assumed to be high. We found that in some relatively isolated cases, values have been taken onto the IDBR that are clearly incorrect. In addition, we found cases where IDBR processes have resulted in missing or invalid data at the time of survey selections. Some cases involved negative turnover values; there were others where the number of employees was greater than the total employment (defined as employees plus working proprietors) and others where the region code was missing. These are relatively minor quality problems, and are probably quite easy to resolve.

**Recommendation 19** *Business Registers Unit should extend the range of validation checks on inputs to ensure that statistical data present on the IDBR are of adequate quality. (Priority: 3)*



## **Chapter 11**

### **Management and Communication with Users**

#### **11.1 Introduction and Summary of Main Quality Issues**

In this chapter, we consider a range of findings relating to the IDBR's management, communication between BRU and Register users, and documentation about the IDBR.

Overall, provisions have been made to ensure that members of staff receive the training appropriate to the job. BRU's own training unit supports this. Members of staff should be encouraged to experience as broad a range of register operations as possible. Managers are supported by a wide and growing variety of management information, and this can be improved if it's more heavily focused on the impact of the Register's quality. Recently, a monthly report was introduced which highlights the state of the Register, and the effect of updates to it. Further development of this report will make it more useful to recipients.

Documentation produced for those who maintain the Register is good, but not for those who use it. BRU communicates effectively with most users, but communication needs to be improved between BRU and DETINI.

Three committees are involved in the IDBR's management, although none has taken responsibility for ensuring that recommendations generated by methodological studies are evaluated and implemented. There is a risk that lack of clarity regarding these groups' responsibilities could adversely affect the IDBR's quality.

#### **11.2 Staffing, Communication and Training**

This review has confirmed that the IDBR is a vital part of the economic statistical system. As such, it is important to ensure that the specialist function of maintaining a business register is reflected in the staff recruitment strategy. The maintenance of a comprehensive training programme is particularly important, and members of staff should also gain experience in a broad range of areas connected with the Register. These principles were reiterated at the recent Roundtable in New Zealand.

The BRU has recognised the importance of training. In particular, it is important that staff within the BRU have an in-depth knowledge of a range of register operations, the business world, and survey operations. Consequently, BRU has set up and maintained its own training unit. The unit was one of the first to be established within a work area. It provides training for three groups of customers: BRU staff, other ONS staff, and staff in OGDs. We are pleased to see that an enhanced training programme was instigated for members of staff working on the 2000 ARI.

We found that members of staff often tend to become experts in quite narrow fields. Because many register procedures interlink, it is important that staff within BRU gain experience of working in a wide range of register operations.

We emphasise that well-trained staff are important to the quality of the IDBR and economic statistics in general. Consequently, the staffing of BRU's training unit should be maintained at present levels.

**Recommendation 20** *Business Registers Unit (BRU) should develop a specific recruitment policy and training programme that reflects its specialist function. This policy should include arrangements to ensure that members of staff are encouraged to move within BRU and across Prices and Business Group to ensure a breadth, as well as a depth, of expertise. (Priority: 2)*

The most important area of communication that BRU should address is that with colleagues in DETINI. Those charged with maintaining Northern Irish units on the IDBR should be regarded as an extension of BRU, and have an important and well-recognised role to play in ensuring the consistency of National Accounts, and other economic statistics for the whole of the UK. Proper access to computer systems and documentation is essential if the IDBR is to meet DETINI's needs and produce information of a standard that other users require. Improved electronic communication is particularly important in allowing easy access to the IDBR manual and other documentation.

In addition, legal constraints imposed by the Statistics of Trade Act 1947 (SoTA) and the corresponding statistics of Trade and Employment (Northern Ireland) Order 1988 regulate the use of, and access to, the IDBR. These make it much harder to update the Register and conduct surveys from it. A change to the SoTA that provides DETINI with the access already given to GB government departments would facilitate improvements in efficiency and quality.

**Recommendation 21** *The Business Registers Unit should improve electronic communication with the Department of Enterprise, Trade and Investment Northern Ireland to ensure consistent practice is adopted in the two departments. This should include consideration of how present legal constraints can be overcome in future. (Priority: 2)*

It's clear that many users don't understand what constitutes a good register and there is a need for education on both sides. Register staff should communicate with users in language they can understand. More importantly, producers of statistics based on samples drawn from the IDBR, or direct analysis of the IDBR, need to have a good understanding of the sampling frame, and some of the general principles of maintaining and using sampling frames. Otherwise, there is a risk of misuse and poor quality statistics. The present Statistical Analyst scheme operating in ONS Newport has been used to educate staff about the IDBR. The parallel Data Analyst Scheme, which is being developed, provides a further opportunity to educate people about frames and the IDBR.

**Recommendation 22** *ONS should strengthen staff training with regard to the principles of maintaining and using a business register. The use of the Data Analyst and Statistical Analyst schemes for this purpose is endorsed. (Priority: 3)*

### 11.3 Documentation

The IDBR has an extensive operations manual, which is produced for people who update the Register and describes the processes involved in doing so. The manual is comprehensive, and is updated frequently as a result of the audit of amendments described in section 10.2 and other developments. The manual is only available within ONS because it uses the Office's internal communication software. The manual is supported by an area of the National Statistics website that provides information for general users.

Documentation for other users is at present more limited, although it is published on the National Statistics website, making it widely accessible. Notably, there is a lack of detail on how the Register works and is maintained. We reported in Chapter 2 that users required better documentation. One recent development in ONS was to consider a way of presenting information on the IDBR's internal data flows pictorially. This has been shelved at present, due to lack of resources and adequate software, but it could be a valuable tool. As part of the recommendation to improve documentation, consideration should be given to developing this work.

Another valuable tool would be a frequently asked questions document or database. Ideally, such documentation would be available over the Government Secure Intranet so that it could be made available to a wide range of users.

**Recommendation 23** *Additional documentation should be produced for users of the Register. Part of this documentation should take the form of a frequently asked questions document. (Priority: 1)*

### 11.4 Management Information

Management information is important for monitoring the Register-updating processes. During this review, it was pleasing to find that an increasingly wide range of information is being produced, both for publication and for use by BRU staff who manage Register maintenance. The information we saw related to the quantity of amendments; it would be better if such information focused on the Register's quality.

Since the beginning of 2000, BRU has produced a monthly overview of the IDBR's structure, noting any major updates that might cause discontinuities in survey estimates. The report is circulated to ONS survey managers and to relevant staff in OGDs via the IDBR Management Committee.

The report is a useful vehicle for identifying major changes to the IDBR that could have an effect on survey results. It presents summary tables by a range of categories such as industry, region and legal status, and includes notes to explain any large changes.

As part of this review, we consulted a small sample of users within ONS to gauge the report's usefulness. We found that it wasn't extensively used. Some respondents commented that they looked at the report only when a particular problem arose; others said that the report should make it easier to identify in advance changes to the Register that may affect survey operations or estimates. An interest in a longer time series was also expressed.

The report's use is limited, particularly in OGDs, because it contains potentially disclosive data and is therefore marked 'Restricted – Commercial'. It is currently impractical to implement disclosure-control techniques because the report's timeliness is important. The lack of automated disclosure-control procedures prevents the production of a timely report that is protected against unlawful release of data.

In Australia, a similar report is produced every quarter. This contains more text than the ONS version, as well as graphs showing a longer time series of some of the main register characteristics. An example is attached at Annex 13.

In the Netherlands, users can access information tailored to their needs through a central, networked query-based system. An extract from the business register is held centrally, and tabulations can be specified directly.

Users will clearly benefit from being warned of the impact of updates to the business register and being able to query changes to it. However, at present, the system needs to be reviewed, in consultation with users. Two particular areas of interest are the relevance to users and the restricted nature of the report for users in OGDs.

**Recommendation 24** *The existing quarterly and monthly management reports should be focused on measuring the impact on quality of changes to the Register. Business Registers Unit should continue to develop its monthly report of changes to the size and structure of the Register in consultation with users. (Priority: 2)*

## **11.5 User Interface**

An important communication channel is the online interface between the IDBR data and the online user. At present, this is still keyboard driven, due to the lack of software tools within the Ingres environment.

A more user-friendly interface might be very beneficial. The greatest benefit is that it allows much more information to be displayed on screen at any one time, and this can be presented in a more logical and structured way. This would result in quicker updating and fewer errors, which would reduce the risk to quality.

**Recommendation 25** *An improved online interface should be developed to make it easier for authorised users to access and update the Register. (Priority: 2)*

## **11.6 Evaluating and Overseeing the Implementation of Recommendations from Methodological Studies**

In recommending improvements to the IDBR, we should have a clear idea of the effect of changes on users and on the Register itself. It is therefore important to carry out an impact analysis of proposed changes to the IDBR or Register processes before making those changes.

**Recommendation 26** *ONS should ensure that the impact on outputs is assessed for any proposed changes to the IDBR. (Priority: 1)*

There is no single group responsible for overseeing the implementation of recommendations from studies that affect the IDBR. Recent ONS studies include the Kokic/Brewer review of operational aspects (see Chapter 12), the Births and Deaths study and the Short-Term Inquiry Project (STIP). In the cases of the latter pieces of work, there has been no mechanism to ensure that any recommendations are implemented. This carries with it an increased risk of the IDBR failing to keep up to date with the most recent developments or best practice.

It is important that such a mechanism be established. Ideally this would be within ONS's Register strategy group, which would be restructured to include a non-ONS representative. The immediate tasks should be to develop action plans arising from the recommendations of the Births and Deaths and STIP projects. The remaining recommendations from the Kokic/Brewer study are included in the recommendations of this review.

**Recommendation 27** *ONS's Register strategy group should take responsibility for overseeing the implementation of recommendations from projects relating to the improvement of the IDBR. The membership of the group should be reviewed and include a non-ONS member. (Priority: 1)*

## **Chapter 12**

### **Implementation of Recommendations from Kokic/Brewer Study**

#### **12.1 Introduction**

In 1996, Phil Kokic and Ken Brewer, both of them from the University of Southampton, were commissioned to review some aspects of the IDBR's operation. The report,<sup>16</sup> which contained a general review of the IDBR, addressed four main areas:

- sample selection, particularly sample stability;
- identification and treatment of duplicate reporting units;
- problems associated with births and deaths of businesses; and
- updating the Register.

They also made a number of recommendations relating to the way the IDBR is maintained and used as a sampling frame.

This section reports the extent to which the recommendations have been implemented or, if they weren't, examines whether they are still relevant. We make recommendations for action or further work as appropriate. In this section, we summarise the main recommendations requiring further work by ONS. A summary of the recommendations from the Kokic/Brewer report appear in Annex 14.

#### **12.2 Sample Selection**

Kokic and Brewer examined a range of aspects relating to sample design and selection. Various recommendations were made to ensure that sample membership is more stable (that is, to prevent businesses moving in and out of sample in an uncontrolled way) and to minimise the response burden and spread it evenly throughout the business community. Various recommendations haven't yet been implemented and some need tidying up.

First, it was noted that under a rotational sampling scheme, high sampling fractions lead to long periods in sample followed by only short periods out of sample – 'on holiday'. It was noted that it would be better in such cases to ensure a minimum holiday period is allowed between consecutive inclusions, or that the business should be included in the sample all the time. Therefore, it was proposed that where the sample design process suggested a sampling fraction of between 0.5 and 0.75, the sampling fraction should be set to 0.5 to ensure that the holiday period matches the time in sample. Similarly, where the sampling fraction suggested was between 0.75 and 1, the stratum should be completely enumerated. Although this rule is generally implemented, it isn't documented and it should be. In exceptional situations, it may be prudent to relax the rule. For example, if the rotation period were long – five years, for example – it wouldn't be sensible to insist on a full five-year holiday if it entailed the loss of a substantial degree of efficiency or precision.

ONS currently operates a partially co-ordinated overlap system to reduce the burden on businesses. Surveys were started off at distinct parts of the PRN line, but are allowed to sample from the whole line. The effect is that, over time as individual sampling intervals move at different speeds, different samples may overlap. ONS should review the

requirements for control of overlap in order that an appropriate overlap-control system can be designed and implemented.

Turning to sample design, the Kokic/Brewer report made a range of recommendations aimed at ensuring that sample selection and overlap control systems are as easy to implement as possible. These focused on ensuring consistency of stratum definition, and the effect that small strata can have on sample stability. Some progress has been made towards standard strata – employment (or, occasionally, employees) is used across the board as the stratification variable and, in most cases, the same stratum boundaries are used.

Things have moved on since then, and it's unclear whether the use of employment strata for the collection of variables related to turnover is ideal. A sampling and stratification review has been proposed by ONS, and this should be given high priority. The review will include research into the use of more than one size variable (or a composite variable) for stratification; standardisation of strata (both for size and industry); identification of optimal industrial stratum definition; and the best way to stratify small businesses given the inaccuracy of industrial classification.

Finally, an anomaly may occur in sample selection where the net gains (new births added minus deaths removed) to a sampling interval exceeds the number of reporting units due to be rotated out of the sample. In such cases, the sample will be adjusted temporarily by rotating out the reporting unit(s) that were included most recently. This causes confusion among respondents and extra work for those responsible for collecting and validating data. It can also lead to a reduction in the quality of statistics, for example by reducing the overlap between consecutive samples. A range of fixes is available and, although this would add to the processing time of sample selection, one should be implemented.

**Recommendation 28** *ONS should change sample selection procedures to ensure stability of sample membership for businesses newly included in samples. (Priority: 2)*

### **12.3 Duplicate Units**

Duplication arising through the non-matching of VAT and PAYE units was addressed by the Kokic/Brewer report. Recently, ONS has taken this further. See Chapter 9 of this report for more details.

The Kokic/Brewer review also commented on the difficulties of imputing employment values for VAT-based enterprises where only turnover information is available. ONS reviewed the methods used, but found that it wasn't possible to make any substantial improvements to the method. Our research for this review suggests that the present procedures are subject to wide variation, but are approximately unbiased. ONS's proposed sampling and stratification project will evaluate how best to stratify small enterprises, given the known variability inherent in the current imputation process.

### **12.4 Births and Deaths**

The Kokic/Brewer report outlined the best practice for dealing with births and deaths in

sampling and estimation. ONS followed up their comments with a study of births and deaths. The recommendations of that study haven't yet been implemented, nor is there any plan to do so. ONS's current practice is not accepted statistical best practice because deaths are fed back to the IDBR from sample sources, although research for one particular survey has suggested that the bias introduced by this may be quite small. It is important that ONS address these issues. Detailed recommendations from the study are included at Annex 15.

**Recommendation 29** *ONS should set aside resources to study further the impact of births and deaths of enterprises on statistical outputs. (Priority: 1)*

## **12.5 Updating the Register**

The review pointed out that it's only acceptable to feed back information from surveys where samples are drawn independently from ONS's statistical surveys. We have seen in Chapter 6 that even the ARI is not run in this way. In addition, our analysis as part of this review has shown that, in some cases, there is a good deal of classification change from surveys. We haven't had time to evaluate whether bias is being introduced into surveys because of this feedback, but we highlight the possible risk. It is our intention that the ARI's expansion and the proposals we've made for using the ARI to update the Register will mean that feedback from surveys is no longer needed.



## **Chapter 13**

### **Concluding Remarks**

The report contains 29 recommendations that will develop the IDBR as a key element of the ONS infrastructure, delivering a service that responds to increasing user demands, while minimising the burden on business. This will be achieved by working more closely with the administrative departments, improved use of technology and taking account of the best practices of other statistical offices. By implementing these changes, the IDBR will continue to deliver a high quality service to its users.

## Notes and References

1. ONS (2001). *A Quality Strategy for National Statistics*. ONS: London.
2. Perry J (1995) The Inter-Departmental Business Register. *Economic Trends* 505, November 1995.
3. Council Regulation on the statistical units for the observation and analysis of the production system in the Community – EEC 696/93
4. ONS (2000). *PA1003 Size Analysis of United Kingdom Businesses*. ONS: Newport.
5. Small Business Service (2000). *Statistical Bulletin Small and Medium Enterprise (SME) Statistics for the United Kingdom 1999*. SBS: Sheffield.
6. Small Business Service (2000). *Statistical Bulletin Business start-ups and closures: VAT registrations and de-registrations 1980–99*. SBS: Sheffield.
7. ONS (1997). *UK Standard Industrial Classification of Economic Activities 1992*. TSO: London.
8. Council Regulation on Community coordination in drawing up business registers for statistical purposes – EEC 2186/93
9. ONS (1998). *Survey of Business Register Quality*. Report for Eurostat. ONS: Newport.
10. Davies P and Smith P (eds.) (1999) *Model quality report in business statistics*, vols. I–IV. ONS: Newport.
11. Fenton T, Hedlin D, Perry J and Pont M (1999). *Births and Deaths of Business Units*. Report for Eurostat. ONS: Newport.
12. HM Treasury (1999). *Modern Public Services for Britain: Investing in Reform*. HM Treasury: London.
13. Cruikshank D (2000). *Competition in UK Banking – A Report to the Chancellor of the Exchequer*. TSO: London.
14. Jones T, Francombe K and Perry J (1994). *Population Units in the Major Business Surveys*. Internal Report. CSO: Newport.
15. see Chapter 5 of (4).
16. Kokic P and Brewer K (1996). *Operational Aspects of the IDBR*. Internal Report. ONS: Newport.

Select the HAND tool and click on the title below to enlarge the view.  
When the cursor becomes a pointed finger items clicked on will come into view.  
Use Bookmarks to navigate to sub-levels within Annexes (where marked).

### **List of Annexes**

- 1 Scope of the National Statistics Quality Assurance Review of the IDBR
- 2 Bodies With Access to the IDBR
- 3 User Consultation and List of Respondents
- 4 Extract from Eurostat's Statistical Programme Committee Report on Register Quality
- 5 IDBR Quality Questionnaire
- 6 Estimating by IDBR Undercoverage Arising from Birth Lags
- 7 Topics Included in Statistics Netherlands' Business Register Customer Satisfaction Questionnaire
- 8 Comprehensive Business Directory: ONS Managers' Briefing Paper
- 9 CLAMOUR leaflets
- 10 Description of Permanent Random Number Sampling in ONS
- 11 Extracts from Statistics Canada and ABS Progress Reports at the 2000 Roundtable
- 12 Unregistered Businesses
- 13 Extract from ABS' Quarterly Business Register Quality Report
- 14 Summary of Recommendations from Kokic/Brewer Report
- 15 Recommendations from the Births and Deaths Study

## Glossary of Abbreviations

ABI	Annual Business Inquiry
ACOP	Annual Census of Production
AES	Annual Employment Survey
ARI	Annual Register Inquiry
BCG	Business Competitiveness Group
BPT	Business Profiling Team
BRU	Business Registers Unit
BST	Business Structures Team
CBU	Complex Business Unit
CH	Companies House
CLAMOUR	CLAssifications MOdelling and Utilities Research
CSO	Central Statistical Office
D&B	Dun and Bradstreet
DETINI	Department of Enterprise, Trade and Investment Northern Ireland
DETR	Department of the Environment, Transport and the Regions
DfEE	Department for Education and Employment
DSI	Distribution and Services Inquiry
DTI	Department of Trade and Industry
ESA	European System of Accounts
Eurostat	The Statistical Office of the European Community (also SOEC)
HMCE	HM Customs and Excise
IDBR	Inter-Departmental Business Register
IR	Inland Revenue
ISRU	Inquiry Specific Reporting Unit
KAU	Kind of Activity Unit
LFS	Labour Force Survey
MAFF	Ministry of Agriculture, Fisheries and Food
MG	Methodology Group
MIDSS	Monthly Inquiry to the Distribution and Services Sector
MWSS	Monthly Wages and Salaries Survey
NACE	Nomenclature générale des activités économiques dans les Communautés européennes.
NAW	National Assembly for Wales
NSI	National Statistical Institute
OGD	Other Government Department
ONS	Office for National Statistics
PAYE	Pay As You Earn
PBG	Prices and Business Group
PDC	Precision Data Coder
PPI	Producer Price Index
PRN	Permanent Random Number
PRODCOM	PRODucts of the European COMmunity
QIDSS	Quarterly Inquiry to the Distribution and Services Sector
RQS	Register Quality Survey
RRC	Regional Registration Centre
RSI	Retail Sales Inquiry
SIC(92)	Standard Industrial Classification (revised 1992)
SoTA	Statistics of Trade Act 1947

SPC	Statistical Programme Committee
STES	Short-Term Employment Survey
STIP	Short-Term Inquiry Project
TCN	Trade Classification Number
UNECE	United Nations Economic Commission for Europe
VAT	Value Added Tax

## Glossary of Terms

Administrative sources	Providers of data from administrative systems. Administrative sources that supply the IDBR are HMCE, IR and CH.
Administrative data	Data from administrative sources.
Auxiliary information/variables	Employment, turnover and industrial classification information on the IDBR that is used to stratify survey samples and improve the precision of survey estimates
Birth	A newly-registered business.
Burden on business	The burden incurred by businesses arising from the requirement to complete statistical survey questionnaires.
Current variables	IDBR variables updated whenever changes occur.
Death	A newly-deregistered business.
Deregistration	The process of reporting a business close-down to the administrative department.
Duplication	When more than one copy of a record appears on the Register.
Enterprise	The smallest combination of legal units that is an organisational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources. An enterprise carries out one or more activities at one or more locations. An enterprise may be a sole legal unit.
Enterprise Group	An association of enterprises bound together by legal and/or financial links. A group of enterprises can have more than one decision-making centre, especially for policy on production, sales and profits. It may centralise certain aspects of financial management and taxation. It constitutes an economic entity which is empowered to make choices, particularly concerning the units which it comprises.
Frozen variables	IDBR variables updated once a year.
Inquiry Specific Reporting Unit	Reporting units set up specifically for use by just one inquiry. They are also established to distinguish between NI and GB sections of an enterprise.
Local Unit	An enterprise or part thereof (eg a workshop, factory, warehouse, office, mine or depot) situated in a geographically identified place. At or from this place economic activity is carried out for which – save for certain exceptions – one or

	more people work (even if only part-time) for one and the same enterprise.
Matched	IDBR units (usually one VAT unit and one PAYE unit) that have been linked together.
Matching	The process of linking VAT unit and PAYE unit records.
Non-representative member	A member of a VAT group not nominated to liaise with HMCE to pay tax due.
Overcoverage	When some units in the sampling frame aren't in the population of interest.
Profiling	The process of establishing the correct structure, and best reporting arrangements for complex businesses.
Proven (enterprise)	Enterprises that have undergone proving and for which local unit information is stored on the IDBR.
Proving	The process of validating the information sent to ONS from the administrative sources, and of establishing the local unit structure for new enterprises.
Registration	The process of reporting a new business start-up to the administrative department.
Reporting Unit	The unit upon which surveys report their data. In general, the reporting unit is the same as the enterprise.
Representative member	The member of a VAT group nominated to liaise with HMCE to pay tax due.
Rotation	The process of systematically dropping businesses from a sample and replacing them with others.
Roundtable	The International Roundtable of Business Survey Frames: an informal meeting, generally held annually, of countries with an interest in developing their business registers.
Standard survey population	The range of enterprises from which most of the main ONS surveys select samples and produce survey estimates.
Undercoverage	Where units of interest aren't included in the sampling frame.
Unmatched	IDBR units (usually one VAT unit and one PAYE unit) that should be linked, but aren't.
VAT Group	A group of businesses under common ownership registered together for VAT payment purposes.

## Some Useful Websites

(Addresses as at 8 February 2001)

ONS	<a href="http://www.statistics.gov.uk">www.statistics.gov.uk</a>
IDBR	<a href="http://www.statistics.gov.uk/nsbase/themes/commerce/Services/idbr.asp">www.statistics.gov.uk/nsbase/themes/commerce/Services/idbr.asp</a>
DETINI	<a href="http://www.detini.gov.uk">www.detini.gov.uk</a>
Scottish Executive	<a href="http://www.scotland.gov.uk">www.scotland.gov.uk</a>
National Assembly for Wales	<a href="http://www.wales.gov.uk">www.wales.gov.uk</a>
HMCE	<a href="http://www.hmce.gov.uk">www.hmce.gov.uk</a>
Inland Revenue	<a href="http://www.inlandrevenue.gov.uk">www.inlandrevenue.gov.uk</a>
Companies House	<a href="http://www.companies-house.gov.uk">www.companies-house.gov.uk</a>
DTI	<a href="http://www.dti.gov.uk">www.dti.gov.uk</a>
DETR	<a href="http://www.detr.gov.uk">www.detr.gov.uk</a>
Small Business Service	<a href="http://www.businessadviceonline.org">www.businessadviceonline.org</a>
UNECE	<a href="http://www.unece.org">www.unece.org</a>
Eurostat	<a href="http://europa.eu.int/comm/eurostat">europa.eu.int/comm/eurostat</a>
CIRCA (European Information Database)	<a href="http://forum.europa.eu.int/dsis">forum.europa.eu.int/dsis</a>
HM Treasury (Invest to Save)	<a href="http://www.hm-treasury.gov.uk/isb">www.hm-treasury.gov.uk/isb</a>
CLAMOUR	<a href="http://www.statistics.gov.uk/nsbase/methods_quality/Clamour.asp">www.statistics.gov.uk/nsbase/methods_quality/Clamour.asp</a>
CBD	<a href="http://www.statistics.gov.uk/about_ns/e_business_annex_1_7.asp">www.statistics.gov.uk/about_ns/e_business_annex_1_7.asp</a>
International Roundtable 2000	<a href="http://www.stats.govt.nz/nzroundtable">www.stats.govt.nz/nzroundtable</a>
Dun and Bradstreet	<a href="http://www.dnb.com">www.dnb.com</a>
Search Software America	<a href="http://www.searchsoftware.com">www.searchsoftware.com</a>
Inference Group Pty	<a href="http://www.inferencegroup.com.au">www.inferencegroup.com.au</a>



## Scope of the National Statistics Quality Assurance Review of the Inter-Departmental Business Register

### INTRODUCTION

1 This paper describes the scope of the review of the Inter-Departmental Business Register (IDBR). The paper has been produced following comments received from interested parties, including the group that has been set up to steer the review. The paper sets out the organisation of the project and a brief description of the broad areas that will be reviewed, along with those that have fallen outside the scope of the review.

### BACKGROUND TO REVIEWS

2 The White Paper *Building Trust in Statistics* announced the introduction of a rolling programme of reviews of key National Statistics outputs. The review of the IDBR falls within that programme and within the Commerce, Energy and Industry theme.

### ORGANISATION OF THE PROJECT AND PERSONNEL

#### Review Team

3 Mark Pont of the Office for National Statistics' (ONS) Methods and Quality Division (MQ) will be the full time Review Team leader and project manager (65 days). He will be assisted by Ray Chambers (University of Southampton, 5 days, in a quality assurance role), John Perry (20 days, as owner of the system) and Gareth Jones (say 10 days, representing survey users). Additional resources for analysis will be provided by Business Registers Unit (BRU) and MQ staff (40 days each).

#### Project Board

4 The project will be overseen by a Project Board comprising:

Tim Jones, Methods and Quality Division (chair) [Marta Haworth from August 2000]  
John Kinder, Prices and Business Group, ONS (PBG)  
Graeme Walker, PBG  
The Review Team

5 The rôle of the Project Board is to:

- agree the Project Definition Document in its entirety and any subsequent changes to it;
- commit the necessary resources;
- arrange dissemination of progress and the final report;
- approve the completion of each stage of the project and the plans for the next stage. At completion to formally sign off the project;
- to advise the Project Manager immediately if it is aware of any problems which may affect the outcome of the project; and,

- to ensure that the products are delivered to agreed standards, on schedule.

## **Steering Group**

6 In addition, a Steering Group will:

- provide direction and guidance to the Review Team;
- provide an external perspective on the review;
- assure itself that the appropriate processes have been followed through the review, and the appropriate stakeholders consulted; and,
- offer advice and guidance on issues brought to it by the Review Team.

7 The Steering Group is made up of:

Susan Linacre, ONS (chair)

Janet Dougharty, Department of Trade and Industry (DTI)

Geoff Noon, Business Statistics Users Group/Machine Tool Technologies Association (MTTA)

Iain Bell, Department for Education and Employment (DfEE)

Frances Pottier, Department of the Environment, Transport and the Regions (DETR)

Hayley Harris, HM Customs and Excise (HM C&E) [Vivienne Burch from October 2000]

Mike Pepper/John Kinder, ONS

Tim Jones, ONS [Marta Haworth from August 2000]

Mark Pont, ONS

Members of the Review Team as necessary

## **TIMETABLE**

8 The first phase, to scope the review, will be completed by end-May 2000. The review will commence immediately after that. Emerging findings will be presented in mid-September 2000 and the final report at the end of October 2000.

## **LINKS/DEPENDENCIES**

9 This review is closely linked with the work of various groups along with other pieces of research and review work ongoing in BRU and other areas. Of particular note are:

- IDBR Management Committee
- IDBR Strategy Group
- Development of the ARI 2000
- Quality work funded by the Statistical Office of the European Community (Eurostat)
- Efficiency reviews of Complex Business Unit and register updating procedures
- Short Term Inquiry Project
- Work on the Single Business Register/Comprehensive Business Directory
- The Eurostat-funded CLAMOUR project
- Ongoing reviews of registration thresholds by supplier departments

## SCOPE

10 Reviews cover three areas:

- fitness for purpose of the outputs;
- appropriateness of the methodology and processes; and,
- quality of the inputs and burdens on suppliers.

### *Fitness for purpose of the outputs*

11 The IDBR has too many customers to consider the fitness of the product for all. The review will concentrate on the major outputs that affect the vast majority of customers. The main customer/output areas are:

- PBG surveys for sample selection and provision of population auxiliary information, and other data relating to the sampling frame, such as the extent and impact of reclassifications;
- DfEE for local area analyses of employment;
- Input/Output for structural tables (and National Accounts Group more generally as users of survey outputs);
- DTI for business demographic analyses, and energy surveys;
- DETR for construction statistics;
- various departments for sub-national estimates (eg DETR Town Centres Project, Scottish Executive, National Assembly for Wales); and,
- the Statistical Office of the European Community (Eurostat) for analyses, and compliance with regulations.

12 By focusing on the IDBR as an output for these main areas, we will be addressing the main processes involved in the register. The choice of main customer areas will not restrict the breadth of consultation on the review. Wide consultation will take place with users within ONS and in the rest of the Government Statistical Service via the IDBR Management Committee, and by other regular communication.

### *Appropriateness of the methodology and processes*

13 The review will provide information for developing a strategy for improving register quality (being taken forward by the ONS Register Strategy Group) and for determining where development money should be spent. The review will also aim to define quality measures which can be used to monitor the quality of the register in future. The review will not be developing any new analyses of the register other than those needed for evaluating and measuring the quality of the IDBR and its processes. We will review whether the IDBR is fit for the purposes for which it is used, rather evaluating whether the uses made by customers are appropriate.

14 There are many processes associated with the register. We will review some of the

processes involved in updating the IDBR from external and internal sources, sample selection, and quality control (including register proving and quality surveys) to establish the effect they have on the quality of the register. Assessment of quality of such processes can be against a number of other “best practice” benchmarks such as the Eurostat manual on business registers, documentation from the series of meetings known as the International Roundtable on Business Survey Frames, recommendations from internal audits (of BRU and the Comprehensive Business Directory), Kopic and Brewer’s report to ONS entitled *Operational Aspects of the IDBR*, and the published literature (notably the publications from the International Conference on Establishment Surveys (ICES) and the forthcoming ICES II). For the purposes of the review we will be taking the present structure of the register database as given, and do not intend to actively consider any alternatives.

## **FORMAT OF THE PROJECT AND ANTICIPATED PRODUCTS**

15 A list of subprojects has been identified. These are listed below along with a priority marking (A = essential and will be reviewed, B = regrettably not to be reviewed due to time constraints and C = definitely won't be reviewed), and the approximate division of time into the various subprojects that will be tackled. A report for each subproject will constitute the product in each case. Two additional products are proposed that do not fall neatly within the main subprojects.

16 Following the description of the subprojects is a tentative outline timetable and indication of the task leader for each of the work packages. Note that this has been produced without full consultation with the whole project team, and is intended to be indicative. There are no clear dependencies between the various strands of work. It is expected that they will each proceed in parallel for the duration of the review period. A further more detailed product list along with detailed resource plans will be produced in due course.

### **Subproject 1            User requirements. Priority A. 15 per cent.**

17 The purpose of this subproject is to establish the requirements of users and non-users. In particular, we will establish what quality improvements users would like to see from the IDBR, and what the benefits to those users of better quality would be. For non-users, we will establish what quality improvements are necessary for them to be able to use the IDBR. Within the banner of quality improvements, we will also address access restrictions, including legal and confidentiality issues.

### **Subproject 2            Annual Register Inquiry (ARI). Priority A. 35 per cent.**

18 The most important part of the work will be to review the ARI (which includes the quarterly register proving inquiry) as the ARI is fundamental to the maintenance of the IDBR. The ARI is the main source of updates of business structures and auxiliary information on the IDBR. As such, the ARI has a large impact on the quality of the register. We need to establish how good the ARI is, and what may be done to improve the quality of the ARI. The main issue is the delineation of units (which relates to the work of the CBU), and the identification of new local units. We will consider the most appropriate ways to collect the data about local units, exploring the possibilities for using new technology and

other data sources (for example, Trade Associations) where possible, and associated costs. We will also consider whether complete censuses of businesses are required, and whether there would be a net benefit in conducting such a census.

19 We also need to determine how we might use the ARI in future as a vehicle for measuring some key aspects of register quality – coverage, business structures, industrial classification and employment. This will necessitate a thorough review of the design and operations of the ARI.

### **Subproject 3 Coverage – small units and matching. Priority A. 5 per cent.**

20 Undercoverage arises for various reasons. Of most importance is that due to units operating below the registration thresholds for the relevant administrative sources, which although believed to be a small proportion overall, can represent a much more significant chunk in some industries (for example, construction). Overcoverage can be caused by our not being able to find a match for units identified by both administrative sources. Consequently, we exclude certain units from the register whilst a match is found. This can result in further undercoverage. A large Eurostat-funded project has just been completed which has reviewed and attempted to improve the matching process. We will carry out an objective audit of the research work and consider the impact on quality, and identify the implications of various ways of treating duplicate units.

### **Subproject 4 Business structures. Priority A. 5 per cent.**

21 An accurate local unit structure of businesses is important for small area estimates produced from the Annual Business Inquiry and the Short-Term Employment Surveys required by DfEE and others. The review will cover the quality of auxiliary information stored at local unit level, the interaction between local unit and reporting unit classifications, and the annual data confrontation exercise which aims to ensure coherence between local unit and reporting unit classifications. We should consider whether the kind of activity unit (KAU) structures we have at present meet the requirements of the European regulations, and whether the structures meet the needs of the NAW and SE for regional sampling and analysis.

### **Subproject 5 Complex businesses – profiling and sampling. Priority A. 10 per cent.**

22 The Complex Business Unit and the Business Structures Team (part of BRU) have been running for several years. We should review the implementation in the light of the requirements of the Population Units paper and the developing needs of survey statisticians. Of particular importance are (a) the range of enterprises that are in scope for CBU treatment, (b) the way that the IDBR is able to delineate complex businesses from other enterprises for sampling purposes and to report this information to survey users. The way that specific surveys treat complex enterprises in sampling and estimation, and the data collection methods for such units are specifically excluded from the scope of this review.

23 Within this subproject we also propose to review inquiry specific reporting units. There are many such units and they require additional maintenance and cause extra

complication in sampling and estimation. In addition, the compliance cost to individual enterprises can be high. We should review in general whether such structures are necessary.

## **Subproject 6            Industrial classification. Priority A. 20 per cent.**

24     We propose to review three areas relating to industrial classification: rules, use and quality.

25     First, we will review the implementation of the NACE classification rules on the IDBR. This will incorporate a review of the rules used to prioritise updates from different sources and a review of the quality of data from each of the sources used. The review will cover the rôles of those sources that can classify units only in part of the economy, for example PRODCOM (a regular survey collecting information on PROducts in the European COMMunity) updates only production industries, which alone would lead to an outflow of units from those industries. PRODCOM has a different way of classifying businesses than other sources which leads to inconsistency. Similarly, DETR can classify only construction businesses. Again, this would result in a net outflow of business from that sector, and the different perceptions of the classification of businesses could lead to an inconsistency of treatment between surveys

26     The review will also cover the use of classifications in general by surveys, especially in areas where quality is poor, such as the miscellaneous categories, and the use of classifications at the finest level of detail. We will also consider whether the rules and processes involved in the maintenance of current and frozen fields which have been in place for several years remain appropriate (that is, whether they improve the quality of statistical outputs and improve the efficiency of statistical processes).

27     The third area for this subproject relates to quality. Work is currently being undertaken in BRU based on the Register Quality Survey and recent ARIs to evaluate the quality of classifications and the sensitivity of survey results to misclassification. We will build on this as appropriate.

28     Under the banner of quality, we will also be considering the quality of industry codes received from Inland Revenue and HM C&E.

29     In the case of VAT classifications, the Register Quality Survey conducted by ONS in 1997 found that about one-third of all classifications from HM C&E did not agree with up-to-date classifications collected by ONS. For PAYE, classifications are not aligned with the UK's Standard Industrial Classification (SIC92). Inland Revenue continues to use a 2-digit Summary Trade Classification (STC) or 4-digit Trade Classification Number (TCN). An SIC92 code is imputed from the Inland Revenue code for use in sampling and estimation. The industrial classifications of over 80 per cent of units representing over a third of employment are taken directly from one of the administrative sources. Department of Enterprise, Trade and Investment Northern Ireland (DETINI) in particular have found that the poor quality of this classification affects the quality of employment statistics relating to the start-up of businesses as the only classification available for these units is from the administrative sources at registration.

30     This subproject will also consider the use of new technology to classify businesses.

**Subproject 7            Development of quality measures. Priority A. 5 per cent.**

31     The aim of this task is to develop and specify (rather than calculate) a range of indicators that can be used to describe and monitor the quality of the IDBR. The development of quality measures will be based on the research in the other areas of the view in addition to reference to published literature.

**Additional products**

**Subproject 8            Audit of the implementation of the recommendations from the Kokic/Brewer review. Priority A. <5 per cent.**

32     Two recent documents have presented overviews of the implementation of the recommendations from the review. We will bring these together as a single product and identify a list of actions, decisions and research that need to be undertaken in order to sign off the review.

**Subproject 9            Benchmark against other European registers. Priority A. <5 per cent.**

33     This product will report the quality of the IDBR against other European business registers. The analysis will be based on the most recent questionnaire of current practice in member states.

**Summary of subprojects and subproject leaders**

<b>Task</b>	<b>Leader</b>	<b>Duration</b>
1 User requirements	Mark Pont	Jun to Jul
2 Annual Register Inquiry	Mark Pont	Jun to Sep
3 Coverage – small units and matching	Mark Pont	Jul to Sep
4 Business structures	Mark Pont	Jun to Sep
5 Complex enterprises – profiling and sampling	Mark Pont	Jun to Sep
6 Industrial classification	Gareth Jones	Jun to Sep
7 Development of quality measures	John Perry	Aug to Sep
8 Audit of the implementation of the recommendations from the Kokic/Brewer review	Mark Pont	Jul
9 Benchmark against other European registers	John Perry	Jul

**TOPICS THAT WILL NOT BE ADDRESSED BY THE REVIEW**

**Geographic variables. Priority B.**

34     The quality and level of geographic referencing information is important in the production of small area estimates (DfEE and others). The quality of the current geographical

details via the postcode will be covered in subproject 1 on the ARI. Development of alternative referencing systems is being taken forward by the ONS Geographic Referencing Strategy.

### **Public/private sector (re)classification. Priority B.**

35 Any classification discrepancy is particularly noticeable in the Average Earnings Index where public and private sector indices are both required. At present, a dedicated ONS team is responsible for public/private sector classifications. This issue will not be addressed by the review. Issues of continuity and the methodology to deal with changes in classification will be addressed in subproject 5 on industrial classification.

### **Coverage – industry. Priority B.**

36 In a National Accounts context, the main area where miscoverage could occur is with agriculture. This is particularly likely with agriculture-related businesses such as retailers of agricultural produce etc. There are long-standing plans to merge information from agricultural census conducted by the Ministry of Agriculture Fisheries and Food (MAFF) with the IDBR. Better integration seems likely to lead to an improvement in the quality of statistical outputs. This issue can be taken forward outside the scope of this review.

### **Coverage – lags. Priority C.**

37 A review has already been completed on the treatment of birth and death lags on the register. An important quality measure could be the level of undercoverage due to missing births or the responsiveness of the register to changes in the patterns of birth and death lags, in particular in rapidly growing or declining industries. The MQ report based on the recent research suggests a way that undercoverage can be modelled.

### *Quality of inputs and burdens on suppliers*

38 The appropriateness and quality of inputs from IR and HM C&E will be reviewed. The proposed detail of the work has been described in subproject 6. In addition, the rôle of other register sources, such as Companies House and Dun and Bradstreet.

39 We will consider whether any improvements to our use are possible, and if so, how these can be made, and what the impact on users will be. We will focus our efforts on those areas where users have more scope to utilise administrative data direct.

40 In considering the requirements of the ARI as an input to the IDBR, we will need to consider the costs (both to ONS and to business) in relation to any possible benefits arising from the ARI.



## Bodies With Access to the IDBR

<b>Bodies with Access to the IDBR</b>	<b>Information protected by Statistics of Trade Act 1947</b>	<b>Information protected by Finance Act 1969</b>
Any person (including TECs) carrying out any research or survey <u>on behalf of the Chancellor</u> into training, industrial relations and employment and related matters	Disclosure permitted by Employment & Training Act 1973 – as amended by Employment Act 1988	Not applicable (see local education authorities below)
Education authorities in GB and the NI Training Authority (NITA)	As above (but NITA excluded)	Disclosure permitted by Finance Act (No. 2) 1987
Local authorities within the meaning of the Town and Country Planning Act 1990 and of the Town & Country Planning (Scotland) Act 1997	As above	As above
Development corporations within the meaning of the New Towns Act 1981 and the New Towns (Scotland) Act 1968	As above	As above
Environment Agency	Disclosure permitted by Environment Act 1995	As above
Scottish Environment Protection Agency	As above	As above
ACAS	Disclosure permitted by Trade Union & Labour Relations (Consolidation) Act 1992	As above
NI departments (see note)	Disclosure <u>not</u> permitted	As above
Welsh Assembly	Disclosure permitted by the Welsh Devolution Act 1999	As above
Scottish Executive	Disclosure permitted by the Scottish Devolution Act 1999	As above
OGDs except Northern Ireland departments (see note below)	Disclosure permitted with a 1947 Act Direction	As above
Health & Safety Executive and Health & Safety Commission	Disclosure permitted (with a 1947 Act Direction) by Health & Safety at Work Act 1974	As above

Note: Information protected by the VAT Act 1994 may be disclosed to OGDs and NI departments

User Consultation and List of Respondents

USERS OF THE IDBR

Contact name, phone number and email address:

Organisation:

**1 What use do you make of the IDBR?** (please state all that apply such as direct receipt of tabular analysis, receipt of disaggregate data for further analysis, running a statistical survey, mailing list/cross-checking against other registers, indirect user of IDBR data (eg from a survey or analysis conducted by someone else), administrative purposes etc)

**2 Which populations or domains are you interested in?** (please state all that apply, and give details or industrial sectors, geographical areas, size etc eg "Financial intermediation businesses with less than 100 employment operating solely in Northern Ireland")

**3 What three aspects of the quality of the IDBR make it a particularly useful tool for you?**

**4 What aspects of quality restrict your use of the IDBR?** (please list all that apply and please be as specific as possible eg availability of relevant size information (turnover, employment etc), out-of-date, coverage issues, business structural issues, contact details, sampling mechanism, legal/access restrictions, industrial classification, area classification etc)

**5 What three individual quality improvements would enhance your use of the IDBR most?**

**6 If you wish to make any other comments please do so here.**

Thank you

## NON-USERS OF THE IDBR

Contact name, phone number and email address:

Organisation:

**1 Why don't you use the IDBR?** (please state all that apply eg legal/access restrictions, quality, unaware of what the IDBR can provide etc)

**2 What use(s) would you make of the IDBR if it were suitable/available to you?** (please state all that apply such as direct receipt of tabular analysis, receipt of disaggregate data for further analysis, running a statistical survey, mailing list/cross-checking against other registers, indirect user of IDBR data (eg from a survey or analysis conducted by someone else), administrative purposes etc)

**3 If there are any aspects of quality that prevent you from using the IDBR, please list them?** (please list all that apply and please be as specific as possible eg availability of relevant size information (turnover, employment etc), out-of-date, coverage issues, business structural issues, contact details, sampling mechanism, industrial classification, area classification etc)

**4 If you wish to make any other comments please do so here.**

Thank you.

## Respondents to the consultation exercise

Wendy Fader	ONS
Ian Richardson	ONS
Derek Bird	ONS
Ann Lewis	ONS
Ann Payne	ONS
Tessa Staples	ONS
Sanjiv Mahajan	ONS
Richard Dagnall/Phil Lewis	ONS
Ian Hill	ONS
Robin Youll	ONS
Frances Pottier/Jackie Sanders	DETR
Janet Dougharty	DTI
Mike Janes	DTI
Arthur Barnett	DTI
Margaret Sims	DTI
Willie Lister	DTI
Ian Gallagher	DETINI
Patrick McVeigh	DETINI
Martin Monaghan	DETINI
Lisa Taylor	DETINI
Hayley Harris/Vivienne Burch	HMCE
Ian Dale	DTI
Tony Williams	DFID
Graham Collett	MAFF
Iain Bell	DFEE
Stephen Hall	DETR
Michael Clary	DTI
David Williams	DETR
Nick Bonney	Orange
Alex Morton	South Lanarkshire Council
Peter Sturman	Tyne & Wear Research
Jill Tuffnell	Cambridgeshire County Council
Chris Evans	Agricultural Engineers Association
David Fleming	Applied Business Statistics Limited
Melvyn Hayward/Lisa Goodyear	Brick Development Association
Harvey Collyer	Iron and Steel Statistics Bureau
Geoff Noon	Machine Tools Traders' Association
Chrys Rampley	Road Haulage Association
Jo Piggott	Schema Associates Ltd
Marek Wojtan	Biscuits, Cake, Chocolate and Confectionery Alliance
Dev Virdee	ONS
Alison Bromley	MAFF
Philip Turnbull	Bank of England
David Vincent	ONS
Richard Bailey	IR

### Extract from Eurostat's Statistical Programme Committee Report on Register Quality

"The first three years after adoption of the Regulation, the efforts of the NSIs of the MS were aimed at the creation of BRs or the adaptation of BRs to the requirements of the Regulation. At the same time Eurostat tried to facilitate this by several measures.... ..and paid a specially attention to methodological issues, since this was recognised as a potential bottleneck for the implementation of the Regulation.

At the beginning the main aim was that the required types of units and variables were recorded, and that the relevant economic activities were covered. Given the starting position, ensuring that certain types of unit and certain economic activities were covered was bound to precede considerations of quality.

With this pragmatic approach much has been achieved.....generally speaking the implementation of the Regulation has been a success: the MS (at least the fourteen to which no derogations apply) all have a BR with the most important economic units and variables and a good overall coverage. Terminology has been harmonised. However, it has also become clear that the quality of the BRs is less than perfect in a number of cases and countries are far from harmonised in the delineation of the statistical units recorded. After the initial years of implementation, the time has come to focus on improving the quality of BRs; in fact, a number of MS have already plans for improvements. Eurostat initiated the discussion with the MS on a quality policy in June 1997.

The Regulation does not specify quality requirements. Although the requested registration of units, the variables and coverage are formulated in absolute terms, a 100% quality is clearly not the objective of the Regulation. Not only is it obvious that such quality would be extremely cost-inefficient and would cause a huge response burden on businesses, but also Article 5 of the Regulation gives some very general rules for register updating which imply a less than perfect quality. According to this article, "As a general rule, information obtained from administrative files or annual surveys shall be updated annually, other information being updated every four years". A procedure for implementing rules is established by the Articles 8 and 9.

It is the intention of Eurostat to reach an agreement with the MS on certain quality requirements and on the way to implement them. In accordance with the subsidiarity principle, the MS will remain responsible for designing their own updating strategy taking into account considerations of cost efficiency, data availability, national needs and business response burden, provided that the quality requirements are met. Those requirements would guarantee a sufficient level of quality and data comparability as needed for the European system of business statistics. This implies that quality requirements would pertain to the output of BRs rather than to updating procedures; as a consequence, they would focus on statistical units (enterprises and local units) rather than on BR input units (legal units).

- The quality requirements would have to respect the following principles;
- Subsidiarity as just explained;
- All quality requirements must be measurable;

- The quality requirements must be based on the needs of the European system of business statistics, in particular data comparability between countries and comparability between different types of business statistics;
- The quality requirements will be restricted to the units and variables whose registration is compulsory according to the Regulation, and to their coverage; there will be no requirements for optional data.

Quality requirements are presently being developed; only a general idea of what they will look like can be given for the time being. A distinction is made between requirements concerning the coverage of the BR, the delineation of statistical units, and the variables recorded.

The quality requirements concerning the coverage of BR must differentiate between units of different size, and will depend on the economic activity carried out. Obviously the coverage of large enterprises has to be more complete than for smaller enterprises, not only because of their economic significance, but also for reasons to lighten the administrative burden on SMEs. But for those services industries where small enterprises dominate, the norms have to take this feature into account as well. The requirements would typically specify a minimum percentage of coverage for each stratum. This percentage would ideally be expressed in terms of value added.

The requirements about coverage can not be limited to the areas about which business statistics are compiled traditionally. The Regulation on structural business statistics has a much wider scope already .... and the BR Regulation itself stipulates a wide coverage, which was chosen with an eye on the future. Moreover, BRs are also important for other statistics than structural business statistics, such as business demography, employment statistics and, indirectly, national accounts. For similar reasons the coverage requirements also have to include small units; BRs are a crucial tool for SME statistics.

Although the level of coverage of units is easy to define, it is very difficult to measure, because this supposes knowledge about the extent to which units are not registered. For this other data sources are needed, and they may not be readily available.

The need for quality norms concerning the delineation of statistical units, in the first place the enterprise, comes from its impact on statistical measurements.[Earlier parts of this report] also showed the need for quality norms in this area.

As was the case with coverage, and largely for the same reasons, the quality requirements will have to differentiate between enterprises of different size, and will depend on the economic activity of the enterprise; they are meant to specify a minimum degree of correct delineation for each stratum. Unfortunately, developing quality standards for unit delineation is very complicated. It is very difficult to define what is a correct delineation and to measure it, considering the fact that BR updating procedures vary widely among countries and depend on the use of national data sources. There are also a number of technical problems when this aspect of quality is to be monitored statistically. And it is clear that in order to formulate quality standards for the delineation of units in some services industries, methodological recommendations are a prerequisite.

Quality requirements for variables can of course only refer to variables whose registration is compulsory. The most important requirements will concern the economic activity (NACE

Rev.1 code at four-digit level) and the size-class in terms of persons occupied, for the enterprise and the local unit.

Quality standards for the economic activity code can be formulated with relative ease, although their measurement is not as easy. Quality requirements for economic activity can be differentiated according to size-class. They would have to apply to the entire BR for the same reasons as given at the coverage requirements.

Requirements concerning the correctness of the size-class have to take into account the specific factors that determine the quality of this variable, in particular the availability of administrative files. In some countries it is very difficult to obtain data on the number of persons occupied as distinguished from the number of employees; in some MS data from social security registers can only be obtained with a long delay. Such factors are bound to influence the quality requirements; and for reasons of business response burden they should.

Several steps can be distinguished in implementing the quality policy. The first step is, as discussed above, to formulate and agree on the quality requirements. These requirements have to be realistic; the level of ambition must be a feasible aim within a period of five to ten years for a large majority of the MS. The ambitions cannot take into account only the need for good quality BRs, but must also be based on the present state of the BRs as described in chapter 2 and improvements that have already been planned. Some of the planned improvements are:

- Concrete plans to include additional NACE Rev.1 sections to their BRs;
- The last missing statistical unit (in one MS) will be added;
- Plans to add one or more new variables;
- To improve the quality by the use of additional administrative sources.

The second step is monitoring the quality of the BRs; if quality standards are agreed on, their implementation has to be checked. Ideally the quality of BRs is systematically measured, for instance by means of a regular quality survey 16 , but unfortunately this tends to be very costly.

In a number of cases the quality will have to be assessed by indirect methods, such as comparison with data from other sources or analysis of the way administrative files are updated. Although measuring the quality of BRs is often problematic, it is important to agree not only on quality standards, but also on ways to assess the actual quality of BRs. The next step is to agree on a procedure of reporting about the quality of the BRs to Eurostat. The contents and frequency of such reports need to be established; they need to be standardised. This, again, may be a complicated and time-consuming affair, as experience with the Regulation on structural business statistics has shown.

Nevertheless, a reporting arrangement is an essential part of any quality policy. If there is agreement on quality requirements, the measurement of quality and the reporting procedure, this should be formalised in accordance with the provisions of the Regulation. Article 8 of the Regulation stipulates that implementation rules shall be adopted in accordance with Article 9, which describes the procedure to be followed. As part of this procedure the SPC will give its opinion before the implementation rules are finally adopted. The implementation rules, consisting of quality requirements, the quality measurement and the reporting arrangements, will be subjected to this procedure.

In addition to the formal implementation of the quality policy as described above, there is more that can be done. Eurostat tries to stimulate the spread of good practices; there are many things that MS can learn from each other in the field of BR quality improvement, in particular in making optimal use of administrative files and keeping the administrative burden on businesses as low as possible.....

In a study on the implementation of enterprises in the business registers in three MS..., the effects on some key statistical variables were investigated. The study did not include the most simple cases but there are more complicated cases than those studied. The study showed that for all 21 enterprise groups investigated, the delineation was done differently in the three countries according to their profiling rules, resulting in inconsistent statistics.

The differences in the delineation have a very serious impact on data comparability, and it urgently calls for standardised application of concepts and definitions in order to reach harmonised and comparable business statistics.

Between Eurostat and all Member States there is a coherent view on the need for quality aspects in the Business Registers, but many Member States have also expressed requests for further guidelines and harmonised definitions.

The first steps have been taken towards collecting documentation on the profiling rules in the Member States and to develop a "case law" in order to establish methodological support for a more harmonised European statistics. The need for application of quality standards is recognised and supported by the Member States"



## IDBR Quality Questionnaire

Below are a number of questions that will give us an insight into the best ways of measuring the quality of the register. In order for BRU to address these quality issues could you possibly complete the questions below which cause you concern in terms of **quality** and **impact**.

At the end of each question there is space for additional comments on each particular aspect. Please feel free to comment.

A sliding scale (1-5) using a tick box method has been applied whereby 1 equals excellent and 5 equals poor in terms of **quality**. From an **impact** perspective 1 equals very important and 5 equals minimal importance.

For example, if you think that the quality of the SIC classifications (a) on the IDBR is excellent then place a tick in Box 1. If the quality of SIC classifications (b) has minimal impact on you as a user then you need to place a tick in box 5.

**Please complete the following details and return to Claire Powell by the 29<sup>th</sup> of October, 1999.**

**Name:**

**Department:**

### Question 1

a) What do you think of the quality of SIC classifications on the IDBR?

1	2	3	4	5
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

b) How much impact does the quality of SIC classifications have on you as a user?

1	2	3	4	5
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Additional Comments....**

### Question 2

a) What do you think of the quality of employment counts on the IDBR?

1	2	3	4	5
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

b) How much impact does the quality of employment counts have on you as a user?

1	2	3	4	5
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

**Additional Comments....**

**Question 3**

a) What do you think of the quality of the IDBR in terms of undercoverage (this refers to the treatment of inquiry stop markers 6 & 7 on the IDBR)?

1	2	3	4	5
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

b) How much impact does undercoverage have on you as a user?

1	2	3	4	5
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

**Additional Comments....**

**Question 4**

a) What do you think of the quality of the IDBR in terms of overcoverage (covers issues such as birth and death lags, falsely active units and unit matching issues)?

1	2	3	4	5
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

b) How much impact does overcoverage have on you as a user?

1	2	3	4	5
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

**Additional Comments....**

**Question 5**

a) What do you think of the quality of the IDBR in terms of covering structure changes?

1      2      3      4      5  
           

b) How much impact does structure changes have on you as a user?

1      2      3      4      5  
           

**Additional Comments....**

**Please use the space below to comment on any areas/aspects of the IDBR you feel should be addressed....**

### Estimating IDBR Undercoverage Arising from Birth Lags

In order to predict the number of missing businesses we need to rely on a model. Our approach is to formulate a model by looking at the number of births and the length of birth lags in historical data and to use this model to predict forthcoming numbers and lags. Clearly, this approach cannot take unforeseeable structural changes into account. The methodology introduced below can, however, accommodate economic cycles and seasonal effects that have been observed in historical data. As the usable data we have go back only to 1995 we have not attempted to model business cycles.

**Table 1. Number of births per lag (in months) and birth month. Unobservable cell counts in the triangle. Partially unobservable cell counts are indicated with a  $\geq$  symbol.**

	Lag						Sum
	0	1	2	...	38	>38	
<b>Jan, 95</b>	5,444	4,982	1,910	...	$\geq 6$	–	16,054
<b>Feb, 95</b>	5,333	4,069	1,280	...	–	–	13,425
M	M	M	M	O	M	M	M
<b>Jan, 98</b>	7,783	4,102	$\geq 1,346$	...	–	–	13,231
<b>Feb, 98</b>	7,075	$\geq 3,087$	–	...	–	–	10,162
<b>Mar, 98</b>	$\geq 5,888$	–	–	...	–	–	5,888
<b>Sum</b>	226,582	156,517	61,346	...	6	–	549,386

Table 1 shows the birth lag distribution for businesses born between 1 January 1995 and 31 March 1998. The rows of the table show the numbers of businesses that were born in each month. The columns are birth lags measured in months. If the time from actual birth to IDBR birth (when the business came onto the frame) is less than one month then the birth lag is counted as zero, and so forth. This implies that the longest birth lag we can observe is 38 months. The counts represented by dashes in the table are *unobservable* (unless we have data that go beyond our final date, which is 31 March 1998). The fact that we cannot observe all births that have occurred implies that our data are *truncated*. The problem is to estimate the number of businesses that have been subjected to truncation, that is, to estimate the undercoverage due to birth lags. On 31 March 1998, the undercoverage is the sum of the unknown counts of Table 1. To briefly outline the approach we have chosen, first we fit a log-linear model to the fully observed counts in Table 1, then we use the model to predict the unobserved and partially unobserved cell counts. We neglect unobservable counts with 38 months birth lag or longer.

## Topics Included in Statistics Netherlands' Business Register Customer Satisfaction Questionnaire

The main subjects of the Statistics Netherlands' user questionnaire are listed below. There are two questions on every topic – one about the standards required, and one about perceived standards. (The DGC is Statistics Netherlands' equivalent to ONS's Business Registers Unit.)

<b>DGC should ...</b>	<b>Opinions about the performance of DGC</b>
Provide an up to date business register	Information on the business register is up to date
Provide requested data promptly	Information is delivered promptly when it's requested
Consult its customers	DGC consults its customers
Provide a business register that contains correct information	Information in the business register is correct
Be willing to listen to suggestions from its customers	DGC is willing to listen to customers' suggestions
Ensure that information in the business register is accessible	Information in the business register is accessible
Give high priority to customers' requirements	DGC gives customers' requirements highest priority
Consult customers when changes are made to the business register	DGC consults customers when changes are made to the business register
Endeavour to understand its customers' requirements	DGC makes an effort to understand customers' requirements
Provide metadata with data	Requested data are provided with metadata
Ensure that the customer knows which member of DGC staff to contact in case of a problem.	Customers find it easy to trace the appropriate (responsible) employee when problems occur
Select the most appropriate communication channel when talking to customers	DGC uses the most appropriate method of communication when talking to customers
Ensure that the data provided exactly match the request	The data that are delivered match the request exactly
Communicate with customers regularly	DGC communicates regularly with its customers

Communicate with customers frequently	Communication between DGC and customers is frequent
Allow customers to contribute to decision-making	DGC allows customers to contribute to decision-making
Supply information to customers concerning the quality of information on the business register	DGC supplies information to customers concerning quality of business register data
React promptly to customers when problems occur	DGC reacts quickly to customers when problems occur
Ensure customers know what facilities are available	DGC ensures that customers know what facilities the business register offers them
Co-operate with customers	DGC always co-operates well with its customers
Abide by its agreements with customers	DGC always abides by agreements with its customers
Communicate fully with customers	There is full communication between DGC and customers
Provide clear information relating to the information stored in the business register	Information about the data stored in the business register is readily available
<b>Members of staff in DGC should ...</b>	<b>Members of staff in DGC ...</b>
Always do their best for customers	Always do their best for customers
Always be ready to help customers	Are always willing to help customers
Above all, consider their customers' requirements	Consider their customers' wishes
Take the lead in solving customers' problems	Take the initiative when it comes to finding solutions to their customers' problems
Deliver the requested information at the agreed time	Deliver requested information at the agreed time
Always make time for customers	Always make time for their customers
Always listen carefully to customers	Always listen carefully to their customers
Be willing to reach an agreement with the	Want to reach an agreement with the

customer at the time of their request

customer when a request is made

Be interested in hearing customers' wishes

Are interested in their customers' wishes

Possess considerable statistical knowledge

Possess considerable statistical knowledge

## Comprehensive Business Directory: ONS Managers' Briefing Paper

### Purpose of this Briefing

This paper outlines the work that has been undertaken to research the prospect of a single business register for general government use and the current work on the proposed option for a Comprehensive Business Directory (CBD).

It is relevant to raise awareness now in ONS as the emerging findings of this programme of work has led to a Capital Modernisation Fund bid to assist in developing the proposal for a significant Infrastructure development programme. In addition the project is closely aligning itself to other cross-government initiatives and therefore it is becoming a focus of attention.

### *Introduction*

#### **The Government's Manifesto**

The Government's General Election Business Manifesto 'Equipping Britain for the Future' stated the need to 'reduce the demands for information that central Government makes on business by developing a common database available to all Government departments'. A whole range of government information and communication technologies makes it ever easier to gather, process and disseminate information and these need to be harnessed to deliver the manifesto objective.

#### **Background: The Business Information in Government (BIIG) Study**

The 1999 Business Information in Government (BIIG) study - referred to in the Modernising Government White Paper - investigated the prospect of a common database. The research investigated the potential for a single business register to facilitate the delivery of services to the UK business community and the need for a single business identifier. Recommendations would be made proposing the most appropriate model of such a central facility. The research was funded under the Invest to Save budget, project managed by the Office for National Statistics (ONS) in partnership with Inland Revenue (IR), HM Customs & Excise (HMCE), Companies House and Department of Trade and Industry (DTI).

The research identified a number of significant conclusions that were notified to the Information Age Government Champions, which led to a secondary study project regarding the implications for future policy for electronic service delivery. The key findings suggest that:

- Government should consider the full range of issues relating to the management of business information in Government in light of the broader agenda of 'Modernising Government'.
- A single business register system would not be capable of supporting all government purposes in every business operation.
- Business identification could not be dependent upon a single identifier.
- New views could be formed of the business community through controlled data matching and sharing.



- There exists a need for an ‘intelligent gateway’ to distribute commonly used data received from businesses to relevant government departments.

The study concluded that a single database and numbering system meeting all government services is not feasible but a central facility bringing together disparate data for a number of determined purposes through an intelligent and secure process is achievable.

### *The Solution: The Comprehensive Business Directory*

#### **What is the CBD?**

The conclusions from the BIIG study have since been taken forward, specifically to establish a central facility, termed the Comprehensive Business Directory (CBD). It will provide a range of services by acting as a business data locator containing information on what and where data, relating to the business community, is used and held in administrative data sources.

The facility would operate in a controlled environment to protect the use and release of information by the participating departments or local authorities. This recognises that each participator is strictly governed by legal and policy controls. However, within these safeguards, the CBD facility will need to have the flexibility and capacity to allow different viewers to interpret data for their own purposes. For example, one such view may meet the needs of the Treasury Banking Review team. An extract from their official response makes mention:

"The Cruickshank Banking Review (Recommendation 39) highlighted the essentiality for a Comprehensive Business Register. It stated that the register should include information on the location of business, sector, date of business trading, turnover and VAT record. HM Treasury have accepted the concept of a register that draws on sources of information from various government departments. It has asked the Office for National Statistics (ONS) to co-ordinate an inter-departmental project to make this facility available across Government."

#### **CBD Partners**

This inter-departmental project will be led by 4 stakeholder departments: Office for National Statistics (ONS), Inland Revenue (IR), HM Customs & Excise (HMCE) and Companies House. Once the legal constraints and process controls of divisions within the stakeholder departments have been addressed the partnership will expand to those potential partners established from the BIIG study. They include Office for Government Commerce (e-procurement), DTI/SBS, HSE, MAFF, DfEE, Norwich City Council and Local Authority (IDEA).

#### **The Benefits of the CBD**

It is not possible in this paper to state the benefits of a central facility for each government department. However, a number of key benefits have been identified and include:

- Greater availability of information within government.
- Improved information quality derived from more data sources.
- Reduction of burden placed on business-government interaction.
- Provide added value to existing data sets.
- A single point of access to publicly available business data sets (e.g. telephone directory information).
- Centrally managed security safeguards for data sharing across the GSI.
- Enabling on-line services to businesses.
- Improved government response and effectiveness in business-with-government service by supporting new process working.

The above benefits highlight what can be offered across Government once a joined-up CBD service is established.

#### *CBD Project Coverage FY2000/01*

The project may be seen as an IT development initially but the issues it raises require closer examination of government processes, legal and policy structures and the standards that would need to be adhered to that are being set within the Office of the e-Envoy. The project has two stages that will embrace the majority of these issues. In addition the project's personnel will participate in forums and with project teams working on other government change processes.

#### **Technical Development Stage**

There is no equivalent facility available in central Government, and therefore the CBD project in 2000/01 is a beacon project that will test the feasibility of building a virtual data warehouse service, providing controlled access to administrative business data. The proposition is to establish a set of intelligent matching and data managing tools on the Government Secure Intranet (GSI). With this strategic facility being hosted on the GSI, it will maximise use by all GSI members thus, enforcing the coherence of public sector work. Specific issues to be explored are:

Metadata: Acting as a data locator the CBD will require context information about the data that it finds. It will store:

- Information about each data item, including format, allowable values and derivation.
- Technical information about where the data item can be found and how the hub is to access it.
- Privacy information that informs the hub who is allowed to have access to the data item, and for what purposes.

Work is progressing on this component with close liaison with departmental and other inter-departmental projects with a metadata interest.

Matching: Data from different sources might contain identifiers (such as Company Registration Number) that allow records to be clearly matched. However, matching will require sophisticated software tools that can match upon secondary data items contained in

their record set. The CBD team are working very closely with the Data Matching project of the IDBR to evaluate a number of data matching packages, and are building a "demonstrator" matching facility to help evaluate the business benefits of matching to stakeholders.

**Data Management:** The hub needs to be built on effective data management tools, able to interface with contributing departments' datasets, and deliver the data seamlessly. The data management tools need to work very closely with the metadata tools, particularly those aspects relating to relevant permissions on the data. The CBD team is conducting an appraisal of appropriate data management tools, looking at what is being used for analogous projects across the public and private sector, and by discussions with potential suppliers.

### **Legal and Policy Stage**

The project will research and assess the legal implications of a CBD facility. It will look at privacy and information technology security issues, the Data Protection Act, the Human Rights Act and key legislation and directives in UK and Europe with regard to data access, specifically for business data.

Two experts in the fields of E-privacy and departmental legislation have been recruited to the team to facilitate a closer investigation of the issues that a CBD facility raises. The stage will specifically look at the legal powers and legislative structures underpinning each of the partner departments viewed from the proposed operational model for a CBD.

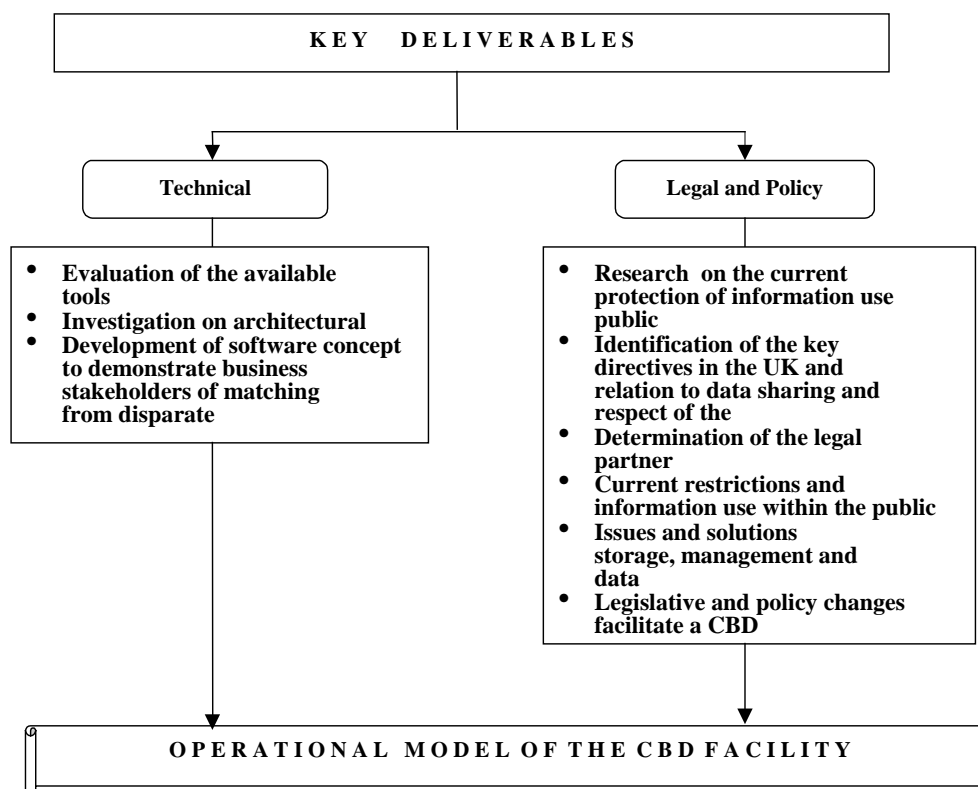
The conclusion to the investigation will be the identification of the legislative and policy changes required for a CBD service to proceed.

Progress is being made on the following areas of the legal/policy work:

- A review of legislation and policy within the partner departments relevant to the operation a CBD.
- An impact assessment of the wider E-government policy implications of data sharing across government.
- A report that draws a comparison between privacy and information technology security.
- A report addressing legislation and regulatory issues and recommendations.

### **Development of the CBD: Key Deliverables**

The key deliverables for the CBD project have been highlighted in the diagram below. The product checklists for the technical development and legal/policy research are presented in Appendix A and B, respectively.



### *Other Government Related Initiatives*

There are a number of related government initiatives aimed at information knowledge and handling. Following from these there will be consequential changes in departmental IT systems and the legislation. For this reason it is imperative that the CBD is recognised as a beacon project and that the CBD team should work closely with the key initiatives.

A proposal is with the Office of the e-Envoy that suggests an alliance between the projects outlined below, so as to benefit from shared ideas, knowledge and resources, reduce duplication of effort, and develop the projects in relation to a common programme schedule. Some of the key projects are:

- *Policy Framework on Data Sharing* led by the Performance and Innovation Unit (PIU), Cabinet Office. This project aims to establish a government-wide framework for data sharing (citizen level). The project covers data sharing within the public sector and between the public and private sectors.
- *Information Asset Register (IAR)* led by HMSO, Cabinet Office. A meta-database will be built to register different government department's information resources (intelligent electronic mapping).
- *Catalogue of Data Sources (CODS) and GUIDE project* led by HMCE. A meta-database will be developed to allow data access and matching across government for fraud investigation and compliance.

- *Review of Government Processes* led by IR. This will establish a map of all processes across government with aid from the deliverables of the IAR, GUIDE and CBD projects.
- *Government Suppliers Information Database* led by OGC. This is an electronic service that allows organisations to register once in government, their interest in being a supplier of goods or services to government.

**CL**assifications **MO**delling and **U**tilities **R**esearch (CLAMOUR) is a European collaborative project that aims to find a better way to classify businesses. This will lead to improvements in the collection and dissemination of information about businesses.

The purpose of this project is to ensure harmonisation within the members of the European Union (EU), in both the collection and management of data.

### **Who are involved?**

The project is part of a European Commission research and development programme. Five EU countries are jointly leading the work. These countries are:

- Denmark
- Finland
- France
- Netherlands
- United Kingdom

A French company, Lexiquet, is also part of the project team. The project will produce a range of outputs that will be of benefit to classification systems across the EU as a whole.

### **The schedule of activity**

The ideas for the project developed during 1999 and were formally accepted by the European Commission in January 2000. Work began shortly after and the project will conclude in December 2001.

### **Research Activities**

The work has been divided into four related areas of research.

The four research areas are:

- User Needs
- Foundations
- Linguistics
- Future Tools and Systems

### ***User Needs***

Information about businesses is collected and classified so that an accurate picture of national economies can be produced. Both government and research organisations also have information to inform policies.

It is important to establish how information will be used and what form of classification provides the best value for users. Each National Statistics Institute (NSI) will interview a sample of users of government information. Questions will ask how

information is currently used and what changes in use may occur in the future.

In planning for the future, care must be taken to protect the needs of those who need consistent information across time. Requirements for change must also give consideration to the burden on information suppliers – often those in the business community.

### ***Foundational work***

The foundational work examines the fundamental building blocks of classification and is divided into three work packages:

- Constructing a model of the structure and activities of businesses
- Applications of the model for classification systems
- Applications of the model for statistical units

The outcome of the *User Needs* work will feed into the *Foundational Work* by informing those responsible of the applicability of NACE (the European classification standard) and how business structures and activities should be described.

## **Linguistics**

To improve the speed and the quality of information classification the project will develop linguistic methods and tools that are able to recognise the exact, or at least, the best meaning of descriptions. Existing tools are often deficient in the recognition of text written in free language.

It should be possible to identify an industrial organisation by the profile of the processes, skills and other factors. Within a given language and across the European community there are nuances that imply different meanings. This work will look at how best these differences can be minimised and perhaps lead to interactive systems providing intelligent guidance. The linguistics work covers four work packages:

- Context based semantic disambiguation
- Compound term processing
- Highly structured description
- Assisted matching between correlations

## **Future Tools & Systems**

Resulting from the preceding work will be information that will inform those involved in developing new products and services.

A key piece of work is the development of an electronic questionnaire that will assist in data collection from businesses.

The electronic questionnaire will be tested extensively in at least three member states.

## **Further information and contacts**

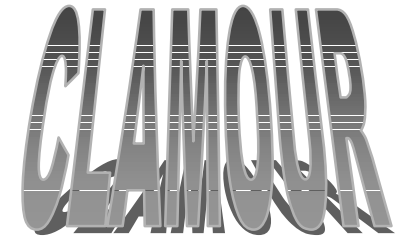
CLAMOUR website:

[http://www.statistics.gov.uk/methods\\_quality/clamour.asp](http://www.statistics.gov.uk/methods_quality/clamour.asp)

Contact us by email:

[clamour.coordination@statistics.gov.uk](mailto:clamour.coordination@statistics.gov.uk)

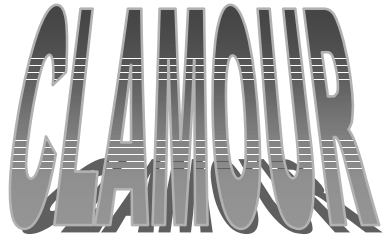
Contact by telephone: 0044 1633 812371



*CLassifications MOdelling and Utilities Research*

A European Collaborative Project to find a better way to classify businesses

“Making the most of your data”



## A dissemination and use plan for the results

The CLAssification MOdelling and Utilities Research project is a European collaborative project to find a better way to classify businesses.

*The project has three basic objectives split in three sub-projects, each comprising a number of work packages:*

- **Foundations** of activity classifications and the links they have with statistical units,
- **User needs** for activity classifications,
- **Linguistic** methods to improve understanding of the meaning of classification descriptions and other information about businesses.

*The dissemination and use plan covers all sub-projects. Both the content of the various parts and the kind of researchers involved differ between **Foundations** and **User needs** research on the one hand and the **Linguistics** research on the other. Consequently, different approaches are described for the dissemination of results.*

The dissemination and use plan for CLAMOUR has three elements:

- **Key products**
- **Possible users** of these products
- **Communication** with the users to ensure that the products will be used

## Foundations and User needs

### Key products

Reports:

- Model for the structure and the activities of businesses
- Proposal for a system of alternative classifications and a methodology to compare tree structures
- Guidelines on how to construct and maintain an activity classification
- Algorithms to construct statistical units based on the model
- Users' needs

Demonstration system:

- Electronic questionnaire to obtain information from businesses.

Delivery of these products will start in 2001.

Possible users

The two main types of users are:

- Those **in charge of developing or maintaining activity classifications**, who will be interested in:
  - what should a good system of activity classifications look like?
  - how should an activity classification be developed and maintained?
- who are the users of activity classifications and what demands do they have?
- Those **in charge of business registers**, who will be interested in:
  - what information about a business should be included in the business register?
  - how can businesses activities be coded?
  - how should statistical units be constructed in the business register?
  - a tool that can be used to get the necessary information from businesses

## Communication

The following steps are proposed to communicate the key deliverables to the users:

To communicate to those **in charge of developing or maintaining activity classifications**:

- Presentation to the Eurostat NACE/CPA Working Group: October 2000 and further dates
- Presentation to the UN Classification sub-group meeting in New York: 2001
- Presentations of the relevant deliverables to the UN expert group and the classifications subgroup

To communicate to those **in charge of business registers**

- Presentation at the Business survey frames roundtable in Auckland (New Zealand): November 2000 then at following Roundtables in October 2001 and 2002
- Presentation at the Eurostat Working Group on Business Registers: June 2001 then in other working group meetings
- Presentation to the joint Eurostat/UNECE Business Registers Committee meeting (in Luxembourg or Geneva): June 2001
- Presentations for relevant government bodies in the participating countries (e.g. the Chambers of Commerce)

## Use

*The partner countries within the project have identified uses for the research:*

- Finland: The output of the **user needs** research will be used for improving the information value of data used to classify and describe the structure of business life in a changing society. This is particularly important for such new branches as the IT sector and the entire service sector.
- France: The output of the **foundations** sub-project will be an input of the Statistical Unit Working Group within the French statistical system. The business register would be interested in using the electronic questionnaire.



- UK: The output of the **user and business needs** research will be used in the creation and management of a comprehensive business directory for statistical and administrative purposes.
- Denmark: The output of the **user needs research** will be used in the management of the service activities conducted by the Danish business register for statistical, administrative and commercial purposes.
- Netherlands: The model will provide input to the development of an activity description for the single business register that is under development. A system like the Blaise questionnaire may be used for data collection for that register.

## Linguistics for classification

### Key products

Reports:

- Specification on how to remove ambiguity from words by considering the role of a given word and the meaning of the surrounding words
- Recognition and construction of compound terms
- Syntactic analysis of highly structured sentences, found in descriptions of business activity
- Assisted matching between classifications

### Possible users

The three main types of users are:

- **Classifications specialists and classification custodians** in order to verify the coverage and the overlap of the semantic content of their descriptors/headings/explanatory notes. The methods and algorithms could help to verify the integrity of classification descriptors, particularly at the drafting stage.
- Those **in charge of business registers**, and other government bodies that make use of them, will be interested in methods to improve automatic or

software-assisted coding of business descriptions provided by enterprises

- **Classification users and linguistic specialists**, who need to correlate their internal references with external ones, will be interested in how the methods can contribute to the semiautomatic building of correlation tables.

### Communication

The following actions are proposed to communicate the key deliverables to the users:

To **classification specialists and classification custodians**, presentation of deliverables and demonstrations through the implementation of the methods and algorithms in an existing specialised classification software at:

- Eurostat NACE/CPA Working Group: Autumn 2001 and/or following meetings
- UN Classification Group meeting in New York: Spring 2001
- UN Classification Expert Group and its classification subgroup: end 2001 or during 2002

To those **in charge of business registers**, simplified presentation of deliverables and demonstrations through the implementation of the methods and algorithms in an existing specialised classification software at:

- Eurostat Working Group on Business Registers: end 2001 or later
- relevant government bodies in the participating countries (e.g. the Chambers of Commerce)
- Roundtables on business survey frames

To large **businesses** and European professional bodies, presentation of deliverables and demonstration through the implementation of the methods and algorithms in existing specialised classification software

- at Eurostat meetings: end 2001 or later
- through direct contacts with appropriate enterprises

- To **linguistic specialists**, presentation of deliverables and demonstrations through the implementation of the methods and algorithms in existing specialised classification software:
  - in international conferences: end 2001 or later
  - through specific meetings organised by the European Commission

### Use

The partner countries within the project have identified uses for the research:

- France: The results of the Linguistic sub-project will be implemented in the classification software (NOMENCLA) in French and English.
- UK: The results will be used to assess existing tools and to develop better automated tools for classification of business descriptions.

The partners in the CLAMOUR project are:

Statistics Denmark  
 Statistics Finland  
 INSEE, France  
 CBS, Netherlands  
 Office for National Statistics, UK  
 LexiQuest  
 Roland Rousseau

Further information on the project is available on the UK National Statistics web site:

[www.statistics.gov.uk/methods\\_quality/clamour.asp](http://www.statistics.gov.uk/methods_quality/clamour.asp)



The PRN start for the next period is calculated by counting forward ‘round( $n/r$ )’ businesses, and then adding 0.000000001 to the PRN for the last unit dropped. This new number now becomes the starting point for the next period. Then, businesses are selected along the line in the same way until the required sample size is achieved. This will introduce round( $n/r$ ) new businesses to replace those rotated out.

In further periods the only difference is that the number of businesses to be rotated is round( $n/r$ ), with the fractional remainder from this becoming the new  $e$ . This will ensure that rotations are completed in the specified period regardless of the speed of rotation that this entails.

To take the example above with  $r = 5$ , we have  $n/r = 1$ , so we drop unit 8. Our new PRN start is recalculated automatically to be 0.548379319 (0.000000001 after the PRN of the last unit dropped) and we select the next five units: units 2, 7, 5, 6 and 9.

### *Births and deaths*

Births and deaths are dealt with naturally within the system. At each sampling point, only those units that are live in the current period will appear on the PRN line. Sampling carries on in the way described already among those units that are live and eligible for selection within the particular stratum. Reclassifications (of size, industry etc) are regarded as a death from the original stratum and a birth in the new stratum. In general, births and deaths won’t have any effect on the definition of the sampling interval. However, in some cases, rotation can speed up or slow down if there is an imbalance between births and deaths in a particular stratum.

### *Overlap control*

Overlap is controlled by strategic definition of PRN start points. To minimise overlap between different samples (negative coordination), PRN start points are defined to be different from each other. Then, as long as the relative speeds of rotation are the same, sampling intervals move along the PRN line and remain distinct (non-overlapping).

## **Extracts from Statistics Canada and ABS Progress Reports at the 2000 Roundtable**

### **Statistics Canada**

Profiles are initiated when a signal from a data collection centre indicates a potential major re-structuring of a business. For example, when a business sells all or part of its operation, a reaction profile is required to investigate how that acquisition will be treated in the purchasing company. A profile may take hours, days and sometimes weeks to complete, depending on number and complexity of the changes involved and the difficulty contacting the business to obtain the needed information.

Profiles can be divided into 3 broad categories. The first category is extremely large, complex and highly volatile enterprises that are absolutely critical to business surveys in Statistics Canada. There are currently about 500 enterprises in this category. The second category is very large and complex enterprises that have major impact on business surveys. There is a very gray line between category one and category two. Because category one enterprises get special treatment by Business Register staff, the cut-off is established based on the resources that can be brought to bear on each category. The last category is smaller less complicated enterprises.

### **Profiling - Extremely large, complex and highly volatile enterprises**

To deal with the first category of enterprises, a special section was established in the Business Register in January 1998. Seven teams were created. Each team is made up of one business profiler and one business analyst who have a portfolio of 60-70 enterprises for which they are responsible. They maintain on-going profiles for all the enterprises in their portfolio as opposed to periodic reviews of legal and operating structures. By continually reviewing sources of information on these enterprises such as newspaper clippings, Internet sites and survey feedback, profiling teams are able to keep the frame very up to date for these enterprises. When a major change occurs in one of these enterprises, the profiler will visit the business and collect all the necessary information to make the frame up to date.

Statistics on the first full year this unit has been in production are now available and clearly demonstrate that the restructuring of this operation has been tremendously successful.

From January to December 1999 a total of 416 profiles for extremely large, complex and highly volatile enterprises were committed. A total of 11,948 events (business changes) were processed against the business register associated with the 416 profiles committed. This is an average of 29 events per profile. 56% of the companies visited in 1999 had their profiles committed on the business register in less than 40 workdays. 70% were committed within the 60 workdays. 92% were committed within 80 working days. Most of the cases requiring more than 60 working days were companies that had never been profiled before by the Business Register and therefore required much more work to portray an accurate profile. Quality of the work in this new program continues to improve. Because companies are being profiled on an ongoing basis they are more up to date and hence there is less confusion / controversy on what the profile should look like. All profiles are reviewed by survey programs prior to the changes being committed on the business register to ensure that they understand and concur

with the changes proposed to the legal and operating structure of the business. Only 6% of the proposed profile changes were questioned by survey divisions.

### **Profiling - Very large and complex enterprises**

The second category is handled by experienced Business Register staff that conducts reaction profiles by phone as opposed to personal visit to the company. On-going profiles for these enterprises are not maintained. A reaction profile only occurs when triggered by survey feedback. Not much progress was made last year with this initiative. Many experienced Business Register staff were lost last fall due to internal competitions within Statistics Canada. New personnel did not start arriving in the Business Register until May 2000 and it will take another 4 to 6 months of orientation and training before they are fully functional.

### **Profiling - Smaller, less complex enterprises**

The strategy for this category is to have the data collection centres in the regions handle the reaction profile in its entirety. Traditionally, the regions would contact the business in question to collect pertinent information and then pass that information on to the Business Register to update the frame. This caused time delays and sometimes resulted in duplication of work when further information was required from the business and a re-contact with the business was made. Now the regions will not only collect the profile information from the business but will also update the frame directly.

A pilot project was launched in 2 regional offices last July. Based on the results of the pilot project a decision was taken to make this an ongoing project in the regions. We will now have approx. 17 full time interviewers in the 2 regional offices doing small, less complex reaction profiles. We expect that the 2 regional offices will commit approx. 2500 reaction profiles this fiscal year. This will most certainly ease the burden on Business Register staff and allow them to focus on large complex businesses.

### **ABS - Extension of Profiling Program**

The introduction of The New Taxation System (TNTS) is likely to induce changes to the accounting structures of businesses. There has also been a major Review of Business Taxation (such as depreciation and capital gains tax laws) which is most likely to affect how larger businesses arrange their legal structures. The Government is yet to finalise the legislative outcomes from this review, however, both types of restructuring have obvious implications for the statistical units used by ABS.

The main aims of the ABS profiling strategy are to maintain the correct structures of large businesses on its Business Frame, assist in the collection of data from large businesses, and to ease provider load for large businesses. A major focus of profiling during 2000-2001, will be to gather information on actual and intended business restructuring, in the light of TNTS and the foreshadowed changes resulting from the Review of Business Taxation. This information will be used to fine tune changes to the ABS information and units model in the immediate future.

The profiling population is based on an algorithm which was designed to identify significant businesses with complex structures and to manage structural change in a methodical manner. This population of large businesses is maintained by the Large Business Unit (LBU), which has operations in each State and is managed from ABS Central Office in Canberra. The LBU

has four main streams of work - personal profiling; mail profiling; forms co-ordination and Key Provider Management (KPM).

The LBU is nominally responsible for the maintenance of enterprise groups with 100 or more employees. There are close to 5,000 enterprise groups on Inteframe which meet this criterion. Of these, only the most significant and complex are profiled each year. The next level, in terms of significance and complexity are profiled every second year, whilst large, single LE enterprise groups may be profiled less frequently.

In response to anticipated changes to business structures resulting from TNTS, the ABS has allocated extra resources to increase the number of large businesses which it profiles each year. During 2000-2001, the ABS will be personally profiling 650 enterprise groups and mail profiling 1,850 enterprise groups. The comparative numbers of enterprise groups profiled during 1999-2000, were 485 and 1,400 respectively. Mail profiling is seen as a particularly effective and efficient method of profiling medium to large businesses, with relatively straight-forward business structures. By way of comparison, mail profilers update an average of 125 enterprise groups per year, compared with approximately 15 enterprise groups per personal profiler, although it should be noted that no co-ordination activities are undertaken for mail profile groups. Accordingly, the ABS has gradually increased the size of its mail profiling population.

## Unregistered Businesses

Industry SIC(92) division	All businesses			Unregistered businesses		
	Number (000's)	Employment (000's)	Turnover (£bn ex VAT)	Businesses (%)	Employment (%)	Turnover (%)
01	166.5	417.1	24.3	0.0	0.3	0.4
02	10.1	17.6	0.7	68.2	42.5	15.7
05	8.8	17.2	1.3	49.2	32.2	11.2
10	0.7	15.0	1.7	72.5	4.3	0.8
11	1.0	30.0	24.0	59.8	2.6	0.0
12	0.0	0.0	0.0	0.0	0.0	0.0
13	0.0	0.1	0.0	0.0	0.0	0.0
14	2.1	38.5	4.4	34.7	2.7	0.6
15	14.7	511.0	62.2	44.0	1.5	0.3
16	0.0	9.6	9.5	0.0	0.0	0.0
17	17.3	177.0	10.6	62.6	6.8	2.2
18	16.5	148.6	7.5	55.0	7.4	2.5
19	2.3	27.3	1.6	41.5	4.4	1.2
20	18.9	92.3	6.2	52.4	11.7	4.0
21	3.0	106.5	12.2	0.0	0.0	0.0
22	55.3	391.8	31.6	45.5	9.6	2.1
23	1.3	27.6	13.4	79.9	5.6	0.2
24	9.2	282.7	49.6	53.4	2.2	0.3
25	13.4	258.5	19.9	45.8	4.4	1.4
26	9.2	147.4	11.4	37.5	2.5	0.8
27	4.0	138.0	22.5	32.6	1.2	0.2
28	49.5	411.5	28.1	38.3	5.1	1.6
29	24.7	407.2	36.4	36.8	2.7	0.8
30	3.3	49.7	9.6	37.9	2.9	0.4
31	14.7	210.3	19.4	57.7	5.1	1.2
32	5.1	134.1	22.3	36.1	1.8	0.2
33	7.9	155.9	11.7	18.7	0.8	0.2
34	6.2	228.4	41.1	47.7	1.7	0.3
35	11.6	175.6	27.4	71.4	5.3	0.8
36	42.7	231.5	14.7	53.0	11.0	3.8
37	1.2	11.2	1.5	4.2	0.4	0.1
40	0.2	89.8	37.3	0.0	0.0	0.0
41	0.1	48.9	7.1	0.0	0.0	0.0
45	683.5	1524.4	123.6	69.9	32.5	11.6
50	93.4	600.0	137.7	18.2	3.3	0.4
51	133.4	1090.1	339.5	8.6	0.9	0.1
52	306.4	2725.6	214.9	24.0	3.4	1.3
55	154.4	1597.9	50.7	22.3	3.7	2.0
60	162.4	625.6	33.9	69.3	18.9	8.8
61	2.9	19.7	3.6	56.7	11.1	1.2
62	1.5	89.8	16.4	31.0	0.6	0.1
63	31.7	297.4	42.0	42.6	5.5	1.2
64	27.2	505.8	40.6	57.4	5.4	1.4
65	21.7	577.3	1397.0	56.9	2.3	0.0
66	5.8	224.7	230.8	68.9	2.2	0.0
67	32.0	240.5	442.1	54.7	8.6	0.1
70	85.6	311.1	34.8	27.5	12.6	3.4
71	16.7	150.7	19.1	12.5	2.1	0.4
72	169.6	454.3	37.2	29.2	12.1	2.9
73	11.5	49.2	3.2	75.0	20.9	6.2
74 (excluding 745)	517.1	2180.9	135.3	42.9	11.7	4.3
80	107.9	255.5	8.0	87.7	40.6	27.2
85	203.5	2107.3	68.3	73.4	9.0	4.8
90	6.2	45.6	3.7	69.2	9.9	3.5
91	23.2	92.7	4.4	68.8	20.1	8.5
92	230.0	619.7	42.5	72.2	28.9	10.2
93	127.9	352.9	13.2	33.2	12.3	8.3

Source: Small Business Service and IDBR, 1999

## Extract from ABS' Quarterly Business Register Quality Report

**Business Demographic Quality Measures****June Quarter 2000**

1. This report presents an agreed set of quality measures for use by ABS management. The graphs presented are:-

**Quarterly series:**

Graph 1 Number of GE Births from ATO and Number of MU Births and Deaths from Inteframe, March 1996 to June 2000

Graph 2 Survey of Employment and Earnings - Business Provision, June 1996 to June 2000

Graph 3 Total Number of Live GEs on the ATO Tape and Total Number of Live MUs on Inteframe, December 1995 to June 2000

Graph 4 Adjusted LFS and SEE Total Employees, All Industries, Including PSM Adjustments and Inteframe MU Total Employment, September 1994 to March 2000

Graph 5 Death Rate - Deaths in Reference Period, Quarterly Collections, September 1998 - September 2000

Graph 6 Death Rate - Deaths in Reference Period, Annual Collections, 1997/98 - 1998/99

**Monthly series:**

Graph 7 Number of GE Births from ATO and Net Number of MU Births to Inteframe, October 1997 to June 2000

Graph 8 Monthly Retail Trade Survey - Business Provision, July 1997 to June 2000

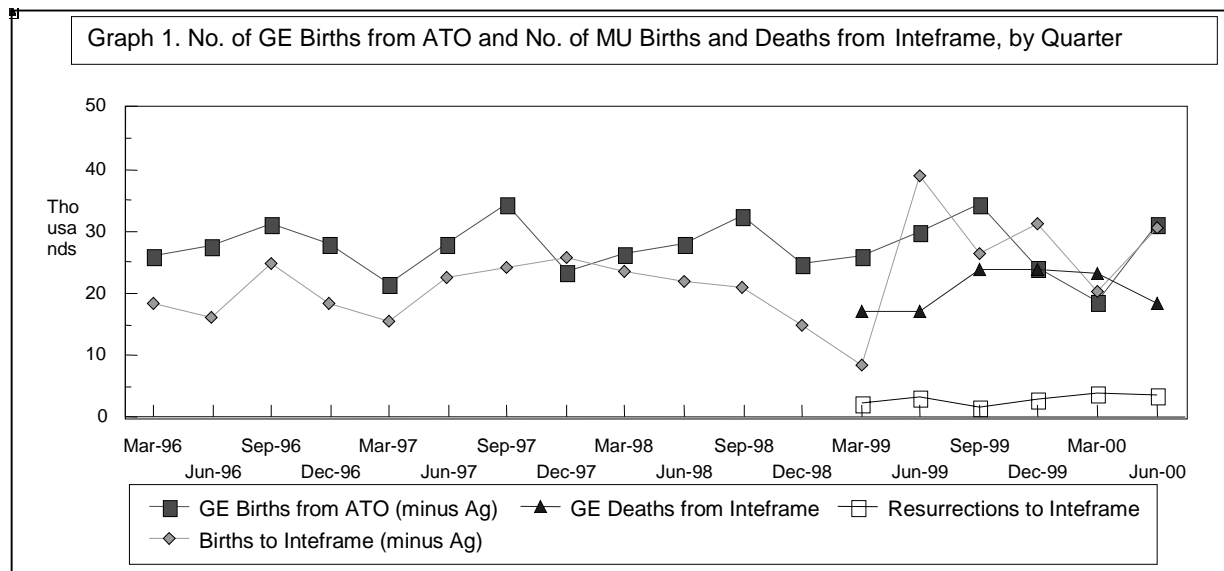
2. Graph 1 compares the flow of businesses into the ATO system and on to Inteframe, and provides an aid in the identification of potential administrative and system induced shocks. The cut over from the Business Register to Inteframe is an example of how birthing irregularities can be highlighted. Graph 1 also shows the number of units deathed on Inteframe, and the number of 'dead' units subsequently resurrected to Inteframe, since March 1999.



3. Graph 2 shows the Business Provision (BP) count for the Survey of Employment and Earnings and facilitates analysis of how well the BP adjusts for the lag effect caused by the time taken to load units to Inteframe. Graph 3 provides an analysis of the estimated number of live businesses on Inteframe. Graph 4 is a comparison of the adjusted employee levels for the Survey of Employment and Earnings and the Labour Force Survey against the level of Management Unit (MU) employment stored on Inteframe. It is presented as a guide to the way 'real' employment and 'benchmark' employment, as measured by Inteframe, are moving.
4. Graph 5, and Graph 6, present a comparison of death rates in sample across ABS collections that draw their frames from Inteframe. The aim is to provide a comparative assessment of the quality control processes used within and across collections.
5. Graph 7, like Graph 1, compares the flow of businesses into the ATO system and on to Inteframe, presented monthly. Graph 8 shows the Business Provision (BP) count for the Retail Trade Survey and like Graph 2 facilitates analysis of the BP adjustment for the lag effect caused by the time that is taken to load units to Inteframe.

### Births of Units (quarterly series)

6. Graph 1 shows the number of GE registrations with the Australian Taxation Office (ATO) and the number of MU births and deaths from Inteframe, by quarter. Although the ATO tapes are matched monthly, processing the GE births and deaths to Inteframe is only performed on a quarterly basis.



7. The level of birthing activity to Inteframe is generally lower than the number of GE births to the ATO. Some GE registrations are the result of business takeovers and as such may not result in the birth of an MU on Inteframe.

8. In more recent times the level of birthing activity to the Business Register/Inteframe dropped off due to the timing lag between the shut down of the IRIS and the start up of Inteframe. In the June quarter 1999 the backlog that had built up was substantially reduced. Business provisions have been used to insulate survey populations from this effect.
9. A series showing the number of MU deaths from Inteframe has been made possible with the decision to implement GE deathing. GE's are deathed where they have been cancelled by the ATO or have not remitted to the ATO for more than 5 quarters; and are linked to a single LE and single MU; and have less than 50 employees on Inteframe; and are not in the Agriculture ANZSIC Division or have a TOLO between 21 and 42 (ie Government Sector). With the move to a PAYG basis for maintaining Inteframe, the 5 quarter non-remit deathing rule ceased with the processing of the July 2000 ATO tape. When a suitable series of BAS data has been received (by August 2001) the deathing rule will be re-assessed.
10. Resurrections, which are MU's which were deathed because they had cancelled or not remitted for 5 or more quarters but have since commenced remitting are now included as a series in their own right. The vast majority of businesses that fall into this category have less than 10 employees (96.8%) and are spread across all ANZSIC industry divisions. The Property and business services (22.6%), Construction (21.1%), Retail trade (15.3%) and industries contain the highest concentrations of resurrected units. A table showing the breakdown by industry division and employment size is available on request.
11. For the January to March 2000 time period an assumption was made that a backlog of group employer registrations was caused by the ATO not processing GE registrations due to the workload of processing ABN registrations. The ATO was able to provide the ABS with details on a weekly basis of the backlog of GE registration applications received and progress on processing them. As a consequence, it was estimated that an additional 700 businesses should have been birthed in January, 1,900 in February and 4,900 in March, making a total of 7,500 for the quarter. These estimates were used to adjust the business provisions (as shown in graphs 2 and 7) but were not added to the series in graphs 1 and 6.
12. After analysis of data received from the ATO a decision was taken by SSB to gradually reduce the March quarter adjustment over the June quarter business provisions. In April the adjustment was reduced to 5000, in May it was reduced further to 2500, and for June it was reduced to zero. The reason for the reduction of the adjustment was that it was deemed the backlog in processing had caught up. This was supported by the data received from the ATO.
13. The graph supports the theory mentioned in paragraph 11 about the possibility of the backlog as the number of GE births from the ATO is at it's lowest point in four years, whilst the number of GE births reported for June 2000 shows the largest rise from March to June in the four years.

## Summary of Recommendations from Kokic/Brewer Report

- S.1. In order not to trivialise sample rotation, any stratification cell with greater than say 75 per cent allocation should be completely enumerated, while any with between 50 and 75 per cent allocation should have their sampling fraction set at 50 per cent.
- S.2. It is recommended that ONS move to a method of synchronised sampling in order to minimise the degree of overlap between different surveys that are run from the IDBR.
- S.3. To ensure that a synchronised sampling system is as simple as possible to implement the following rules should be adhered to at least initially for all surveys that are to be run on the system
  - S.3.1. The size stratification variable should be employment and common size stratum boundaries should be used.
  - S.3.2. A common definition of industry should be used throughout the business surveys for industry stratification. This stratification can be at different levels for the various surveys, but never too fine as this would lead to problems in controlling rotation.
  - S.3.3. No other stratification should be used for the time being, but it may be worth reconsidering this decision later.
- S.4. The use of a common sampling unit is also fundamental to the successful implementation of such a system. Any survey that samples from a unit higher or lower than an reporting unit will not integrate effectively into the overlap control/sample rotation system. These surveys would need to operate outside the rotation system or be redesigned, using reporting units as sampling units, so they could be integrated into the rotation/overlap control system.
- S.5. A number of ONS and OGDs' surveys wish to use local unit as the sampling unit. It should be possible to select samples for these surveys at reporting unit level. To permit consistent comparison of industry estimates with other business surveys, information would then need to be obtained for each local unit comprising the reporting unit. This approach may prove too expensive for reporting units with a large number of local units, in which case subsampling will have to be performed. In certain circumstances it may be possible to impute the local unit information with a reasonable degree of accuracy from historical information. This approach, however, will not be possible for any survey which requires the sampling unit to be at a level higher than reporting unit (for example, enterprise group).
- S.6. On some occasions enterprises consisting of several reporting units. These are currently being standardised through the Complex Business Unit (CBU). This makes sampling of enterprise units difficult, and the subsequent estimation issues complicated beyond practicalities.
- S.7. Dead units may be discovered from sample sources. It is important that these units are not removed from the register immediately otherwise bias may be introduced into the estimation procedure. In this case the following procedure should be followed.
  - S.7.1. If the sample is a complete enumeration the unit can be removed immediately from the sample and the benchmark total for the stratum adjusted.
  - S.7.2. If the unit was selected in a sample that was not a complete enumeration of the stratum to which it belongs no adjustment should be made to the

corresponding stratum benchmark total until its death can be confirmed from VAT or PAYE sources. However, for the purposes of estimation its response should be set to zero. In other words the unit is treated as if it were live, but with zero response.

- S.7.3. While the dead unit remains within the sample interval an additional unit may be selected to compensate for the loss by extending the sample interval forward by one unit.
- S.8. In nearly all cases a change in stratum cell population size due to births or deaths should not imply a change in the allocation to that cell. The second exception to this rule (after the one noted above) will be in completely enumerated strata, where it will be preferable to maintain the cells in this state even when their population sizes increase.
- S.9. To generate a new PRN, an effective pseudo-random number generator which uses the previously generated PRN as a seed should be used. There are random number generators available that pass through sequences of billions of unique “pseudo-random numbers” before repeating the sequence, and which are fairly straightforward to program. Of course, there will be a need to determine the costs and computing resource requirements before adopting this new procedure.
- S.10. Stability of sample membership is an important issue for effective control of sample rotation and response burden. Sample instability can effectively be reduced in a number of ways, several which have already been mentioned in this summary, including the adoption of a synchronised sampling system, common stratification boundaries, the use of employment rather than turnover for size stratification and the avoidance of fine industry stratification.
- S.11. In order to reduce the chance of units coming into and leaving the sample for short periods, ideally the rotated survey interval should be calculated both forward from its beginning and backward from its end, and the interval implying the larger rotation chosen. As the benefits resulting from this change are expected to be relatively minor, it may be preferable to postpone this alteration until a convenient time in the future
- S.12. For the purposes of sample selection and rotation, ideally all units should be held in the same stratification cell until the next annual update, even though the most recent information (whether from sample or non-sample sources) may be indicating that a change of stratum is called for. Current ONS practice follows this guideline fairly closely.
- S.13. If more recent information becomes available from non-sample sources, auxiliary (employment) values for sample units *and* stratum benchmark totals should be updated immediately.
- S.14. Carry-out and carry-in should be used where there is a change in industry, but not if there is only a size change, because aggregation over size band will be performed in any case. The terms to be carried out and in are the products of the observations and their old stratum weights. This procedure will also be unnecessary in situations where industry level estimates are not required. However, it will probably be simpler to adopt a consistent procedure, that is to always perform carry-out and carry-in between industries regardless of the situation since in the later case, when aggregation is performed over industry, it makes no difference to the values of the final estimates.

- S.15. Reclassified units should retain their original PRNs. However, once a unit is reclassified in an annual update, sample membership should be determined afresh, based on the location of the survey intervals in the new stratification cell.
- S.16. Information obtained from sample (that is non completely enumerated) sources should not generally be used to update the IDBR if it will affect fields that are used for sample design or estimation.
- S.17. Information obtained from sample sources may nevertheless be used for updating fields used only for purposes that have no direct bearing on estimation.
- S.18. Information obtained from units covered by completely enumerated strata, either in the AES or in the business inquiries (or, theoretically, in the two jointly) need not be considered as being from sample sources. (This affects industry updating for Categories 1–7 and for the 50+ size band in Categories 10–11 of the AES.)
- S.19. Where complete information is obtained for a stratification cell over a four year period, there is no ideal feedback strategy, but one that works on an annual updating procedure appears less objectionable than any other. (This affects industry updating for Categories 8–9 of the AES and conditionally the 20–49 size band of Categories 10–11.)
- S.20. The general principle stated in S.16 above is relevant to industry updating for the remaining size bands in Categories 10–11, and in the short term to Categories 13–15 in their entirety.
- S.21. It would be desirable to change the sample allocations for Categories 13–15 so that larger enterprises can be completely enumerated (and perhaps also 25 per cent in the 10–19 size band) in which case they would be affected by rules S.18 and S.19 above, respectively.
- S.22. Notwithstanding any of the recommendations numbered S.16-S.21 above, the general principle in recommendation S.16 can be overridden if changes in VAT industry in particular (but also PAYE where relevant) are in fact treated as signals from non–sample sources that the industry classification requires attention and hence industry is effectively updated continuously from non–sample sources.
- S.23. Sample information may also be used for updating if the sample from which the information came and the sample for which it is to be used are selected in a statistically independent manner.
- S.24. The annual business survey samples are rotated only 50 per cent between surveys, and therefore do not constitute statistically independent samples.
- S.25. It would be quite easy to select business survey samples that were statistically independent of each other on an annual basis, but this would involve a departure from PRNs and consequently this change is not recommended.
- S.26. If, however, the sample information were to consist of data that would in any case shortly be provided (exactly or approximately) by non–sample sources, it would be legitimate to update the IDBR from business survey samples.
- S.27. It may be possible to exploit the potential coverage properties of the union of the AES sample and the business survey samples. Further investigation of this approach would be necessary. With this approach it will be necessary to consider the cost of collecting additional variables (or redefining existing ones) so uniformly defined employment and industry variables can be derived for all the major business surveys.

- S.28. An analysis should be made of the 20,000 sample units in category 15 of the 1995 AES with a view to assessing the impact of duplications in the IDBR on employment estimates, and if necessary steps should be taken to reduce their incidence.
- S.29. Some improvements to the current employment imputation procedures used in the production sector may be possible. In particular, cases where the procedure currently produces implausible results should be examined to see whether particular small industries have substantially different turnover to employment ratios from the others in the same high level industry. Ways to prevent enterprises repeatedly missing out on proving also need to be investigated.
- S.30. The following rules should be used regarding the treatment of unproved VAT-only records consequential on imputation.
- S.30.1 If imputed employment is in the range 0–9 assume that the enterprise is real, and add its imputed employment to the relevant stratum benchmark total, but do not proceed to prove.
- S.30.2 If imputed employment is in the range 10–19, provisionally assume that the enterprise is real, add its imputed employment to the relevant stratum benchmark total, and proceed to prove as soon as possible. Following proving, restratify if necessary, correct the addition to the benchmark total and assign a PRN.
- S.30.3 If imputed employment is 20+, provisionally assume that the enterprise is a duplication or otherwise unreal, do not add its imputed employment to any stratum benchmark total, but proceed to prove as soon as possible. If it is found to be real following proving, stratify, add the employment provided to the stratum benchmark total, and assign a PRN.
- S.31. For the proving of PAYE-only units the rules at S.30 can also be followed, except that in this case turnover, rather than employment, is imputed. The proving of turnover for PAYE-only units will not be as important as the proving of employment of VAT-only units if employment is to be used for size stratification. However, what is of greater importance in this case is industry proving, otherwise, depending on the degree of misclassification, there is the potential for significant undercoverage in certain surveys. It is therefore recommended that industry proving of PAYE-only units with imputed employment of 10 or more continue.
- S.32. Given the potential for inaccurate or duplicated information on the IDBR, and of its central importance to a large number of ONS business surveys, consideration should be given to the investment of further resources into its improvement and maintenance. Of course, there will be a need to weigh up the trade-off between this resource investment compared to others in terms of their effect on the accuracy of survey estimates.
- S.33. Although it would be desirable to have as many business surveys integrated into a sample rotation/overlap control system as possible, there still may be good reasons to maintain some surveys outside the system. The AES should probably be considered in this class, although partial integration could be considered as an option. Any survey with very fine stratification should also be included in this class.

**Recommendations from the Births and Deaths Study**

- We recommend the creation of a new variable on the IDBR to contain the historical information relating to the status changes of units. The relevant status changes are: birth; death; merger; takeover; break-up; split-off; joint venture; restructuring; and, change of group.
- We recommend that births of businesses and deaths of businesses be treated as separate issues in matters of the IDBR and in ONS business surveys.
- We recommend that dead units identified in surveys should be treated no differently from other zero responses, subject to making a detailed investigation of the effects on the results using actual survey data.
- We recommend that further consideration be given to the issue of feeding back to the IDBR information about deregistered businesses identified by surveys, based upon a detailed assessment of the bias caused in survey results and weighing this against the benefit, in terms of maintaining the IDBR, of this procedure.
- We recommend that further work be done to incorporate the results of our modelling of under-coverage into estimation processes, by making informed assumptions about the characteristics of new businesses and translating the estimate of the number of businesses missing from the IDBR into a factor by which to increase our estimates.
- We recommend that the further work proposed here is considered as part of any wider work in ONS concerned with the reporting of survey quality.