

Statistics Commission



Report No. 37

Tax Records as a Statistical Resource:  
A Review by the Statistics Commission

November 2007

# Statistics Commission

Report No. 37

## Tax Records as a Statistical Resource: A Review by the Statistics Commission

November 2007

Statistics Commission  
Artillery House  
11-19 Artillery Row  
London SW1P 1RT  
020 7273 8008  
[www.statscom.org.uk](http://www.statscom.org.uk)

© Crown Copyright 2007



# Contents

Foreword	1
Part One – The Review	
Summary and recommendations	2
Section 1: Inventory of publicly available tax data	6
Section 2: Uses and potential uses of tax data	13
Section 3: Barriers to the greater use of tax data	20
Section 4: Breaking down the barriers – some possible solutions	25
Section 5: Conclusions and recommendations	33
Part Two – Evidence on the use and potential use of tax data	
Section 1: Use and potential use of tax data by academic and other researchers	42
Appendix to Section 1: List of respondents	62
Section 2: Use and potential use of tax data by government departments: Office for National Statistics	63
Section 3: Use and potential use of tax data by other government departments (excluding ONS)	70
Annexes	
Annex A: Tax data – what is available: an inventory of tax statistics	79
Annex B: Legal restrictions on dissemination of tax records by HMRC	97



# Foreword

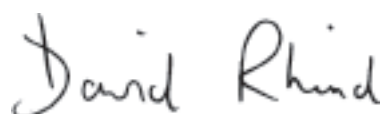
## By the Chairman of the Statistics Commission

Tax records have the potential to be a most important source of statistical information, throwing light on matters from household income and wealth through to population and migration patterns – matters on which official figures are currently relatively weak. Realisation of this potential has been rather less in the UK than in some other countries.

This report was prepared for the Statistics Commission by our head of research, Allen Ritchie, drawing on inputs and comments from many experts. It is a detailed analysis of the case for providing all users of tax-related statistics with more support – and in some cases better statistical data too.

In many European countries, the national statistics office has an absolute right, even a duty, to use the available tax records for statistical purposes. In the UK we do things differently. We have very strong statutory bars on the transfer of confidential information between different parts of government. Whatever the justification for this approach, it has the unfortunate consequence of impeding the statistical service, costing the taxpayer much more for statistics and slowing the delivery and use of such statistics. Unusually, the users whose interests most need to be addressed in the short term seem to be the users within government itself. Readers may be surprised to learn, for instance, that it is more common for academic or private sector researchers to be given access to UK tax records for statistical purposes than for other government departments to be granted this access. The newly enacted Statistics and Registration Service Act contains some provisions specifically designed to help address this problem. We hope that government ministers will be quick to exploit the new freedom the Act allows. The long term goal of course must be to enhance the statistical information available to all users, reduce costs through obviating multiple collection of data and provide a speedier service.

The report addresses its recommendations mainly to Her Majesty's Revenue and Customs, the primary tax collecting department of government. We would like to acknowledge the substantial assistance that officials in that department provided in the preparation of this report. Their openness to our enquiries and understanding of the issues bodes well for resolving the various concerns raised in the report.



Chairman, Statistics Commission

November 2007

# Part One: The Review

## Summary and recommendations

1. This review looks at the potential value for statistical purposes of tax records and data derived from tax records. Most of the research was carried out between June 2006 and March 2007.
2. Tax records represent an important source of data for the compilation of official statistics. In particular, data on tax receipts provide key statistics for analysis of the government's fiscal policy. Information on the tax bases for the different taxes – incomes, wealth, expenditures – is used to compile a range of statistics, including the national accounts and regional economic accounts. Income tax records are the basis for official statistics on the distribution of income; inheritance tax records underlie statistics of personal wealth.
3. The official statistics and analyses derived from tax records are, in turn, an important source for economic and social research. But of equal or greater potential value for research purposes are the microdata underlying the statistics – the tax records themselves. However, much of the information in individual tax records is confidential and the various tax authorities go to great lengths to protect the confidentiality of such records. This is reinforced by tight legal restrictions on the transfer of potentially disclosive data, even between one government department and another undertaking related analytical work.
4. The result is that tax microdata can only be made readily available when the records can be anonymised, and the risk of disclosure effectively eliminated. However anonymisation of data can take time and consume resources, and there is currently only one tax microdata set that is made available for public use (the Survey of Personal Incomes, a 10 per cent sample of income tax returns). Otherwise microdata are usually only provided to non-government researchers working under contract to a government department on a specific approved project, often in a controlled access environment with restrictions on the dissemination of research results as well as a complete prohibition on further dissemination of the records.
5. It is a curiosity of such arrangements that, because government departments cannot enter into legally binding contracts with one another, there can be in practice less access to such records for officials undertaking research on behalf of government than for academic or private sector researchers.

## Key issues

6. The review looks at a number of questions:
  - Who currently uses tax data and for what purposes?
  - What further potential uses might there be?
  - What are the barriers that inhibit these potential uses?
  - What might be done to overcome these barriers, and thus allow more extensive use of tax data for statistical and research purposes than at present?

## The nature of the review

7. Central to the review was the need to gather evidence on present uses of tax data and on the potential for further use. This was approached by circulation of questionnaires to a range of academic and other researchers, and to a number of government departments that were known users of tax data, as well as some interviews with users in the Office for National Statistics (ONS). An inventory of currently published tax statistics and data was also prepared and we have included an annex on legal issues (Annex B), which sets out in more detail the legal position regarding the transfer of data from tax records outside HM Revenue and Customs (HMRC), the main department responsible.
8. The structure of this report is as follows:
  - **Part One** is a report of the review, in five sections. The conclusions and recommendations are summarised below and set out in more detail in **Section 5**.
  - **Part Two** covers the detailed results from the evidence-gathering process of questionnaires and interviews.
  - **Annex A** is an inventory of publicly available tax data.
  - **Annex B** sets out the legal position on the dissemination of data from tax records by HMRC.

## Recommendations

9. The conclusions and recommendations of the review are set out in Section 5 of Part One. The recommendations are summarised below. They are mostly addressed to HMRC who are responsible for collection of the large majority of tax revenues and produce the bulk of the published tax statistics.



10. The first three recommendations relate to the current availability and utility of tax data:

- **Recommendation 1.** HMRC should try to find out more about the users and use of the tax statistics that they publish on their website, and about the needs of those users. To this end, HMRC should consider setting up a user group for tax data.
- **Recommendation 2.** HMRC (and where appropriate other producers of tax statistics) should ensure availability of a comprehensive set of statistics on UK tax revenues and the tax base, by taking action to fill in the gaps in the published statistics identified in this review; and should consider what further statistics could readily be compiled and published from tax records, taking account of the needs of users inside and outside of government.
- **Recommendation 3.** HMRC should seek to improve the accessibility of published tax data, having regard to the principles set out in the Statistics Commission's report *Data on Demand – Access to Official Statistics*. In particular, HMRC should address specific concerns about website navigation and multiple sites identified in this review.

11. The next three recommendations make some suggestions as to what might be done to better support research that uses tax records:

- **Recommendation 4.** HMRC should consider the cases made in the review a) for a larger Survey of Personal Incomes (SPI) in terms of sample size, and b) for developing the SPI into a longitudinal dataset, perhaps through adoption of a rotating sample.
- **Recommendation 5.** HMRC should establish a secure datalab facility for researchers outside government to access HMRC tax microdata for approved research purposes. Approval of individual proposals for access should be on the basis of statistical criteria and on whether or not the research is likely to be of general public interest.
- **Recommendation 6.** HMRC, in conjunction with the departments responsible, should consider the feasibility of enhancing existing official longitudinal microdata sets, as a possible means of meeting the demand for longitudinal data on incomes and tax for research purposes. Specific datasets of interest are the Work and Pensions Longitudinal Study (WPLS) maintained by the Department for Work and Pensions and the ONS business microdata available through Business Data Linking.

12. The final recommendation relates to the provisions within the Statistics and Registration Service Act 2007 which enable data sharing amongst government departments.
  - **Recommendation 7.** HMRC and ONS should make use of the provisions in the Statistics and Registration Act to provide a legal gateway for the release of tax data to ONS for new statistical purposes, where this is not covered by existing legislation.

## Section 1: Inventory of publicly available tax data

13. Annex A to this report presents an inventory of statistics and analyses compiled and published from tax records. This section summarises what is available, and highlights some gaps, and some problems of data accessibility.
14. In the UK context, statistics on taxes, or derived from tax data, are usually compiled and published by the government department that collects the taxes in question. To some extent this is inevitable given the tight restrictions on the transfer of data to other parts of government. In many European countries the national statistical office has a legal right, and duty, to access such records for statistical purposes but there is no such provision in the UK. Most taxes are collected by HM Revenue and Customs (HMRC). The main exceptions are vehicle excise duty, which is collected by the Driver and Vehicle Licensing Agency (DVLA); business rates, which are collected by local authorities (although then passed to central government for redistribution); and council tax, also collected by local authorities and retained by them.
15. Statistics on business rates and council tax are compiled and published by the Department for Communities and Local Government (DCLG) in respect of tax collected in England, by the Scottish Executive for Scotland, and by the Welsh Assembly Government for Wales. The Office for National Statistics (ONS) publishes a number of analyses of aggregate tax receipts, based on data supplied by HMRC and the other tax collecting departments. Apart from this, all publicly available statistics and analyses of taxes for the UK are compiled and published by HMRC.
16. The statistics derived from tax data can be categorised into three types –
  - **statistics on tax receipts**, including details of repayments;
  - **statistics on to the tax base** for the different taxes, including data on taxpayer numbers; and
  - **statistical analyses based on tax data** that go rather further than just extracting details about the tax base.

These are considered in turn below and followed by some observations on the position in relation to national insurance contributions and non-HMRC taxes.

### Statistics on tax receipts

17. The analyses of aggregate tax receipts published by ONS show receipts broken down into five categories based on national accounts – ‘taxes on production’, ‘taxes on income and wealth’, ‘taxes on capital’, ‘other taxes’ and ‘compulsory social contributions’. HMRC publishes statistics for aggregate tax receipts, broken down by individual tax – but only for taxes collected by HMRC. The only place where aggregate receipts data on a tax-by-tax basis for non-HMRC taxes (vehicle excise

duty, council tax, business rates, etc) are published alongside data for HMRC taxes is in the Treasury's Budget and Pre-Budget Reports. These show annual financial year data for tax receipts by tax, but only for the most recent year, together with forecasts for the current year and year ahead.

18. Further details of the composition of tax receipts are published on the HMRC website for the main HMRC taxes (income tax, corporation tax, VAT, fuel duties, stamp duties, tobacco duty) as well as for some of the smaller revenue-raisers (capital gains tax, inheritance tax). The details provided vary from tax to tax, but include separate identification of gross receipts, repayments and net receipts (income tax and VAT); separate figures for accruals/liabilities and receipts (corporation tax, air passenger duty and some other 'environmental' taxes); and a breakdown by industry or trade group (corporation tax, VAT).
19. As well as data on tax receipts, HMRC also publishes estimates of the cost of the main 'tax expenditures' and structural reliefs, covering the most recent year and the current year (projected) only. The estimates cover the cost of reliefs for income tax, corporation tax, VAT, capital gains tax, petroleum revenue tax, inheritance tax and stamp duties, as well as national insurance contributions.

## Statistics on the tax base

20. Statistics relating to the tax base are published for all HMRC taxes, with the notable exception of national insurance contributions. Tax base statistics are of two main kinds – statistics on taxpayer numbers and their distribution with respect to some specific aspect of the tax base (eg income band, industry groups); and statistics derived from data provided on returns claiming particular tax reliefs.
21. Statistics of numbers of taxpayers are published for income tax, corporation tax, capital gains tax and inheritance tax, and of numbers of registered traders for VAT, air passenger duty, landfill tax, climate change levy and aggregates levy. For the excise duties (alcohol, tobacco, fuels) statistics on tax base numbers typically relate to volume indicators for 'clearances' or 'quantities released for consumption'.
22. HMRC publishes a variety of statistics derived from data on tax reliefs for income tax, including statistics on savings and investments in various schemes offering tax relief (eg employee share schemes, tax-free savings accounts) and on tax relief given on 'approved pension schemes'. Statistics relating to corporation tax reliefs include capital allowances due by type of asset and industry group, and research and development (R&D) tax credits.

## Statistical analyses based on tax data

23. The main examples of statistical analyses published by HMRC that go further than summarising tax base information are statistics on the distribution of income, derived from income tax data; estimates of the distribution of wealth (from inheritance tax data); and analyses of property transactions, derived from returns for the purpose of assessing liability for stamp duties.
24. Most of the statistics on income distribution are derived from the Survey of Personal Incomes (SPI), which is an annual sample survey taken by HMRC statisticians from the individual income tax records held by the department. The SPI is the data source for a wide variety of statistics on personal incomes and income distribution, and analyses of the distribution of tax liabilities and of taxpayers.
25. The SPI is the only current example where a tax microdata set is made publicly available. HMRC releases a public use version of each year's SPI on the UK Data Archive at the University of Essex. Data in the public use version are anonymised so that it is not possible to identify any individual.

## National insurance contributions and non-HMRC taxes

26. For non-HMRC taxes, and for national insurance contributions (NICs), the general position is that data are less readily available to the public than for most HMRC taxes. Whilst there are aggregate data on NICs receipts on the HMRC website, there is no further breakdown of these totals. Neither is there any published information on the tax base for NICs. The position is similar for vehicle excise duty (VED). DVLA does not publish any regular statistical series for VED receipts, or for the VED tax base. However, a financial year figure for aggregate receipts can be extracted from DVLA's Annual Report and Accounts.
27. The situation is better for business rates and council tax. Statistics on business rates receipts (tax collected by local authorities, net of reliefs) and on council tax income (net of benefits) for England are published by DCLG. Similar data for Scotland are published by the Scottish Executive and for Wales by the Welsh Assembly Government. All three countries make data for receipts, by individual local authority, available as well. (The system is different in Northern Ireland, where domestic rates have been retained and both domestic and business rates are set partly by the Northern Ireland Assembly and partly by local councils.)
28. There is also a substantial amount of information available on the tax bases for business rates and for council tax – though data in many cases only relate to the most recent year. DCLG for England and the respective devolved administrations for Scotland and Wales each publish analyses of numbers of chargeable dwellings by council tax rate and individual local authority; information on council tax rates (in the

form of the Band D rate for each local authority, together with national and regional averages); and annual statistical releases on collection rates.

29. In England and Wales, the property valuations for business rates (the rateable values) are set by the Valuation Office Agency, an executive agency of HMRC. HMRC has published various analyses of the most recent rating lists (2005), showing property numbers and total rateable values analysed by property type and by region, and the distribution of properties over rateable value bands. For Scotland, the Scottish Executive has published an analysis of numbers of properties and rateable values by property type, based on the latest 'valuation roll'.

## Availability and accessibility of tax data

30. Compilation of this inventory highlighted a number of problems relating to the availability of data – particularly gaps in series – and the accessibility of data on websites. The same issues also emerged from the responses to the questionnaire issued to academic researchers – discussed in Section 2 of this review and reported at greater length in Part Two.
31. On the *availability* of tax data, the main gaps identified were in respect of **vehicle excise duty and national insurance contributions**. In the case of vehicle excise duty, the tax collecting department (DVLA/DfT) did not publish any regular statistics for tax receipts or the tax base. The Treasury's Budget and Pre-Budget Reports include estimates of receipts for both taxes, but only for the most recent year. There does not seem to be any obvious reason why these statistics should not be published. For national insurance contributions, HMRC publish aggregate receipts data, but no further details, or any data on the tax base. Again there does not seem to be any apparent reason why more information should not be published.
32. A rather different 'gap' concerns the absence of easily accessible data for business rates and council tax at a GB level. Separate data for England, Scotland and Wales are readily available but located on different websites, and not always on a fully comparable basis. Again the Treasury's Budget and Pre-Budget Reports include estimates of total (GB) receipts for both taxes, but only covering the most recent year.
33. There are also some problems with the *accessibility* of tax data. Two of these concern navigation through the HMRC website – or more accurately websites. HMRC was formed in 2005 from the merger of the Inland Revenue and HM Customs & Excise, but data for ex-IR and ex-C&E taxes continue to be located on two separate websites. Data for ex-IR taxes can be found on the statistics pages of the main HMRC website, but data for ex-C&E taxes are located on a completely separate website, along with the UK trade statistics also produced by the ex-C&E side of HMRC. There appear to be no links between the two websites. A separate issue regarding the (main) HMRC website is the absence of a link from the HMRC homepage to the statistics pages – though the statistics homepage can at least be

'discovered' through the website search facility (which is not the case for the ex-C&E statistics site).

34. These specific problems of navigation need to be addressed. More generally, however, there is scope for some improvement to the accessibility of HMRC's publicly available tax data, and indeed of UK tax data more generally. The Statistics Commission's June 2007 report *Data on Demand – Access to Official Statistics*<sup>1</sup>, proposed eight principles of statistical dissemination for producers of official statistics to follow when publishing statistics on the web. These principles stress the importance of presenting data on the web in a format designed for the web. Whilst presentation of HMRC tax data is in line with some of the principles – for example, presentation of statistics on the main HMRC website statistics pages is broadly in line with principle 7: "Data should be presented in a layered or hierarchical way to allow users to drill down to the level of detail they desire" – it falls short on some of the others (eg principle 5, "User needs, interests and capabilities should determine the design and operation of statistical dissemination over the Internet").
35. Principle 8 is that there should be a single point of entry – a portal – giving access to official statistics. Whilst this principle is framed in the report in terms of all UK official statistics, it can equally be applied to a specific area of statistics. Most of the problems with availability and accessibility could be addressed at least in part through having a single portal giving access to tax data. The links that would be required from the portal to the data should deal with the navigation problems with the current HMRC website(s). And a portal should make it easier to plug the gaps for GB level business rates and council tax statistics by combining data from individual country websites. Even where having a single point of access to tax data does not directly address the issue, as with the gaps in data for national insurance contributions and vehicle excise duty, it might indirectly help through highlighting the problem.

## International comparisons – the US experience

36. Some respondents to the review questionnaire compared the availability of tax data in the UK unfavourably with that in other countries (see Part Two, paragraph 42). Two respondents specifically mentioned the United States Internal Revenue Service (IRS) as providing a much wider range of information on tax and taxpayers than is available in the UK.
37. In broad qualitative terms, what the IRS makes publicly available through their website is not dissimilar to what HMRC makes available – a range of statistics, in tabular form, on tax receipts and the tax base for the various taxes that they collect, together with some further analyses derived from tax data, including income and wealth

<sup>1</sup> *Data on Demand – Access to Official Statistics*, Report No. 34, Statistics Commission (2007)  
<http://www.statscom.org.uk/uploads/files/reports/Report%2034%20Data%20on%20Demand.pdf>

distribution. The IRS also makes publicly available, in anonymised form, microdata from individual income tax returns (though microdata from corporate tax returns are not publicly available).

38. However, in quantitative terms, much more is available from the IRS than from HMRC, particularly in respect of corporate incomes and tax. For example, HMRC publishes aggregated statistics for the main elements of the corporation tax liability computation broken down into 14 industry groups, three of which relate to the manufacturing industry; the IRS provides a similar analysis broken down into 65 industry groups for manufacturing alone. The range of analyses is also wider. IRS provides much more in the way of geographic analyses based on individual income tax data, including income distribution data by ZIP code. They also use income tax address data to compile data on internal migration flows. And they provide substantially more in the way of information about tax compliance. (Though it should be noted that not all this information is freely available from the website – IRS charges for all microdata and for some analyses, including income distribution by ZIP code and migration flows.)
39. The IRS website is also arguably more user-friendly than that of HMRC. The navigation problems with the HMRC website noted in the previous section do not arise – there is a direct link from the IRS home page to the TaxStats pages, from where all the free-of-charge data and details of chargeable data and microdata can be accessed. (By contrast, whilst the HMRC website provides a full description of the SPI dataset as metadata for the income distribution analyses, there is no mention of the public use version that is available through the UK Data Archive.) However the IRS website is not a portal to all US tax data – only to data on federal taxes collected by the IRS and other federal agencies (the Alcohol and Tobacco Tax and Trade Bureau, the Customs Service). There are no data or links to data on the sales taxes and local income taxes levied at state and local level.
40. In conclusion, the IRS appears to publish substantially more data on taxes, and a wider range of analyses from those data, than does HMRC, on a website without the navigation problems of the HMRC site. The IRS site also provides full details of what is available in the way of tax microdata, and how to get hold of it. There are some important differences between the UK and US tax and statistics systems that need to be recognised. One is that the IRS charges for microdata and some analyses, whereas all data that HMRC publishes are available free of charge. Another is that universal self-assessment for income tax in the US (ie all taxpayers complete a return) means that IRS data on taxpayer addresses – which are a key input to most of the geographical analyses – are likely to be of better quality than HMRC address data from PAYE returns. Nevertheless there may well be something for HMRC to learn from how the IRS presents and publishes tax data.



## **Statistical uses of tax data – the US experience**

As in the UK, tax data from the US Internal Revenue Service are used by other Federal Government agencies for a variety of purposes. The US Census Bureau has told us of ten areas of its work (listed below) that draw directly on federal tax data. The extent of this use, alongside data from other administrative sources, is a good illustration of the potential value of tax records in a range of statistical applications.

Census Bureau applications:

- Survey applications branch
- Census applications branch
- Intercensal population estimates programme
- Longitudinal Employer-Household Dynamics
- Periodic demographic surveys
- Small area estimates
- Survey of business owners
- Census/Social Security Administration joint research
- External research
- Economic Directorate

## Section 2: Uses and potential uses of tax data

41. This section pulls together evidence relating to the uses and potential uses of tax data. The main sources were a questionnaire to academic researchers and a similar one to, and interviews with, government departments. The evidence from the questionnaires and interviews is reported in more detail in Part Two.
42. This section starts by asking who the users are and then goes on to describe some of the current uses. It highlights some areas where users have indicated a potential new use for tax data, or a potential to make greater use of tax data.
43. We identified four main user groups – government departments (other than HMRC); academic and other researchers; fiscal analysts and forecasters; and the general public.

### The public's use of tax data

44. Not much appears to be known about the use of tax data by individual members of the general public. HMRC seems to have practically no information as to who accesses the relevant data on their two websites, or for what purposes. We believe that HMRC should try to find out much more about who accesses their data and for what purposes, especially as regards users outside government and the academic research community.
45. One way to do this might be to set up a formal user group for tax and tax-derived statistics; at present there is no such user group. Nor do tax statistics appear to be covered by any of the 17 user groups affiliated to the Statistics User Forum (SUF). We think HMRC should consider setting up and supporting such a group.

### Specialist users of tax data

46. HMRC knows rather more about the more specialist users of its tax data – academics and other researchers, other government departments including ONS – not the least because these users have a tendency to make their views known. The only current example of an HMRC-run user group is for the mostly specialist group of users of SPI microdata.
47. Evidence gathering for this review has focused on the two groups perceived as the biggest users of tax data – other government departments and academics. An email questionnaire on use and potential use of tax data was sent to some 80 academics; and around 25 replies were received. Evidence on use and potential use by government departments was collected through a series of meetings with key users in ONS, and through an email questionnaire sent to six other departments that were known to be users of HMRC tax data. An email questionnaire was also sent to

selected fiscal analysts and forecasters; this received only a limited response. Findings from this evidence gathering are reported in the following paragraphs.

## Uses by government departments – ONS

48. ONS uses tax data in two different ways – as a data source for published statistics, and in compilation of the sampling frame for statistical surveys of businesses. Sample data from income tax and corporation tax records – including a 1 per cent sample from the National Insurance Records System (NIRS) – are an important data source for the national accounts and regional economic accounts; receipts data for taxes and national insurance contributions are used in compiling the public sector finances data releases. These data are non-disclosive – the data passed to ONS are statistics derived from individual tax records, rather than any details from the individual records themselves.
49. In contrast, construction of the sampling frames for statistical surveys of businesses involves the use of disclosive microdata, with taxpayer names and addresses. PAYE income tax and VAT trader records are used in compiling and updating the Inter-Departmental Business Register (IDBR), and in selecting a sample for the Annual Survey of Hours and Earnings (ASHE).

## Uses by government departments (other than ONS)

50. Other than ONS, the most extensive use of tax data for statistical purposes by a government department is by the Department of Work and Pensions (DWP). In its Work and Pensions Longitudinal Study (WPLS), DWP matches employment data from PAYE and self assessment tax records with benefits data from the department's own records. This is the only instance where HMRC provides disclosive tax microdata for statistical purposes to a government department other than ONS. DWP also makes use of the SPI and the NIRS sample as data sources in compilation of statistics that they publish.
51. Current use of tax data by other government departments is more limited. A number of departments make use of SPI data. There is also some indirect use of tax data through the use by departments of the IDBR. More specialist uses of tax data include that by the Scottish Executive in compiling estimates of the Scottish share of UK tax revenues for the annual publication, Government Expenditure and Revenues in Scotland (GERS).

## Uses of tax data by academics

52. Replies to the questionnaire to academics indicated a range of current uses for the tax data published by HMRC and others. The most frequently cited use was for

research and modelling – a number of specific examples are given in Part Two, paragraph 11. Other uses of tax data mentioned by academics were for teaching, or simply to understand the tax system and explain it to others, and for ‘reference purposes’, in commentary on economic developments and in analysis of policy.

53. One respondent to the questionnaire expressed the view that the tax data published by HMRC “are some of the most interesting, and least known, of all that the government publishes”. However there were also some negative comments, in particular about the published data for tax credits, which were seen as not very helpful, and about a perceived lack of readily available long time series from HMRC.

## Uses of tax microdata

54. Responses to the questionnaires indicated that the SPI – the only publicly available tax microdata set – was used by both government departments and academic researchers. Government departments using the SPI included the Department for the Environment, Food and Rural Affairs (DEFRA), the Department for Business, Enterprise and Regulatory Reform’s Enterprise Directorate (formerly known as the Small Business Service, and referred to in this report as the SBS) and DWP. A number of academic respondents had used the SPI for research, or were considering using it – but others were unaware of its existence. (A couple of the latter group added that, now they knew about the SPI, they might make use of it.)
55. The main issue raised by academic researchers in respect of the SPI was the inability for a researcher to link data from one year’s SPI with data from previous years. This was seen as a major restriction on the potential usefulness of SPI data.

## Uses by fiscal analysts and forecasters

56. Our attempt to gather evidence, through a questionnaire, on the use and potential use of tax data by fiscal analysts and forecasters met with only a very limited response. The replies we did receive indicated substantial use by this group of a wide range of the publicly available tax data from HMRC, including detailed tax base data and statistics on income distribution, as well as the time series for tax receipts. They also indicated some use of tax data from DCLG and the devolved administrations, and of the aggregate tax receipts data published by ONS, in particular the monthly *Public Sector Finances* First Release published jointly by ONS and HM Treasury.
57. We believe that overall this group of users is quite well catered for by the currently available data. In contrast to the responses from academics and government departments discussed in the following section, the questionnaire replies from fiscal analysts and forecasters did not include long wish lists of data and microdata that they would like access to (although one respondent would like to have more detailed figures regarding corporation tax reliefs and allowances).

## Potential new uses of tax data – government departments

58. ONS would like to make more use of administrative data for statistical purposes. A current project is exploring possible ways of using administrative data in order to cut back on business surveys. The motive is to reduce compliance costs for business, particularly smaller businesses. In this context, administrative data mainly means tax data.
59. A feasibility study, by an ONS secondee to HMRC, has confirmed that information from corporation tax (CT) returns, supplemented by similar data for the self-employed and partnerships from self-assessment income tax returns, could be a viable alternative data source for some of the financial information currently collected in the second (financial) part of the Annual Business Inquiry (ABI/2). However there is at present no legal gateway that would allow ONS access to these data.
60. Another potential use of tax data to replace business survey activity would involve the use of VAT data as a data source for some of the estimates currently derived from the ONS monthly and quarterly business surveys. A feasibility study is planned to determine whether this might be a viable option, if there were a legal gateway allowing ONS access to the VAT data in question – which at present there is not.
61. The 2004 *Review of Statistics for Economic Policymaking*<sup>2</sup> by Christopher Allsopp saw the greater use of administrative information, including tax records, as a data source as one of the keys to better regional economic data. The ONS Allsopp Implementation Programme includes a programme of work to improve the quality of data in the IDBR on the location of the activity of companies (better data on 'sites'). However this work does not involve, initially at least, a more extensive use of tax data.
62. A new use of tax data by ONS is in selection of the sample for the Wealth and Assets Survey, a new survey of households which aims to measure households' wealth and assets. The tax records used are income tax self-assessment returns and records of investments in certain tax-free savings vehicles. HMRC is a joint sponsor of the Wealth and Assets Survey, the first of which was launched in July 2006.
63. Population statistics are an area where there is potential for the use of tax information as a data source, but as yet little or no actual use. ONS set out in 2003 proposals for an Integrated Population Statistics System, which would "combine census, survey and administrative data to create a single comprehensive population statistics database", but the time horizon currently envisaged for development is fairly long (after the 2011 Census). The General Register Office for Scotland (GROS), who administer the Census in Scotland, has also noted the potential use of tax data, alongside other administrative and survey sources, in compiling population statistics.

<sup>2</sup> Christopher Allsopp, 'Review of Statistics for Economic Policymaking. Final Report to the Chancellor of the Exchequer, the Governor of the Bank of England and the National Statistician', HM Treasury (2004).  
[http://www.hm-treasury.gov.uk/consultations\\_and\\_legislation/allsopp\\_review/consult\\_allsopp\\_index.cfm](http://www.hm-treasury.gov.uk/consultations_and_legislation/allsopp_review/consult_allsopp_index.cfm)

64. These proposals for population statistics revolve around a substantial degree of data linking between tax records and administrative and survey data, which in turn requires data sharing between government departments, including HMRC. The success of the WPLS (see paragraph 50) illustrates that inter-departmental data sharing is feasible. But here a legal gateway exists that permits HMRC to pass on specified tax records to DWP for linking with benefits information. There is no legal gateway at present that would allow HMRC to pass on detailed tax data to ONS for the purpose of improving population statistics.
65. DWP plans a range of further developments of the WPLS, including integration of Child Tax Credit and Working Tax Credit data. More generally the aim is to expand the use by the department of the WPLS and the tax data in it – operationally, in predictive analysis and in research. Outside of the WPLS, DWP is looking to develop a 100 per cent feed of NIRS data in place of the current 1 per cent sample.
66. DEFRA would like to have more information at a ‘suitable geographical level’ from the SPI for their work on rural policy. The department would also like to see the SPI “combining individual tax records at the household level”, as an input to analysis of farm incomes. DEFRA recognises that meeting these demands would require that the SPI be based on a larger sample. The Welsh Assembly Government has also indicated that a larger sample for the SPI would be helpful. The Scottish Executive would like to see a Scottish sample from the SPI – which might also first require a larger sample overall.
67. Collectively the departments surveyed put together quite a long ‘wish list’ for potential new uses of tax data (for more details, see Part Two, paragraphs 91-97). Many of the items on this wish list would require the release to departments of disclosive data from individual tax records. However in a number of instances departments are just looking for statistical information at a greater level of detail than they have access to at present. For example, the Department for Business, Enterprise and Regulatory Reform (DBERR) would like to have tax data “showing breakdowns by industry or size of business for the take up of tax reliefs and incentives”, for use in policy analysis. Another example is what the SBS has termed its ‘ideal scenario’, whereby HMRC would provide “a representative sample of each subset of the SME [small and medium enterprises] sector (industry sector by employment size band) with information that would allow us to mimic the tax computation of the company”.
68. The Scottish Executive would also like more detailed information for their GERS publication (see paragraph 51). The Executive would like specific information, tax-by-tax, on the amounts of tax collected in Scotland or paid by Scottish residents. However it seems unlikely that this demand could be met at present, as certain tax receipts data (eg VAT receipts) are not available broken down on the geographical basis required for the GERS calculations.

69. Departments' wish lists for disclosive tax microdata include tax records relating to farmers' incomes (DEFRA), company level clearances for dutiable oils (DBERR), and VAT records (for the purpose of constructing microdata matrices of user-supplier relationships between firms) (DBERR). In all these cases, a new legal gateway would need to be established through legislation before HMRC could pass on any data to the other department. In similar vein the Scottish Executive has had requests for access to anonymised individual WPLS data for research and analysis purposes turned down on the grounds of absence of a legal gateway.

## Potential new uses by academic researchers

70. Academic researchers' replies to the questionnaire produced a lengthy wish list of data that academics would like to access for research purposes. (For details, see Part Two, paragraphs 27-28 and 32-34.) This wish list predominantly focused on greater access to microdata (tax records) rather than on the provision of further statistical series and analyses derived from those records, although there were some requests for more statistics to be made available. Some respondents were keen to emphasise the potential for important research that might be unlocked if academic researchers were to be given wider access to tax microdata.
71. Among the demands for more statistics, many were for further detail than is currently made available for series already released. Examples include more data at industry level for VAT, for national insurance contributions and for CT receipts and allowances; more data on recipients of tax credits; and "a detailed breakdown of CT payments". A number of respondents expressed an interest in tax compliance and wanted to see the release of more data relevant to compliance issues. One respondent commented that it was often necessary to "rely on NAO reports to pick up useful information about administrative and compliance issues".
72. Demands for access to (disclosive) microdata covered a wide range, but two main themes stood out. Academic researchers wanted access to tax microdata for companies; and there was a strong demand for longitudinal data on tax and incomes relating to individuals.
73. No microdata for company tax payments are currently made publicly available – there is no equivalent for company incomes of the SPI. Releasing a public use version of the SPI data is possible because the tax records are anonymised prior to release; it would be very difficult to anonymise the data effectively in respect of large companies. (This is discussed further in Section 4, paragraph 104.) Anonymised microdata would anyway not meet the purposes of those respondents who indicated that they would like to have tax microdata in order to match with other microdata on companies, eg the data available through the ONS Business Data Linking arrangements (see paragraph 109). Data matching would require names and addresses.

74. As this report has already noted (paragraph 55), academic researchers' replies to the questionnaire highlighted an inability to link data records through time as the main perceived weakness of the SPI; there appeared to be a strong demand for a longitudinal equivalent. However a longitudinal anonymised SPI dataset could not by itself meet all of the demands from researchers in this area. As with company tax data, a number of respondents are looking to match tax microdata on incomes and tax with microdata from other datasets on individuals. Such data matching would require disclosive data, with names, rather than anonymised longitudinal data. An alternative, where the datasets to be matched are held by government departments, is for government to do the matching, and make matched data available in anonymised format. This is what DWP do with matched tax and benefits data in the WPLS.
  
75. A couple of respondents, who were familiar with the WPLS from previous research, saw an extension of the WPLS dataset, both in terms of numbers of individuals covered and the range of data from HMRC tax records included, as a possible way of providing a longitudinal microdata set for personal incomes. They felt that the existing WPLS did not go far enough – only people with a recent benefit history are included, and the information taken from HMRC tax records only relates to employment. (We return to this later in this review report – see paragraph 123.)



## Section 3: Barriers to the greater use of tax data

76. The evidence on use and potential use of tax data discussed in the previous section suggests that there exists a substantial demand from academics and from other government departments for tax data for statistical and research purposes that is currently not being met. This section considers the reasons for this – what are the barriers that prevent the greater use of tax data for these purposes?
77. This section starts with a discussion of users' perceptions of the barriers to greater use of tax data, drawing on the material from the questionnaires and interviews. We then discuss the nature of legal restrictions on the dissemination of tax records data by HMRC. These restrictions, which are probably the main barrier to a wider use of tax data, prevent HMRC from passing disclosive tax records data to anyone else, even for purely statistical purposes within government, except where this is specifically permitted in legislation. The discussion here draws on the note on the current legal position on dissemination of tax data records, attached at Annex B.
78. The section concludes by discussing other kinds of barriers to the greater use of tax data, including resources, HMRC governance arrangements (in particular its status as a statutory body with defined functions) and departmental culture.

### User perceptions of the barriers

79. It is instructive to begin a discussion of the barriers by considering the perspective of the users of tax data. As part of the evidence gathering process, we sought views from academic researchers and from other government departments.
80. The academic researchers who responded to the questionnaire recognised the existence of legal restrictions on the dissemination of data by HMRC, and their importance as a barrier to greater use of tax data. Nevertheless the general view of the researchers appeared to be that there were probably ways round these restrictions, if HMRC wanted to find them. Some academics thought that resource constraints might be important when it came to providing users with new data or datasets – providing researchers with what they want might well prove to be resource intensive for HMRC, especially as the alternative of simply handing over the raw data and letting the user carry out the necessary calculations might be ruled out by legal restrictions. Many thought that HMRC culture might be a problem, at a senior level if not at working level.

81. The other government departments consulted all regarded legal restrictions as a barrier to the greater use of tax data. Although DWP has a legal gateway that enables them to receive tax microdata from HMRC that are used for the WPLS, the department observed that the statutory gateway places restrictions on the use of the data and on the acquisition and use in the WPLS of data from other departments. All the other departments reported examples where legal restrictions had prevented their access to tax data. (See Part Two, paragraph 102, for examples.)

## Legal restrictions on the dissemination of tax records by HMRC

82. The legal position regarding the dissemination to third parties, including other government departments, of tax data information and records held by HMRC is set out in Annex B, which takes the form of replies by HMRC to a set of questions from the Statistics Commission.
83. The Commissioners for Revenue and Customs Act<sup>3</sup>, which established HMRC as a statutory body from the merger of Inland Revenue and Customs & Excise, also set out the legal position on further dissemination of tax data. Section 18 of that Act first establishes HMRC's obligation not to disclose any of its information to anyone outside the department, and then goes on to prescribe certain exemptions to that overriding principle.
84. What this means in practice is that non-disclosive information, including anonymised microdata, can usually be made available to third parties. The tax data provided to ONS for use in compiling the national and regional accounts, including the NIRS sample, are of this kind, as are the anonymised income tax records made publicly available through the public use version of the SPI. But making disclosive information available, even to another government department, requires an enactment of Parliament. A statutory gateway that allows HMRC to pass on the data has to be established through primary legislation.
85. The recent Statistics and Registration Service Act<sup>4</sup> includes a number of clauses regarding data sharing between government departments, which have implications for the legal position regarding dissemination of tax data by HMRC. An enabling clause<sup>5</sup> allows for the statutory gateway required before disclosive tax data can be passed to another government department to be henceforth provided through secondary legislation under this Act, rather than always requiring new primary legislation. This clause applies to data passed from a government department to the new Statistics Board; a further clause<sup>6</sup> deals with data passed from the Statistics

<sup>3</sup> Commissioners for Revenue and Customs Act 2005 (c.11).  
[http://www.opsi.gov.uk/acts/acts2005/pdf/ukpga\\_20050011\\_en.pdf](http://www.opsi.gov.uk/acts/acts2005/pdf/ukpga_20050011_en.pdf)

<sup>4</sup> Statistics and Registration Service Act 2007 (c.18).  
[http://www.opsi.gov.uk/acts/acts2007/pdf/ukpga\\_20070018\\_en.pdf](http://www.opsi.gov.uk/acts/acts2007/pdf/ukpga_20070018_en.pdf)

<sup>5</sup> op cit, Section 47.

<sup>6</sup> op cit, Section 51.

Board to other government departments. The data sharing clauses of the Act also cover sharing of non-disclosive information, where the Act reinforces HMRC's current practice of being ready to supply non-disclosive aggregate or anonymised information for the purpose of producing National Statistics.

## **Implications of legal restrictions for other government departments**

86. In general the need for new primary legislation for each new instance of data sharing is likely to be quite a strong deterrent to the more extensive use of tax records as a statistical data source outside of HMRC. The discussion in Section 2 of other government departments' current use of tax data included a few examples where HMRC pass on disclosive tax information to another government department for statistical purposes – PAYE and VAT data to ONS for the IDBR, PAYE information to ONS for the ASHE sample, income tax data to DWP for incorporation into the WPLS. In all these cases a statutory gateway enabling HMRC to do this had been previously created through primary legislation. However the political will to put through the necessary legislation may not always be there.
87. The Statistics and Registration Service Act should make it easier than at present to secure a statutory gateway for new instances of sharing of tax records data between HMRC and ONS (which will become the Executive Office of the new Statistics Board). (Although any secondary legislation relating to HMRC data will need the agreement of HMRC Commissioners and the Treasury, as well as the minister for the Cabinet Office, as the minister with residual responsibilities for the Statistics Board.) However it is less clear how or whether the Act will help facilitate more sharing of tax data between HMRC and government departments other than ONS, as the relevant clauses specifically relate to data sharing between the Statistics Board and government departments, and not between two government departments. (This is discussed further in Section 4.)

## **Implications of legal restrictions for academic researchers**

88. The statutory bar that the Commissioners for Revenue and Customs Act places on HMRC disclosing information outside of the department means that access to disclosive tax microdata for researchers working outside of government is tightly restricted. The only way in practice that academic researchers can gain access to disclosive tax data is through carrying out research working under contract for HMRC. A researcher working under contract has to sign HMRC conditions of employment and so effectively becomes an HMRC employee and covered by HMRC disclosure policies for the purpose of the work. The Statistics and Registration Service Act will not change this.

## Other barriers to greater use of tax data – resources

89. On the whole, HMRC publishes quite a lot of different statistics and statistical analyses derived from tax data – although there are a few gaps (see Section 1). So, when a statistical series is not published in the detail that a particular user would like (eg demands for more industry level detail for CT data), it will often be the case that the series is not compiled in a way that generates data at that level of detail – and that to do so would involve a substantial commitment of resources by HMRC. However it may not always be clear to users that the main barrier to getting hold of the data they want is a lack of resources rather than legal restrictions or a lack of interest.
90. The legal restrictions on dissemination of tax records data may also be a factor in certain cases, as they may effectively rule out the simple option of meeting demands for statistics that have not been compiled (but could be) by passing the raw data to whoever wants it, and leaving the user to do the statistical manipulation required. In other cases it may not be a question of a lack of resources to perform the necessary statistical manipulation; the details that users want may not actually be collected at all. HMRC may not need to know that particular information for the purposes of assessing tax liability, and so does not ask taxpayers for it. This will always be an issue with administrative data that are used as a data source for statistics – should forms be designed in order to collect data that do not meet any administrative requirement, but might be of use for statistical purposes?
91. This applies to the data demands of government departments as well as to those of academics. The Scottish Executive's demand for data on taxes paid in Scotland for their GERS publication (see paragraph 68) is a good example. As a tax collector, HMRC is only interested in whether or not taxable income was earned, or taxable transactions took place, in the UK – not in the whereabouts of the income earned or the transactions that took place. The only geographical information that HMRC will normally hold is for taxpayer addresses. Where taxes are paid by individuals (eg income tax, national insurance contributions) address information may be of some use in compiling a geographical analysis of tax receipts. But where companies make tax payments (corporation tax, VAT) it may be much less useful. For example, HMRC may only have a single address – typically the head office – for a company that operates on several sites in the UK, and that sells its products or services throughout the UK.

## HMRC culture as a barrier

92. A number of respondents to our questionnaires suggested that HMRC culture might be a barrier to the more extensive use of tax data for statistical purposes. The ‘culture’ that matters is that of the department’s top management. Some academic respondents observed that requests for access to unreleased microdata were invariably turned down, even though the officials they dealt with were often sympathetic. This pointed, they suggested, to resistance amongst senior management. But it could equally be that the less senior officials were not fully familiar with the legal position.
93. There is some evidence that HMRC takes a more restrictive stance on release of data for statistical purposes than other government departments. One respondent compared HMRC unfavourably with DWP on this: “the DWP has a long established culture of producing statistics ... [and of] commissioning and publishing research. HMRC lacks this track record/culture.” However this may reflect governance factors as much as, or more than, culture. HMRC is a statutory body with its functions defined by statute (for most government departments, their functions are defined by administrative fiat), which may restrict the scope for flexibility in interpreting the scope of its functions.
94. This may matter when it comes to agreeing to the release of microdata for specific academic research projects. The usual way to deal with legal restrictions that prevent the release of disclosive microdata to third parties is to release data to researchers under contract for specific approved projects. ONS do this for business and census microdata through their Virtual Microdata Laboratory (VML). DWP do it for WPLS data (which include some tax information). There needs to be a process for approving research proposals; for legal reasons, this typically includes a departmental benefit test – the research project needs to benefit the department releasing the data.
95. HMRC’s governance status as a statutory body, with functions defined by statute, may make a departmental benefit test more restrictive for HMRC than it is for other departments, for whom functions are less closely and strictly defined. The departmental policy agenda may also restrict the range of research projects HMRC feels able to approve, particularly in comparison with ONS, whose main function is to compile statistics and make these publicly available, and which – unlike other statistics-producing departments – does not have a departmental policy agenda that is separate from its statistical agenda.

## Section 4: Breaking down the barriers – some possible solutions

96. This section discusses some possible solutions to the problems presented by the existing barriers to greater use of tax data that are outlined in the previous section. It starts by considering whether there are any alternatives to tax data for the kind of uses that respondents to the review questionnaires had in mind.
97. The section goes on to discuss a number of options for making tax data more readily available for statistical purposes – more publicly available anonymised microdata; wider provision of microdata for specific research projects under controlled access conditions (ie a datalab); more prior data matching and linking by government to create longitudinal datasets like the WPLS. Finally it looks again at the legal position in the light of the recently enacted Statistics and Registration Service Act, and considers whether the Act makes any difference.

### Are there alternatives to use of tax data?

98. Evidence gathering for the review indicated a strong and widespread demand for greater access to tax data for statistical purposes. However, as the previous section has demonstrated, there are a number of difficulties involved in meeting these demands, particularly where doing so would involve granting access to disclosive microdata. So are there alternative ways in which these demands could be met?
99. Users have indicated that they would like to have access to microdata from company tax returns. Company tax returns are provided in confidence and the usual legal restrictions on disclosure apply. However companies are legally required to compile company accounts, which will include at least part of the information on their tax return. These accounts are published; details from them also need to be filed with Companies House, a government agency. So why can't academic researchers use (publicly available) company accounts data instead of (confidential) company tax microdata?
100. There are two reasons why researchers do not use company accounts data. Firstly company accounts data are not publicly available in a very user-friendly form. Companies House makes information available at the level of the individual company, but does not make accounting data available in a form suitable for use as a statistical data source in research (unless the user is only interested in a specific named company). For microdata and statistical data across a broad range of companies, researchers need to go to a commercial organisation such as DataStream – who will charge for these data.

101. Secondly company accounts do not provide the full range of information that is available from the tax returns. Moreover, for small companies, Companies House requires only summary accounts data to be supplied. ONS has cited this as the main reason why they did not consider Companies House data to be a viable alternative source for the financial data collected in business surveys, but opted for company tax data instead (see paragraph 59).
102. Inheritance tax (IHT) provides another example of an area where much of the information available from individual tax returns is already in the public domain, in the form of probate returns. However, as with company accounts, probate data are not readily accessible in a user-friendly way. The situation is in fact significantly worse than for Companies House data, where information for different individual companies is at least accessible through the same web portal. Individual probate records are only available from the local probate office. Moreover no commercial organisation appears to be in the business of pulling together probate data and supplying those data as a statistical source – in the way that DataStream does for company tax data. It is also the case that some of the information available from the IHT forms is not available from probate returns.

## More anonymised datasets

103. One way in which tax microdata might be made publicly available to a wider audience, staying within the legal constraints on dissemination, is through anonymisation of the data. The SPI illustrates how this can be done. The safeguards that are applied to the SPI dataset before it is passed to the UK Data Archive (anonymisation of all data records, suppression of some outliers through combining individual records) provide a model for the kind of disclosure controls that need to be applied.
104. Could this model be applied in areas other than personal incomes? Would it be feasible, for example, for HMRC to produce and release an equivalent of the SPI from a sample of company tax returns (a 'Survey of Company Incomes' or SCI)? A major obstacle is that effective anonymisation of tax returns is likely to prove much more difficult for company incomes than for personal incomes – it will be very difficult to hide the identities of the larger companies in the sample. Suppression of the details of the larger companies through combining individual records – the method followed for the SPI – is likely to produce a dataset that is in effect a 'survey of company incomes' for smaller companies, together with aggregated data for larger companies. This seems likely to be of limited usefulness, and unlikely to justify the resources that would be needed for development and operation. (Though such a dataset may be of some use in certain situations; an SCI of the SME sector is essentially what the SBS have asked for – see paragraph 67.)

105. The SPI in its current form does not provide the longitudinal microdata that many would like to have – and it will not be able to do so as long as it remains based on a sample that is drawn afresh each year. It would be technically possible to move to a longitudinal SPI, either through fixing the sample over a number of years or moving to a 100 per cent sample – although the latter in particular is likely to be very resource intensive for HMRC. One suggestion that might enable a longitudinal dimension to be added to the SPI would be to use a rotating sample, whereby some of the sample is retained from one year to the next, whilst the remainder is drawn afresh.
106. We were struck by the strength of the demand for longitudinal data on personal incomes and tax indicated by the questionnaire replies, and we think that HMRC should give serious consideration to the possibility of developing the SPI into a longitudinal dataset, perhaps through moving to a rotating sample. However, we do not think that a companies' equivalent of the SPI – an SCI – is likely to be feasible, given the difficulty of effectively anonymising the tax data records of larger companies.

## Working with data at HMRC

107. An alternative to releasing more data (anonymised or not) as a means of providing greater access to tax data for statistical purposes is to provide individual researchers with controlled access to microdata for specific approved projects. HMRC already do this to some extent, as do some other departments that compile and publish statistics (the experiences of two of these departments – ONS and DWP – are discussed further below). Providing controlled access to data can either involve requiring researchers to work on data under controlled access at a 'secure' location (a datalab), or else providing the data for researchers to work on in their own environment, with legal constraints on the wider use and dissemination of these data. A key principle of controlled access is that researchers have access to potentially disclosive data, but the results taken away must be non-disclosive.
108. The legal constraints faced by ONS (which derive from the Statistics of Trade Act<sup>7</sup>) in regard of the dissemination for research purposes of business microdata collected in business surveys are not dissimilar to those faced by HMRC in respect of tax microdata. The ONS solution has been to set up a secure datalab, where researchers can work with business microdata for approved projects under secure conditions – and only take away their results, not the data. The ONS Virtual Microdata Laboratory (VML) is used by ONS to provide access to microdata from business surveys, and from certain other sources through the Business Data Linking project; the VML is also used as means of providing access to census sample microdata, and to other datasets where confidentiality reasons prevent data being released through other channels.

<sup>7</sup> The Statistics of Trade Act 1947, amended 1990  
[http://www.England-legislation.hmso.gov.uk/si/si1990/Uksi\\_19902597\\_en\\_1.htm](http://www.England-legislation.hmso.gov.uk/si/si1990/Uksi_19902597_en_1.htm)



109. The ONS Business Data Linking arrangements have a number of key features. Access to data is only provided for specific research proposals, which need to be authorised in advance by the ONS Microdata Release Panel. The research has to be of benefit to ONS, as the commissioning department. Researchers who are not civil servants or contracted to other government departments carry out their research under contract to ONS, so they become in effect ONS employees for the duration of the project. Their current employing institution is required to have an institutional data access agreement with ONS. For researchers from other government departments, access is only given under a Ministerial Directive that specifies the uses to be made of the data. Access to data is only possible on-site at a specific secure location (the datalab). And all results released are scrutinised by trained analysts against the possibility of accidental disclosure of information. (ONS also trains all datalab users on how to avoid accidental disclosure.) Many of these features – eg the requirement for employment contracts, use of a departmental benefit test in the approval process – are necessary for compliance with the legal restrictions on ONS in the Statistics of Trade Act.
110. HMRC statisticians are looking at the possibility of developing a secure datalab at HMRC to provide access to tax data. If established, an HMRC datalab seems likely to have similar features to, and operate along similar lines as, the ONS Business Data Linking arrangements.
111. The views about datalabs expressed by academic respondents to the review questionnaire were mixed. Some were keen; others were firmly against, on the grounds that it would “drastically reduce the value [to the researcher] of the data”. Even amongst those prepared to accept controlled access through a datalab if that was what was required, a datalab was seen as less than ideal; respondents said that they would prefer to take the data away, and sign a legally binding undertaking not to disclose any further.
112. In contrast to ONS for business microdata, DWP has not gone down the datalab route in respect of WPLS data. Researchers who are granted access to WPLS microdata for approved research by DWP are not usually required to access those data using DWP facilities only. However DWP generally only provide WPLS data to researchers in anonymised form. This is feasible because the data relate to individuals’ incomes and benefits, not companies, and is acceptable to the researcher because data matching (of DWP and HMRC microdata) has already been carried out.
113. Notwithstanding the adverse views on datalabs expressed by some of our academic respondents, we believe that – given the sensitivity of tax data and the disclosure risks inevitably associated with any form of direct release – an HMRC datalab along the lines of the ONS VML may offer the best way forward for allowing non-government researchers greater access to HMRC tax microdata.

114. All applications for access to ONS microdata held in the VML are required to go through a formal approval process. There are similar vetting procedures for access DWP's WPLS microdata. Partly for legal reasons, a key feature of these procedures is a 'departmental benefit' test – to be approved for access, a research proposal needs to be of benefit to ONS (or to DWP for WPLS microdata).
115. As discussed in the previous section (paragraph 95), it is possible that a departmental benefit test applied to the release of confidential data held by HMRC might prove more restrictive than a similar test applied to data held by ONS or indeed to data held by DWP. Unlike most government departments, HMRC is a statutory body, with functions clearly defined by statute. And unlike ONS, which as the Executive Office of the new Statistics Board will soon be in a similar position of having its functions defined by statute, the specified functions of HMRC are primarily concerned with administration and policy (of the tax system) and not the production and dissemination of statistics and the promotion of statistical research.
116. The Statistics Commission accepts that any release for statistical purposes of disclosive microdata originally supplied in confidence has to be properly justified and tightly controlled. We believe that release of data through a secure datalab facility, with all proposals for access subject to a formal approval process, should be able to provide that justification and control. However we are uneasy about a situation in which the strictness of the controls operated partly depends on a departmental benefit test, the effectiveness of which is likely to vary between departments, depending on how tightly that department's functions are defined in law and on the extent to which that department's policy functions are distinct from its statistical functions.
117. The Commission believes that providing access to microdata for statistical research, whether through a datalab facility or not, should be regarded as a statistical function which comes under the auspices of the new Statistics Board. We fully accept the necessity of a formal approval process for research proposals looking for access to disclosive microdata, but believe that the criteria against which research proposals for access are assessed should be statistical – eg do the proposals represent a proper use of the data?, is the research likely to be of general public interest? – and should be similar for all official microdata, regardless of which department holds the data.
118. Other government departments might also benefit from the establishment of an HMRC datalab facility. At present any project of another department requiring access to tax records needs to be undertaken by seconding someone to work as an HMRC employee, so that they can have fairly unrestricted access to tax records. For example, ONS feasibility studies of the possible use of tax data to replace some business surveys were carried out by someone on secondment to HMRC. In future such investigations might be carried out through use of a secure datalab facility. However access through a secure datalab should not be seen as a permanent solution in cases where a department needs access to tax records data on a regular basis.

## Meeting the demand for longitudinal data – more data matching by government

119. Improving access to microdata through a datalab facility may in itself be of only limited use in meeting the strong demand for longitudinal data indicated by the research for this review. Even where sufficient data have been released to make linking of individual records over a run of years technically feasible, putting together a longitudinal dataset for research could be a major task for an individual researcher. And it is the release of data with disclosive identifiers (eg names and addresses), which researchers would need for data matching, that poses the big disclosure risks.
120. A better way of providing more longitudinal data for research purposes might be for government statisticians to do the linking first. This could also help meet government departments' own needs for data for research and analysis. There are already some official longitudinal datasets in existence; DWP's WPLS is an example of an official longitudinal dataset that incorporates some tax data. A number of government departments have recognised the potential for making better use of the longitudinal information available within government in their own analytical work; to this end, six departments, including HMRC, ONS and DWP, have recently set up a project to look at Analytical Data Integration in Government (ADIG). The ADIG project is looking at the feasibility of creating integrated longitudinal cross-government datasets for policy, research, analysis, evaluation and decision making.
121. ADIG is not intended to lead to the creation of a mega database of government information. What is proposed instead is a federated approach, whereby raw data stay secure within the collecting departments, but are linked through a secure hub, which provides a facility for integrating and matching data as required. Datasets for specific purposes can then be pulled together, drawing on the linked data. As these purposes will usually be research or analysis, the presumption is that these datasets will not as a matter of course be updated or retained. It is intended that data extraction and manipulation should take place within a secure datalab environment, and that, once the necessary data matching has been carried out, datasets for research and analysis will be anonymised. All research proposals for linking and using datasets within the hub will be required to be vetted; the approval process should include an assessment of the ethics of linking and using data in such an environment.
122. As well as meeting the needs of government research and analysis, realisation of the vision of the ADIG project could go a long way towards helping meet the demands of non-government researchers for longitudinal data – but only as long as non-government researchers are given access in some form or other to the individual datasets that are created. In cases where anonymisation is fully effective in removing all disclosure risks, this might be done by making the individual datasets, or versions of them, publicly available through an independent third party, eg the UK Data Archive, in the same way that HMRC currently make available a public use version of the SPI. Where anonymisation is not fully effective in removing all disclosure risks,

access to non-government researchers could be given through a secure datalab environment, with research proposals subject to a prior approval process. This is the model proposed earlier (paragraphs 109 to 113) for HMRC tax data.

123. Realisation and implementation of the ADIG project is for the future; in the meantime, progress on meeting the demand from non-government users for longitudinal tax data seems most likely to come through enhancement of an existing dataset such as the WPLS. This is likely to be a more attractive option, from the point of view of resources, than creating a new longitudinal dataset from scratch. The WPLS currently matches DWP's own departmental benefit records with employment data from HMRC tax records, with strict safeguards on access to and use of the resulting matched dataset. Some academics suggested, in their questionnaire responses, that WPLS might be usefully extended towards universal coverage of all PAYE taxpayers rather than just those with a benefit history (see paragraph 75). We think that this idea is well worth considering. An enhanced WPLS could provide a useful source for longitudinal data on personal incomes and tax.
124. What about company incomes and tax? The ONS Annual Respondents Database (ARD) is a form of 'Companies Longitudinal Survey' that links successive years of business survey data (from the ABI and predecessors). The ONS Business Data Linking (BDL) project provides a facility for matching ARD data with data from other business surveys (including survey data from DTI). The ARD and the additional BDL data are accessed through the ONS VML.
125. At present, no tax data are incorporated in the ARD/BDL data. Tax data are used in compiling the IDBR, but the data on ARD/BDL are from business surveys, and not taken directly from IDBR. However the developments discussed in Section 2 (paragraphs 58-60) – in particular the proposed replacement of the ABI/2 survey with data from tax records – would integrate tax data into the ARD/BDL dataset for the first time, and thus represent a move for the ARDBDL data in the direction of becoming an equivalent of the WPLS for companies. We would regard this as a favourable development.
126. More generally we would encourage ONS and HMRC to consider the feasibility of enhancing the ARD/BDL data with some data from tax records. In proposing this, we recognise that longitudinal business data will inevitably be disclosive in some parts, and that access to the data will probably need to be under datalab conditions.

## Overcoming the legal barriers – the impact of statistical legislation

127. The need to have primary legislation in place before tax microdata can be passed to another government department for statistical purposes is not an insuperable obstacle to extending the use of tax data in research. For example the WPLS was set up quite recently following this route. However a degree of political commitment will normally be required to carry through the necessary legislative changes.
128. As noted in the previous section of this review report (paragraph 85), the provisions of the Statistics and Registration Service Act regarding data sharing will make obtaining legal cover for new instances of HMRC sharing data with the Statistics Board easier, by allowing this to be provided through secondary legislation under the Act, rather than requiring new primary legislation to be passed. But the requirement for legal cover is not being removed altogether. Moreover any secondary legislation in respect of specific instances of new data sharing involving HMRC as one of the parties will need the agreement of HMRC Commissioners and the Treasury, as well as the Minister for the Cabinet Office (as minister for the Statistics Board). The provisions of the Act are probably of most practical significance for the ONS proposals to use administrative (tax) data in order to cut back on business surveys (see Section 2). This should now be able to move forward under cover of a statutory instrument, rather than requiring new primary legislation – always assuming that HMRC Commissioners have no objections.
129. It is less clear how the provisions of the new legislation will impact on departments other than ONS. The Statistics and Registration Service Act provisions that allow new instances of data sharing to be covered by secondary legislation apply specifically to data sharing between the Statistics Board and other government departments, including HMRC. The Act makes no mention of data sharing between two government departments. It may be possible, for example, for DBERR to use the secondary legislation route to get legal cover for access to HMRC microdata, by going through the medium of the Statistics Board – ie two statutory instruments are passed, one to allow data to be shared between HMRC and the Board, and then a second to allow the Board to share the same data with DBERR. Even if this could be agreed, it would still need the formal agreement of HMRC Commissioners, the Treasury and two Departmental ministers – Cabinet Office (for the Statistics Board) and DBERR.

## Section 5: Conclusions and recommendations

130. This section brings together the main conclusions from this review, which has looked at the use and potential use of UK tax data, and sets out the Commission's recommendations. These are mostly for HMRC, but in some cases are also directed at the other departments involved in the production and dissemination of tax data.

### Knowing the users

131. The review gathered evidence from users on the use and potential use of tax data and focused on two main groups of users – other government departments, who use tax data both as a data source for other statistics and for policy analysis and research, and academic researchers. This was partly because these two groups were seen as the key users of tax data – but it was also because very little appeared to be known about other users outside these two groups, in particular about the use of published tax data by the general public. It became clear at an early stage of the review that, outside of other government departments and apart from some academics with whom HMRC economists and statisticians were in regular contact, HMRC had little knowledge as to who used the data that they made available on their website.

132. We believe that HMRC need to do more to find out about who uses the statistics that they publish on their website, and for what purpose, especially in respect of users outside of government and the academic research community. In this context we note the absence of any recognised user group for tax statistics; none of the groups currently affiliated to the Statistics User Forum (SUF) appears to cover tax data. HMRC might want to consider setting up such a group.

133. So our first recommendation is:

- **HMRC should try to find out more about the users and uses made of the tax statistics that they publish on their website, and about the needs of those users. To this end, HMRC might want to consider setting up a user group for tax data.**

### Coverage of published statistics

134. Although tax records are administrative data, collected for a purpose other than compiling statistics, they nevertheless provide a rich data source for compilation of statistics about the revenues collected and the bases on which the different taxes are levied (eg personal incomes, company incomes, consumption of specific goods and services), as well as the raw data for statistical analyses of income and wealth distribution. The Commission is keen to promote the use of administrative data such as tax records as a statistical data source, whilst recognising that information will not always be collected in ways that meet the demands of the producers and users of statistics.

135. The inventory of published tax data reported in Annex A illustrates that HMRC and the other tax collecting departments publish a substantial volume of statistics and analyses derived from tax records. Nevertheless there are some clear gaps in respect of the data that are made available for specific taxes – particularly in respect of non-HMRC taxes. The biggest of these gaps is the absence of any detailed statistics on National Insurance Contributions; other gaps relate to Vehicle Excise Duty revenues, and to data relating to council tax and business rates at a GB level. In none of these cases is there any question of the data not being available, and we would invite HMRC, along with the other departments involved, to take action to address these gaps in the published statistics.
136. At a more general level, we would note that, whilst HMRC publish a wide range of statistics and analyses on their website, as well as providing public access to some personal income tax microdata, their counterparts in some other countries manage to publish more. One of the respondents to our survey of academic users remarked on the generally wider range of tax data for the US available on the Internal Revenue service website – including anonymised personal tax microdata similar to the SPI. (Some of the IRS data are, however, only available on payment of a charge.) HMRC may want to consider this point.
137. A number of our academic respondents made specific suggestions regarding additional details of tax revenues and/or the tax base that they would like to see released. (Details are in Part Two, Section 1.6.) We would invite HMRC (or other departments where appropriate) to consider whether any of these suggestions could be readily met from the information currently collected, without compromising the confidentiality of individual records.
138. Our second recommendation is as follows:
- **HMRC (and where appropriate the other producers of tax statistics) should ensure availability of a comprehensive set of statistics on UK tax revenues and the tax base, by taking action to fill in the gaps in the published statistics identified in this review; and should consider what further statistics could readily be compiled and published from tax records, taking account of the needs of users inside and outside of government.**

## Accessibility of tax statistics

139. A recent Statistics Commission report, *Data on Demand – Access to Official Statistics*,<sup>8</sup> looked at the accessibility of official data published on the web, and proposed a set of eight principles of statistical dissemination that departments who

<sup>8</sup> *Data on Demand – Access to Official Statistics*. Report No. 34, Statistics Commission (2007)  
<http://www.statscom.org.uk/uploads/files/reports/Report%2034%20Data%20on%20Demand.pdf>

publish official statistics should aim to follow. The web dissemination of tax statistics by HMRC currently departs from these principles in a number of respects. For example, there is little evidence that statistical outputs have been specifically designed for the web, nor that user needs, interests and capabilities have played much of a role in the design and operation of statistical dissemination. We believe that the presentation of tax data on the HMRC website would benefit from some restructuring, and we would commend the principles in our recent report as the basis for doing this.

140. There are a couple of specific issues regarding the HMRC website that we believe need to be addressed urgently. The first is the continuing existence of entirely separate websites for ex-Inland Revenue and ex-Customs & Excise taxes. The second is site navigation, in particular the absence of clear links (a) from the home page to the (ex-IR) statistics home page and (b) from the HMRC site to the separate ex-C&E site. It would also be helpful if all tax data were accessible at, or from, a single location, consistent with the 'single point of entry' principle (principle of statistical dissemination no 8). This could be realised by merging the two HMRC websites, and making data on non-HMRC taxes available on the (merged) HMRC site.
141. Our third recommendation, concerning the accessibility of tax data, is:
- **HMRC should seek to improve the accessibility of published tax data, having regard to the eight principle of statistical dissemination set out in the Commission's report *Data on Demand – Access to Official Statistics*. In particular, HMRC should address the specific issues concerning website navigation and multiple sites identified in this review.**

## Tax microdata – developing the SPI

142. The evidence gathered for this review in questionnaire replies from academics and from government departments indicated a strong demand for greater access to tax microdata for statistical research. However granting access to any personal data that government collects for administrative purposes inevitably raises questions regarding confidentiality and inadvertent disclosure. In the case of tax records, the usual data protection considerations are reinforced by specific legal restrictions which prevent HMRC passing on any data of a potentially disclosive kind, unless a legal gateway for doing so has been created by separate legislation.
143. Release of an anonymised sample of records offers one way of providing researchers with access to microdata, without contravening the legal restrictions. HMRC have followed this route with the release of annual personal income tax microdata in the form of a public (anonymised) version of their Survey of Personal Incomes (SPI).



144. User views regarding the SPI were generally positive, although some believe that it is under-used as a research resource. Questionnaire replies highlighted two kinds of development for the SPI that respondents would like to see. The first is a larger sample, which would permit robust analysis with the data at a finer level of detail than the present 10 per cent sample allows. The second is a move towards a longitudinal survey that would enable SPI datasets from different years to be linked. We would encourage HMRC to consider these demands, and the ways in which the SPI might be developed to meet them. A move towards a longitudinal survey might be realised through a rotating sample.
145. The questionnaire replies also indicated a substantial demand for tax microdata relating to company incomes. We have considered whether this could be met through development of an anonymised sample of tax records along the lines of the SPI – a ‘Survey of Company Incomes’ – but have concluded that disclosure issues in respect of large companies probably rule this out. Suppression and/or combination of individual records for the highest payers, which is what is done for the SPI, is likely to substantially reduce the usefulness of a company incomes equivalent.
146. Therefore our fourth recommendation is:
- **HMRC should consider the cases made in the review (a) for a larger SPI in terms of sample size, and (b) for developing the SPI into a longitudinal dataset, perhaps through adoption of a rotating sample.**

## Access to microdata for non-government researchers – the datalab approach

147. Where microdata cannot be suitably anonymised – for example, because of the difficulty of concealing the identity of large companies – public release of data along the lines of the SPI model will not normally be a feasible option; the danger of inadvertent disclosure will usually be too high. Yet the questionnaire replies from non-government researchers indicate a clear demand for access to disclosive tax microdata for research purposes.
148. One option for meeting this demand is to release specific datasets or parts of datasets to named individuals, under contract terms that restrict use to a pre-agreed statistical purpose and prohibit further dissemination. Questionnaire responses suggested that a number of academics would be in favour of such a model. But any direct release of confidential and/or disclosive microdata carries a risk of inadvertent further disclosure.
149. An alternative to direct release of tax microdata to individual researchers is to provide individuals with controlled access to specific data sets for agreed research purposes at a secure location. This ‘datalab’ approach has been developed by ONS for their Business Data Linking arrangements, which provide approved researchers with

access to disclosive information on businesses from business surveys through the ONS Virtual Microdata Laboratory (VML). ONS also use the VML facility to provide approved non-government researchers with access to a sample of microdata from the Census, and to other datasets where confidentiality prevents release through other channels.

150. The evidence from our questionnaire indicates a substantial demand for disclosive tax microdata for research purposes that is currently not being met. Notwithstanding the adverse views on datalabs expressed by a number of our academic respondents, we feel that – given the sensitivity of tax microdata and the risks associated with direct release – the datalab approach probably offers the best way forward for meeting some at least of this demand. We understand that HMRC are looking at the possibility of setting up a datalab for tax microdata, and would strongly support such a development.
151. The key features of the ONS Business Data Linking arrangements are described in paragraph 109. They include the following:
- access to data is provided for specific research proposals, which are authorised by the ONS Microdata Release Panel;
  - research must benefit ONS;
  - non-government researchers carry out their research under contract to ONS, with whom their employing institution must have an institutional data access agreement;
  - access for other government departments is only given under a Ministerial Directive, which specifies the uses to be made of the data;
  - access to data is on-site at a specific secure location (the datalab);
  - results are scrutinised against the possibility of accidental disclosure of information.
152. Details of how a tax microdata laboratory might operate need to be worked out, and will at least partly depend on what is needed for compliance with legal restrictions. But it seems reasonable to expect that a datalab for tax microdata might operate along similar lines to that for ONS business data. The requirements for contracts and for research to be of benefit to ONS are necessary for compliance with legal restrictions (the Statistics of Trade Act in the case of ONS business data) and similar requirements are likely to be needed for tax data.
153. We have one specific concern about how this model might operate when applied to tax data. Legal compliance requires that research proposals for access to ONS business data are subject to a ‘departmental benefit’ test – the research should be of benefit to ONS. For ONS, whose main business is the compilation and publication of

statistics for public use, the interpretation of the 'departmental benefit' of research proposals is likely to focus on reputational issues regarding the proper use of the data – are the data clearly to be used for statistical purposes, and are the proposals methodologically sound? However, for a department (like HMRC) whose primary business is something other than the production of statistics, the interpretation of 'departmental benefit' might be rather different, for example focussing on whether or not the research proposals will help that department to deliver its policy and/or operational agenda. It may also be significant that, unlike most government departments, HMRC's functions are defined by statute – which may encourage a restrictive interpretation of departmental benefit, through a strict linking of benefits to functions.

154. We do not think it desirable that the decision to grant access to microdata for research should be dependent on whether or not that research fits in with a department's policy agenda. We believe that providing access to microdata for statistical research through a datalab facility should be regarded as a statistical function which comes under the auspices of the new Statistics Board, rather than as a departmental function. The criteria to be applied when evaluating proposals for access to microdata should be statistical, and focus on whether or not the research proposal represents a proper use of the data, and on whether or not the research is likely to be of general public interest. Whilst the access arrangements for individual microdata sets may vary for legal and security reasons, the identity of the department holding the data should not be a factor.

155. Our fifth recommendation is as follows:

- **HMRC should establish a secure datalab facility for researchers outside of government to access HMRC tax microdata for approved research purposes. Approval of individual proposals for access should be on the basis of statistical criteria and on whether or not the research is likely to be of general public interest.**

## Meeting the demand for longitudinal microdata

156. A main theme of the responses to our questions about potential further uses of tax data was a demand for longitudinal data, both for personal incomes and tax, and for company incomes and tax payments. Improving access to microdata through a datalab facility may be of only limited use in meeting this demand. Even where sufficient microdata are made available to make linking of individual records over a run of years theoretically possible, doing the necessary linking to transform the dataset into longitudinal data is likely to prove to be a long and arduous process for an individual researcher. Moreover it can only be done where data are made available with clear identifiers attached – eg names and addresses – which inevitably carries with it a significant risk of disclosure.

157. It may be both more efficient and less risky in terms of disclosure if the government department responsible for the microdata were to first create a longitudinal dataset from the data available, and make this accessible to researchers under restricted use conditions. Government departments have recognised the potential for making better use of the longitudinal information available within government in their own analytical work; to this end, a project has recently been set up to look at Analytical Data Integration in Government (ADIG). The ADIG approach would see microdata remain with collecting departments, but linked through a hub that provides a facility for integrating and matching data as required. Datasets for specific purposes can then be pulled together, drawing on the linked data.
158. If and when it is implemented, ADIG could go a long way towards helping meet the demands of non-government researchers for longitudinal data – as long as non-government researchers are given access in some form or other to the individual datasets that are created. This is for the future; in the meantime, progress on meeting the demand from non-government users for longitudinal tax data seems most likely to come through enhancement of an existing official dataset. This is likely to be a more attractive option, from the point of view of resources, than creating a new longitudinal dataset from scratch.
159. Two examples of official longitudinal datasets are DWP's Work and Pensions Longitudinal Survey (WPLS), which matches DWP benefit records with employment information from PAYE records provided by HMRC, and ONS's Annual Respondents Database (ARD), which pulls together successive years of business survey microdata from the Annual Business Inquiry (ABI) and its predecessors. The ARD can be accessed by researchers through the Business Data Linking (BDL) arrangements on the ONS Virtual Microdata Laboratory (discussed in the previous section), where ARD data can be linked with data from other business surveys. DWP provide non-government researchers with access to WPLS data under contract for specific research projects. Data are usually supplied in anonymised form. Research contractors are required to use a suitably secure environment for their research, but in most cases are not required to visit DWP premises to access the data.
160. Neither of these official microdata sets currently provides the longitudinal data on personal and company incomes and tax that respondents to our questionnaire have indicated that they would like to have. But in each case there may be potential for enhancing the dataset to include more tax microdata.
161. The WPLS provides a comprehensive source of longitudinal data on benefits and employment income, but only for a specific sub-set of the population – people with a recent benefit history. It does not cover people who have not claimed any form of benefit in recent years. A couple of respondents to our questionnaire suggested that the WPLS would be a more useful data source for research if it could be extended to cover all PAYE taxpayers, and not just those with a benefit history. We believe that this idea is worth exploring, and would invite DWP and HMRC to consider it.

162. The ARD/BDL data currently available on the OND VML do not include any tax records information. Whilst tax records are used extensively in compiling the IDBR, the financial and accounting information on companies available through the ARD is derived from survey data (the ABI). ONS have been looking at the possibility of using company tax data, rather than survey data, as a primary source of financial and accounting information, with the aim of cutting back on survey activity and thus reducing business compliance costs. This might open up the possibility of extending the range of financial and accounting data available through the Business Data Linking arrangements to include data on company tax payments. We would invite ONS and HMRC to consider the feasibility of doing this.
163. This review has indicated a clear demand for more longitudinal data on incomes and tax for use in research. We believe that this demand might be most effectively met through enhancing existing official microdata sets with further data from tax records, and we would invite HMRC to explore this possibility with the departments that 'own' the existing datasets – DWP for personal data on the WPLS and ONS for business data.
164. So our sixth recommendation is:
- **HMRC, in conjunction with the departments responsible, should consider the feasibility of enhancing existing official longitudinal microdata sets, as a possible means of meeting the demand for longitudinal data on incomes and tax for research purposes. Specific datasets of interest are the WPLS maintained by DWP and ONS business microdata available through Business Data Linking.**

## Legal gateways and the Statistics and Registration Service Act

165. Both the WPLS and the IDBR draw substantially on tax records as a data source. Compilation and maintenance of these datasets involves HMRC passing disclosive tax records information to DWP and ONS respectively. This is possible only because a legal gateway permitting HMRC to pass on the information required has been created by past primary legislation.
166. Any further extension of the use of tax data by DWP or ONS, or indeed by any other government department, will only be possible if a legal gateway exists, or is created, which allows HMRC to pass on that information to them. ONS are at present unable to move in the direction of making more extensive use of tax data in place of business surveys as they do not have legal cover to receive the data. All the other government departments that responded to our questionnaire reported that legal restrictions were a barrier to realising the full potential of tax data for statistical purposes.

167. The Statistics and Registration Service Act includes clauses that will allow a statutory gateway for a new instance of data sharing between the Statistics Board (ONS) and other government departments to be created through secondary legislation, rather than requiring primary legislation as it has up to now. This has the potential to make the process of setting up new legal gateways a bit easier, at least as regards data sharing between HMRC and the Statistics Board (ONS). This will only be possible with HMRC co-operation; any secondary legislation under the Act involving HMRC data requires the specific agreement of HMRC Commissioners and the Treasury. We hope that HMRC will be prepared to make use of the provisions of the Act to help extend the use of tax data for statistical purposes within government, and thus enhance the value of the statistics and statistical resources that ONS and other government departments are able to provide to the user community.

168. Therefore our seventh and last recommendation is:

- **HMRC and ONS [Statistics Board] should make use of the provisions in the Statistics and Registration Act to provide a legal gateway for the release of tax data to ONS [Statistics Board] for new statistical purposes, where this is not covered by existing legislation.**

# Part Two: Evidence on the use and potential use of tax data

## Introduction

1. Part Two of this report sets out in detail the evidence gathered for the review on the use and potential use of tax data by academic researchers and by government departments. The evidence was gathered through questionnaires and interviews. Section 1 covers use by academic and other researchers; Section 2 covers use by the Office for National Statistics (ONS); and Section 3 use by other government departments.

## Section 1: Use and potential use of tax data by academic and other researchers

2. This section reports on the use and potential use of tax data by academic and other researchers. It draws upon replies to a questionnaire sent to some 80 academic and other researchers in the UK, selected because they were perceived as having a possible interest in tax data.

### 1.1: The questionnaire replies

3. 24 replies to the questionnaire were received, including two collective replies covering a number of people in the same institution. A further three respondents declined to respond, on the grounds that they did not actually make any use of tax data, eg because they were theoretical rather than applied economists.
4. The following paragraphs report and discuss the main points from these replies. For the most part, the evidence is presented in terms of anonymised quotes from the individual questionnaire replies. None of the statements extracted is specifically attributed to individual (or in a couple of cases, collective) responses. A list of those who contributed replies, or contributed to collective replies, appears at the end.
5. Based on the replies received, and following the order of the questionnaire, this section starts with the current uses of published tax statistics by academic researchers. It continues by looking at current and past use of tax microdata – in particular of the Survey of Personal Incomes (the SPI). It then goes on to look at various possibilities for a more extensive use of tax data mentioned by respondents to the questionnaire. Overall the possibilities mentioned amount to quite a long ‘wish list’ for greater access to tax data, in many cases to disclosive tax records and microdata. The analysis of questionnaire replies concludes with a look at respondents’

perceptions as to what constitute the main obstacles to the wider dissemination and more extensive use of tax data for research purposes.

6. Finally this section pulls together comments in the questionnaire replies on a couple of other issues – the accessibility of tax statistics through HMRC and other websites, and respondents’ experience with using tax data from other countries.

## 1.2: The questions

7. The purpose of the questionnaire was to gather evidence from academic and other researchers on their current and potential uses of tax data. There were seven questions in all, each of them ‘open’ and seeking views and comments:
  - the first two questions asked about use of currently available tax data (see Annex A for a full inventory of tax data);
  - the next two questions asked about use of tax microdata, in particular the SPI;
  - the next two questions were about the potential use of tax data – respondents were invited to submit ‘wish lists’ of statistical data series and microdata sets that they would like to see released or be given access to;
  - the last question asked what were perceived as the main obstacles to the release of more tax data – legal restrictions on HMRC, or something else.

## 1.3: Use of currently available tax data

8. Questions 1 and 2 asked about the use of currently available tax data. A list of tax data series currently published by HMRC and others (based on Annex A to this report) was included with the questionnaire; the text explained that the aim was to “get some feel of the use made of these statistics”.

### Question 1

- Do you use, or have you used, any of the published HMRC tax data series available on the HMRC websites?
- If yes, which? Briefly, what did/do you use the data for?

### Question 2

- Do you use, or have you used, non-HMRC tax data? (eg ONS national accounts analyses of aggregate tax receipts, data for council tax and business rates from DCLG and the devolved administrations)
- If yes, which? Briefly, what for?



9. Most respondents to the questionnaire had used HMRC data from the websites. A wide range of statistics from the HMRC websites were mentioned, covering all aspects of tax data – tax receipts, tax base data, numbers of taxpayers, rates of tax and estimates of income and wealth distribution derived from tax data. Specific series mentioned on more than one occasion included inheritance tax and its tax base (numbers of estates); income tax relief on savings schemes, eg ISAs; personal wealth distribution; and corporation tax and its tax base.
10. Fewer respondents had used non-HMRC data; a number of respondents mentioned aggregate taxes data from ONS, and a couple had used council tax data. VED (vehicle excise duty) information and 'national insurance information from the GAD website' were also mentioned.

"I sometimes use ONS data as convenient aggregate summary statistics"

"I have extensively used supply and use tables and the accompanying tax data [from ONS]. Use: constructing economy-wide datasets for use in economic models."

"ONS measures of aggregate tax receipts are used for analysis of trends over time in the size of the tax burden."

"We use data on average council tax rates."

"I have used council tax data ... for comparing cost of living between regions."

11. Statistics were used for a range of purposes, mostly research and modelling. A number of examples are given below:

"[I have] used the data for research purposes in preparing policy reports and academic papers ... including effects of single market on alcohol duty receipts; economics of charitable donations; survey of saving and wealth in the UK."

"Constructing economy-wide datasets for use in economic models"

"I often use the data to find out how much different elements (ie bands) of taxes raise (or how much particular reliefs "cost") as this can be useful when analysing effects of taxes on individuals."

"Corporation tax base data [are] used to estimate trends in industry wide average effective tax rates (ETRs)."

"The material on estates and inheritance tax has been used for reports on wealth, inheritance, etc"

"I have used the tables on the breakdown of corporation taxes to attempt to identify the role of various reliefs (capital allowances, interest paid, loss, etc)".

"[I] use employee share ownership plan statistics very regularly for public presentations and for background material in research papers."

"Betting, gaming & lottery duties (monthly series) [have been used for] ... research into impact of tax on demand for gambling."

"Estimates of the distribution of personal wealth derived from inheritance tax data [are] used to explore issues regarding the size and distribution of bequests."

"Corporation tax receipts and base data [are] used to model the public sector finances within [a macroeconomic model]."

"The sort of things we have used the data for ... include comparing self-assessment and PAYE tax receipts/taxpayers, trying to work out what proportion of total tax receipts are remitted by business, and constructing series of VAT registrations/receipts."

12. As well as research, tax statistics were also used for teaching, or simply to understand and explain the tax system. Some respondents used tax data for 'reference purposes' – in commentary on economic developments and in the analysis of policy.

"As a teacher of UK taxation I draw on these materials for my students and point to this information to give them specific details of how the tax system is constructed, where monies are raised, how allocated etc."

"Published data on income distribution and wealth distribution [are] used in graduate teaching"

"VAT data sheets [are] used to understand the system, and explain it to others."

"When thinking about and discussing policy issues, it is often useful to have simple statistics to hand about, say, the changing number of estates subject to inheritance tax, or the revenue raised from stamp duty on shares. We use the HMRC website for reference purposes like these."

"We use figures on the distribution of tax receipts by ranges of the tax base to cost various policy options."

13. One respondent regarded the tax data published by HMRC as very interesting and useful – although at the same time under-used.

"These data are some of the most interesting, and least known, of all that the government publishes. Information, for example, on the distribution of income tax payments, almost always astonishes audiences."

14. However there were some negative comments about individual time series, in particular for tax credits (with criticism directed at the same time at HMRC's child benefit statistics) and for council tax (where comments highlighted the three websites issue discussed in Part One, paragraph 32).

“Published data on tax credits for families with and without children (post-2003) are much less helpful, comprehensive and comprehensible. We appreciate that the peculiarities of these tax credits mean that lags are inevitable, but it sometimes seems that the paucity and tardiness of tax credits data stifles public debate.”

“It's a disgrace that so little effort has been put into making the data [for tax credits and child benefit] more useful. For example HMRC data have not been spliced to earlier DWP (and even earlier DSS) data ... to give us a longer consistent series ... HMRC does not even seem to provide “real” data – that is inflation adjusted. This matters a lot for making comparisons across successive governments.”

“It is extremely time consuming to construct comparable data [on average council tax rates] between England, Scotland and Wales as you have to look (for some time) on 3 different websites. This is true even of very simple things like the average household council tax bill.”

15. The view was also expressed that HMRC do not provide enough in the way of long time series for their key data series.

“Long time series are not readily available [from HMRC] ... The ONS is much better at providing time series.”

## 1.4: Current use of microdata sets

16. Questions 3 and 4 related to the use of microdata sets. Question 3 asked about use of the Survey of Personal Incomes (SPI), the only dataset underlying published statistics that HMRC make publicly available on a regular basis. The text explained that the aim was “to get some idea as to the extent of use of the SPI through the UK Data Archive, and of the purposes of that use”. Question 4 asked about access to microdata sets other than the SPI; the aim was “to gain some idea of the extent of ad hoc requests for access to other HMRC datasets, and whether these requests had had any success”.

### Question 3:

- Have you ever used the SPI for research purposes?
- Briefly, what did you use the SPI data for?

- Any views on the usability of the SPI dataset at the UK Data Archive? (eg. does the process of anonymising the data for release substantially detract from its usefulness to you?)

**Question 4:**

- Have you been provided by HMRC with any other tax dataset for research purposes?
  - If so, which? Were there any specific conditions or restrictions placed on use of the data?
  - Have you ever had a request to HMRC for access to a tax dataset turned down? If yes, do you know why?
17. The SPI is an annual sample of income tax records that HMRC use to construct a number of the statistical series that they release, including estimates of income distribution. An anonymised version is made publicly available through the UK Data Archive at the University of Essex.
18. A number of respondents had used the SPI for research, or were considering using it, but the majority had not used it. Uses included modelling income distribution, looking at taxable incomes and the labour decisions of the self-employed, and adjusting household survey data to adjust for underreporting by high income individuals.
19. A couple of respondents said that they had never heard of the SPI – but added that now that they knew about it, they might consider making use of it. One respondent expressed the view that the SPI was under-used by researchers, though did not offer a view as to why this might be.
- “Haven’t used this though sounds interesting!”
- “I haven’t used this data-set but it sounds useful.”
- “The SPI is a tremendously under-used resource ... to date [we have] not used this data that much [but] there is much that could be done with it.”
20. The main issue raised in respect of the SPI was the inability for a researcher to link data from one year’s SPI with data from previous years. This was seen as a major restriction on the potential uses of SPI data. In effect respondents expressed a demand for a longitudinal SPI where anonymised individuals could be tracked through time. Anonymisation of the data was not generally seen as a problem *per se* – although one respondent strongly disagreed.

“The key issue for me is whether individuals can be tracked through time. Making the data anonymous while preserving the panel structure is not a problem.”

“It would obviously be possible to answer a much wider range of research questions were it possible to link SPI data across years, or link family members within a given year. For this to occur the sample would need to be fixed over time.”

“Inability to link across years is regrettable and does limit the usefulness of the data. The possibility of matching observations to other datasets would open up many potential areas for research.”

“Anonymising of data *per se* does not negate its value for many uses – particularly related to personal taxes research.”

“Releasing data [is] very helpful and anonymisation [is] a good step as long as the underlying statistical properties of the data are not disturbed.”

“The data is anonymised in such a way as to make them useless for econometric analysis.”

21. Other issues with the SPI included gaps in the historical series of datasets, the levels of disaggregation to which the datasets were available and useable, and the quality of the documentation provided to users. One respondent wanted coverage of the SPI to be extended to include details of tax credits.

“It would be great to have a complete series of SPI datasets available ... at present there is a gap between [19]85 and [19]95.”

“The industry classification does not appear to be sufficiently detailed”.

“The degrading of the detail of the geographic coding attached to each record essentially ruled out any sub-regional analysis.”

“Anonymisation wasn’t an issue, but the fact that some variables are imputed was: it would be nice if the documentation on this was improved explaining in detail how variables were imputed.”

“Something similar [to the SPI] based on families’ tax credit data would also be very useful (or, in fact, the inclusion of tax credit data in the SPI – particularly if the SPI linked couples’ income tax records).”

22. Answers to question 4 regarding other microdata sets confirmed that access was generally only provided when researchers were working for HMRC (or DWP in the case of combined tax and benefit data) on a specific contract. Access in such circumstances was typically accompanied by restrictions on use of the data and on its further dissemination.

“Yes, but this was because of a joint project with the HMRC – I have signed the HMRC confidentiality form, and there are restrictions on data use and dissemination.”

“There was a brief opportunity when researchers in my unit were allowed access to micro-data on CT. But this was only within HMRC and for a limited period.”

“We have been given an extract from the WFTC dataset ... for a DWP-funded project and we had to give the data back when the project had finished.”

“We are using the WPLS, which combines HMRC data on employment spells with DWP data on benefit receipt – this is made available by DWP ... [for use] in DWP-funded projects.”

23. A few respondents reported that they had been given access to unpublished data with no (or at least not many) strings attached. But none of these cases involved access to disclosive microdata – the data provided were either anonymised or aggregates of the actual data records.

“Yes: breakdown of betting duty by (anonymised) company ... No special conditions, beyond keeping the raw data confidential”

“Aggregated data on inheritance tax records for an HMRC project. No conditions were placed on them.”

“Yes, as part of an ESRC project ... As part of our agreement, an anonymised sample was provided to me (ie individual tax return data on the control group + the 4 treatment groups). I regarded this as being part of the project and have proceeded to publish the results.”

24. A number of respondents reported requests to HMRC for access to unpublished data that had been turned down. Others have refrained from asking for data because they expected to be refused.

“[In the past] we were keen to extract more data from IR in particular, and were largely unsuccessful.”

“We made a request for information on how much was being paid into personal pensions as contributions in excess of the pre-April-2001 contribution limits ... We never received these data.”

“Over time, I have made many requests for more detailed data to be made available ... However, no data has ever been provided: this has usually been on the grounds of privacy, that HMRC/IR could not release information about individual (or corporate) taxpayers.”

“We requested additional data [to that available from the website, which] would have enabled us to look more closely at the distributional impact of a move from council tax to a local income tax. However HMRC refused to provide ...”

“I never asked because ... HMRC were not widely known for acceding to such requests owing to the sensitivity of the material”

## 1.5: Potential uses of tax data

25. Questions 5 and 6 asked about potential uses of tax data for research. In effect respondents were invited to set out a ‘wish list’ of data series and microdata sets that they would like to be made available for research purposes. Question 5 asked about additional data series or analyses that respondents would like to see released, whilst Question 6 asked about microdata that respondents would like to have access to. Respondents were also asked about their need for times series and/or longitudinal data and for their views on the acceptability of conditions and restrictions that might be placed on access to microdata for research.

### Question 5:

- What additional data series/analyses would you like to see released by HMRC or by anyone else?
- Briefly, what would you intend to use these data for?
- Does consistency over time matter? Do you want time series or longitudinal data, or would cross-section data for a single period suffice?

### Question 6:

- Please specify which tax datasets you would like to have access to for research purposes.
- Briefly, what would you intend to use these data or datasets for?
- What conditions or restrictions of access would you be prepared to accept? (eg. would anonymised data be OK, or do you need names for data matching?)
- Would access in a controlled environment (eg a datalab) be acceptable?

## 1.6: The demand for additional data series and analyses (not microdata)

26. Overall, respondents appeared to be reasonably satisfied with the statistical series and analyses derived from tax data that are currently made available by HMRC and others. ‘Wish lists’ for the release of more data predominantly focused on greater access to microdata (tax records) rather than on the provision of further statistical series and analyses derived from those records.

“I think the data published is a pretty full and very useful array. I am much more enthusiastic about making the data available to outside researchers than about increasing the range of published series.”

“Current releases are fine, although I would also press for the spatial coding to be given at lowest possible level that confidentiality protection allows.”

27. Nevertheless some respondents did list further statistical series and analyses that they would like HMRC to make available. Most of the demands were for further detail in respect of series that are already released.

“More data at a disaggregate level, ie the 123 industry classification used in the ONS supply and use tables ... VAT payments, and stuck VAT by industry; national insurance payments by industry; a derivation proportioning income tax payments to these industries; corporation tax payments and allowances by these industries.”

“A detailed breakdown of CT payments”

“More information available on the international activities of UK multinationals, [which] is extremely limited at present. Yet this is one of the most important fields for study; how far is the UK tax base mobile out of the UK?”

“More detail on revenues from betting duty. eg separation of duty from spread betting, exchange betting and fixed odds betting.”

“It would seem reasonable for information regarding net tax and benefit schedules (post tax and benefit income as a function of pre tax and benefit income) to be provided by HMRC.”

“More data on recipients of tax credits for families with and without children.”

“SPI data ... with a high level of industry detail ... Names, addresses, etc would not be needed, but it would be useful to keep regional data.”

28. A number of respondents recorded a specific interest in tax compliance – but expressed the view that HMRC did not currently release sufficient data of the kind that might be relevant to administrative and compliance issues.

“Most of my interest would be in the area of compliance, audits, dispute recognition etc ... for both individuals and companies. Very, very little scholarly work has been published on corporate tax compliance – part [of the explanation for this is] the difficulty of obtaining compliance data.”

“Numbers of taxpayers and receipts are often not broken down by administrative category. For example, we can't find anywhere on the HMRC website information about the number of firms using the flat-rate VAT scheme, or the number of individuals subject to income tax self-assessment ... Often



one has to rely on NAO reports to pick up useful information about administration and compliance issues ... Our current interest in this data is for administrative and compliance related issues, but we are strongly of the opinion that having more data available would be beneficial much more widely.”

“[I would like to see HMRC make available] data on compliance (a UK version of TCMP? A similar data set for VAT and excise duty?)”

## 1.7: Time series, longitudinal data and consistency over time

29. Whilst views differed to some extent as to the importance of consistency over time, the general view was that consistent time series were desirable, and probably essential for some purposes.

“Consistency over time does matter, but ... new outputs which are not necessarily backwards compatible with older outputs is acceptable as a trade-off for increased detail on current situation.”

“Ideally consistent information over time is what is most useful ... so that key event activities can be explored to show before and after impacts of different aspects of tax revenues, tax payer behaviour etc. Without consistent year on year, comparable, data this is clearly impossible research to undertake reliably. However, a wider range of data being made available, even if not always directly as comparable year on year as would be ideal, will still be of great value for many research topics ...”

“Single period data would suffice, although time series would be better.”

“Consistency over time is crucial for these [data] to be useful in time series econometric analysis.”

“Consistency over time is very useful, since often we are interested in constructing time series ...”

30. Longitudinal and/or panel data were generally seen as the ideal. Some respondents went further and maintained that such data were essential for certain kinds of analysis.

“The ideal is usually panel data: a cross-section repeated over time. Time series analysis and (to a lesser extent) cross section analysis can also be useful, but cannot generally provide such rigorous examination.”

“A time-series of cross sections would be most useful. We would also want to see how individual experiences have changed over time; a panel would be even more useful.”

“For the finest spatial resolutions, cross-sectional data is fine. For coarser (regional or even national) resolutions, longitudinal data would be helpful.”

“Consistency over time is extremely important. Availability of longitudinal tax record data would particularly enhance the robustness of microeconomic procedures which could be applied to the data.”

“Longitudinal data is crucial for me.”

## 1.8: Access to microdata – the ‘wish list’

31. The main thing to emerge from respondents’ answers to questions 5 and 6 was a widespread desire for much greater access to tax microdata for research purposes. Some respondents were keen to emphasise the potential that existed for important research, which might be unlocked if academic researchers were to be given wider access to tax microdata.

“[I] would celebrate extending access, since I strongly believe it would lead to a wide range of new and important pieces of empirical work which would help greatly the understanding of the economy and its interaction with the tax system.”

“Being able to get wider access to large tax filing data – even if anonymised – will open many new research opportunities ... in areas such as personal tax compliance, corporate tax planning activities, response to changes in tax regulations, international tax activities etc.”

“Tax microdata (and derived employment information) is an invaluable resource for policy evaluation ... to work out the impact of a particular policy intervention, we need to know how individuals have responded – and this often includes changes in employment. Having data for all individuals in the country rather than just a subset (collected via a survey) enables impacts to be estimated more precisely, allows a much broader range of analysis by subgroup, and in some cases allows research questions to be answered that it wouldn’t otherwise have been possible to address ... micro data on Corporation Tax, Stamp Duty and Inheritance Tax would all be useful [as well as income tax records].”

“I am confident that there would be demand for use of longitudinal data or data which could be matched with other data sources if it were to be made available.”

32. The ‘wish lists’ of respondents for tax microdata were long and varied, but two themes in particular emerged. The first – set out in the examples below – was a strong demand for microdata from company tax records, an area in which there are at present no currently available microdata.

“Would be very useful to get company data, longitudinal”.

“Corporation tax statistics at microdata level – ditto for capital taxes.”

“The extent of any adjustments in tax terms to corporation tax returns – on a company level, ideally adjustments by category. In addition an analysis of the quarterly corporation tax payments, again by company level. Such data would allow an examination of companies’ ability to estimate their liabilities and also give an insight into negotiating strategies. Along with an analysis of companies’ financial reports the data would also allow an analysis of how equity markets assess corporate tax risk. Also an analysis of quarterly corporation tax payments could provide insights into capital market induced earnings management. Would require names for data matching.”

“It would be very good to be able to get company data to both link to the existing data [ONS micro data in the Business Data Lab.] and help in building better data sets on growth, survival, entry and exit, competition etc.”

“I would be very interested in any datasets which included microdata on companies, or unincorporated businesses. I would use these primarily for research into the impact of taxes on business behaviour: investment, finance, location, choice of legal form, etc. The aggregate data available in published form is not sufficient to undertake any serious research work; it can only give a guide to broad aggregates.”

33. The second main theme to emerge from the ‘wish lists’ was a demand for longitudinal data from income tax records. A couple of respondents specifically argued for the extension of the DWP Work and Pensions Longitudinal Study to cover individuals who had no recent record of claiming benefits (and who therefore are not currently picked up in the WPLS).

“I could envisage using longitudinal data (on changes in income/tax payment over time), were this released ... [for] sub-district modelling of income distributions etc.”

“Longitudinal data on personal incomes.”

“Ideally an integrated dataset that offers as close to 100 per cent population coverage as possible, with fusion of records for same individual present in more than one tax dataset. [Use:] In the short-term, for modelling tax/income distributions at the sub-regional scale. In the medium term, if made available, I would like to use longitudinal data on changes in income/tax as an input to a microsimulation model, one element of which would be cumulative life earnings/tax payments.”

“There are two important issues. One is the availability of longitudinal data on individuals’ incomes. This would be very useful in a wide range of studies. Second is the ability to match income data onto other datasets. I am not clear of the technicalities of how this can be done, but matching different datasets is immensely powerful.”

“[Extension of the Work and Pensions Longitudinal Study] dataset with HMRC data on earnings ... would be very useful, and the inclusion of information on income taxes at this stage would also be helpful. This dataset is useful for a broad range of economic research, particularly in the area of labour economics.”

“The major microdata source containing HMRC data that we have used recently is the Work and Pensions Longitudinal Study (WPLS) ... The WPLS as it stands currently ... doesn’t contain employment information for individuals without any spells on a DWP benefit/programme since June 1999. This means it can’t be used to address some research questions ... The WPLS would be a much more useful resource if employment information was available for all individuals.”

34. Some other examples of tax microdata on respondents’ wish lists are given below:

“A sufficient longitudinal data set to capture the many changes in tax treatment of pensions ... age gender breakdowns etc [are already provided] and are useful. However if a tax reform affects a particular group ... it would be useful to know how many households were affected ...”

“Administrative data on firms’ use of tax-advantaged employee share schemes”

“Access to information on R&D tax credit claims by firms.”

“Access [to] another dataset which (we think) is called the 10% self-assessment extract to study the self-employed sector.”

“I would like to have used data on private pension contributions in a project on persistency of contributions. Looking at NI contributions would also improve understanding of saving for retirement.”

“Micro data on inheritance tax ... [for] Modelling the evolution of wealth distribution in the UK – are there trends that will lead to greater divergence? What role does IHT play in restraining this?”

“Company data (derived from approvals process) on employee share ownership plans. For research into factors influencing company use of approved share plans. Names of companies would be essential for data matching ...”

## 1.9: Issues around access to data – anonymisation and datalabs

35. Asked whether anonymised data would be acceptable for the research projects that they had in mind, a number of respondents replied that such data were acceptable for many purposes. However respondents also emphasised the importance in some circumstances of data matching – which, if performed by HMRC prior to release, does not necessarily preclude the data remaining anonymised.

“Anonymised data would be absolutely fine... indeed I would be sceptical of those claiming that non-anonymised data were necessary.”

“Anonymous would be fine, but any demographic information would be helpful (gender, age etc) if available.”

“Anonymised OK if HMRC does data matching across HMRC datasets prior to release; otherwise I could envisage ... undertak[ing] such matching myself within a controlled environment (e.g. datalab).”

“In many circumstances, having access to anonymised data would be perfectly acceptable. There are, however, situations in which having access to names and full addresses would be very useful ... In general, linking datasets massively increases their value. However, this need not preclude the datasets being anonymised: the government could provide anonymised linked dataset, as with the linked tax and benefit data in WPLS, or provide eg household or company identifiers that are consistent across anonymised datasets to allow us to do the linking.”

36. Views on the acceptability of controlled access through a datalab varied significantly, ranging from strong enthusiasm for a datalab along the lines of the ONS model to firm opposition to the idea of a datalab, on the grounds that it would “drastically reduce the value [to the researcher] of the data”. Even amongst those respondents prepared to accept controlled access through a datalab if they had to, access through such means was regarded as ‘far from ideal’; respondents would much prefer to receive access under a legally binding undertaking prohibiting further release of the data or use for any other purpose. A couple of respondents stressed the importance to the academic researcher of being able to publish the results of any research carried out with data made available under conditions of restricted access.

“Access in a controlled environment, while inconvenient, would be an acceptable price to pay in order to access the data.”

“Access in a datalab has been worked on extensively by ONS and would, I think, be an excellent model.”

“Access in a controlled environment would be acceptable.”

“Anonymisation fine; datalab difficult – you need to ensure international reproducibility of results.”

“Controlled access is acceptable. However, the first priority for me would be to have anonymised longitudinal data that was not in a controlled area.”

“I would be prepared for restrictions on access, as required. Obviously the ideal would be for full access to the data, available on disk to work on in my office.”

“If the data has to be anonymised, then so be it, but that is much more difficult when considering large corporations than individuals ...I would ... propose instead that restrictions take the form of an undertaking not to reveal any information on particular taxpayers. I would have no objection to such an undertaking. A restriction based on such an undertaking would maximise the chances of a reasonable research outcome from using the data. It might be possible to link the data by company to financial statements, for example.”

“I would strongly prefer that the restriction not be that it can only be used in a controlled environment. That would drastically reduce the value of the data, since it would require large blocks of time to be spent in the controlled environment, and would rule out any use of the data in any other period. I would rather an arrangement whereby I signed a binding undertaking, and accepted that I would be prosecuted if I broke that undertaking.”

“There are real difficulties in operating a “datalab” system. I have used the Business Data Linking lab at the ONS and it really restricts what an econometrician can do. Better than nothing, though!”

“Having access to the data only via a datalab would be acceptable but far from ideal because (a) setting up the data so it is ready to be analysed is immensely time consuming, meaning that we would need to spend periods of weeks in the datalab, and (b) much of the estimation we do is highly computer-intensive requiring expensive computing resources (and even then things can take a couple of weeks to run!).”

“[What is] critical to me as an academic researcher ... [is that] results can be published in an agreed fashion – ideally with this agreement being achieved before the work is undertaken rather than suppressed once the work is done ... Wider restrictions such as need to sign confidentiality agreements, temporary employment relationship ... to allow a researcher general access to HMRC data etc ... are not to me insurmountable ...”

## 1.10: Obstacles to the wider use of tax data for research – legal and other restrictions

37. Question 7 asked respondents about their perceptions regarding HMRC’s attitude to the release of statistical data for research purposes. What did they see as the main obstacle to the release of more tax data and datasets by HMRC? Was it legal constraints – either from general restrictions on the release of disclosive information arising from legislation such as the Data Protection Act,<sup>9</sup> or from specific legislation restricting what HMRC are permitted to do with the tax records data that they collect? Or did they see it more as a question of departmental ‘culture’, in that senior people in HMRC were not really that interested in giving researchers more access to currently unpublished data?

### Question 7

Which of the following do you see as the most important obstacle to the release of more tax data and datasets by HMRC?

- HMRC aren’t really interested in giving researchers more access to tax data
  - Legal constraints on HMRC which prevent their passing on tax records data to others.
  - Restrictions on sharing personal data (eg deriving from the Data Protection Act) which require all data records to be effectively and demonstrably anonymised before they are shared with anyone else.
38. Many respondents recognised that legal restrictions were an important constraint on HMRC’s ability to pass on tax data and tax record information to other parties, particularly outside of government. Nevertheless there appeared to be quite a widespread feeling that ways round these constraints could readily be found.

“There clearly are legal and anonymity constraints, but these are certainly capable of solution.”

“I guess legal constraints are important for both confidentiality and merging.”

“I would have thought that potential legal obstacles can be overcome by appropriate legally binding research access protocols such as those used in the US.”

“I am guessing legal problems but I wonder if the ONS datalab is a way forward of sorting”

“Legal constraints ... has been the answer given to me on previous occasions, but then at other times it has not appeared to be insuperable ...”

<sup>9</sup> Data Protection Act, 1998 <http://www.opsi.gov.uk/ACTS/acts1998/19980029.htm>

“Legal and DPA restrictions are clearly important, but I am sure that ways can be found of respecting all legal requirements whilst enabling much useful research to be done.”

39. The idea that HMRC ‘culture’ might be an issue here appears to have also struck a chord with a number of people. A large number of respondents stressed that HMRC statisticians were generally keen to help academic researchers get access to the data they needed – but at the same time there appears to exist as well quite a widespread belief that ministers and senior officials at HMRC are rather less keen than the statisticians to allow this. One respondent compared HMRC unfavourably with DWP in this respect.

“I think that HMRC are aware of the value of the data they have to external researchers.”

“Seminars suggest that HMRC recognise the needs and the benefits of external researchers having access to data.”

“I don’t doubt there is a genuine interest in facilitating informed research relevant to tax policy.”

“There are individuals in HMRC keen to do this [release more data and microdata for research], but the political will at higher levels of the civil service seems largely to have been absent.”

“Where I have had conversations with HMRC staff they have regularly expressed interest in research ideas but are frustrated at not being able to be more helpful in the research process.”

“I’m not sure [that HMRC are not interested]. Whenever I have discussed this, officials have generally been supportive. But it has never happened, which I think probably reflects reluctance somewhere in the HMRC/Treasury.”

“Traditionally at least, HMRC has not been keen on releasing data.”

“HMRC do seem to have a ‘culture’ sceptical of the benefits of wider publication and dissemination. They are right to be mindful of the needs of data protection, of course, but the needs of outside analysts conducting non-HMRC work would seem to be a low priority.” ... However ... on the one occasion I tried to obtain data before, some time ago, their staff were always open and helpful about the process.”

“Although [legal constraints] ... are given as formal reasons ... [there is a] perception ... that politicians in charge of HMRC would not be supportive of moves to give more people access to more information about tax and taxpayers.”



“I think the main issue here is a cultural one. If you compare HMRC with the DWP, then the DWP has a long-established culture of producing statistics ... [and of] commissioning and publishing research. HMRC lacks this track record/culture ... Government departments tend to be very nervous about handing over any information ... I think HMRC has been even more guarded than others in this regard.”

40. Finally a couple of respondents perceived the main obstacle to be neither legal constraints nor departmental ‘culture’ but rather HMRC resources.

“Where some data processing/analysis is required ... HMRC resources may also play a role [as an obstacle to realising the potential use of tax data]”

“HMRC resource constraints [are the biggest obstacle] ... I suspect many researchers would like to have access to a microdataset based upon harmonised tax records harmonised drawn from all HRMC in-house databases – but this would be a fairly resource intensive product to produce, so is unlikely to happen ...”

### 1.11: Other issues – accessibility of tax data and websites

41. Although not specifically asked about it, a number of respondents used their questionnaire response as an opportunity to offer some comments about the accessibility of tax statistics – and in one case of government statistics in general –through the HMRC and other websites. Some of the comments on tax statistics (on national insurance contributions data and on HMRC tax data on the uktradeinfo.com website) mirror the discussion of issues of data availability and accessibility in Part One (paragraphs 31-33).

“Government statistics are scattered over a number of websites. What would be really helpful is a single page that provided directions to these sites. I’m sure it would not be hard to add an index to such a page or a key word search.”

“I have just tried to obtain data from ... [an HMRC] link ... It caused me to recall what happened last time I tried: a complete failure to obtain the data I wanted. If the websites promise data they should deliver it without too much difficulty.”

“It would be helpful if statistics on National Insurance contributions could be provided much as for income tax, with similar uses. Much less information is currently provided on the NICs Office website and the GAD website.”

“A good deal of data on non-HMRC taxes (VED, council tax etc) is not to be found on the HMRC website. It would be a fine example of joined-up government if these data could be found in the same place and presented in a similar way; at the very least there could be links between them. This also applies to HMRC’s own taxes: [we] only became aware recently and fortuitously

that statistics on ex-HMCE taxes can be found on the uktradeinfo.com website ... there seems to be no indication anywhere on the HMRC site that the information is held there! ... even if you know that information is available on the HMRC website, it is often very difficult to track down: as a trivial example, it is hard to navigate to the [www.hmrc.gov.uk/stats](http://www.hmrc.gov.uk/stats) page without typing it in the address bar.”

“It is difficult to find the statistics web page if you log on to [www.hmrc.gov.uk](http://www.hmrc.gov.uk) main home page. Could HMRC put some kind of link on the main home page to ‘research’ or ‘statistics’ or something?”

“Understanding exactly what the information tells me is not always straightforward. Some of the descriptions aren’t easy to understand, making it difficult to rationalise differences across tables/data sources ... Trying to reconcile figures from HMRC Statistics, HMRC Accounts and Budget tables to calculate total spending on tax credits and its breakdown [is difficult].”

## 1.12: Other issues – comparisons with other countries

42. A few respondents also commented on their experience with using tax data from other countries – specifically on the wider availability of tax statistics for the US, and on the easier access to tax microdata for research purposes that is provided in Denmark.

“An American colleague has been surprised at how little information about tax and tax payers is available in the UK compared to the US ... The range of information available [through the IRS website] is much wider than on the HMRC website. It would be useful if HMRC were to aim towards providing much more of this sort of data.”

“The availability of data from IRS in the USA seems to be much more substantial than we have in the UK – maybe there is some value in exploring that model for potential application/lessons for the UK market?”

“I currently use Danish administrative data that contains the tax returns for everyone since the mid 70s. I can, subject to agreement, merge this data with other administrative data (using the equivalent of a NI number) ... I have restricted access ... The output is monitored but this system works extremely well.”

## Appendix to Section 1

### List of respondents

Replies to the questionnaire were received from the following:

#### Replies from individuals:

Andrew Dilnot (St. Hugh's, Oxford)  
Ian Walker (Warwick)  
Matthew Wakefield (IFS)  
Paul Williamson (Liverpool)  
Karen Rowlingson (Bath)  
Andrew Lymer (Birmingham)  
Adam Blake (Nottingham Business School)  
Kevin Holland (University of Wales, Aberystwyth)  
John Hasseldine (Nottingham)  
Steve McKay (Bristol)  
Jonathan Haskel (Queen Mary, London)  
Sarah Smith (Bristol)  
Richard Disney (Nottingham)  
Frank Cowell (LSE)  
Hamish Low (Cambridge, Dept of Economics)  
Michael Devereux (Said Business School, Oxford)  
Simon Burgess (Bristol)  
Andrew Pendleton (York)  
Gareth Myles (Exeter)  
David Paton (Nottingham Business School)  
Jane Frecknall-Hughes (Sheffield)  
Ian Preston (UCL)

#### Replies from institutions:

IFS (joint reply from Carl Emmerson, Stuart Adam, Mike Brewer, Claire Crawford, Jonathan Shaw, Helen Simpson)  
NIESR (joint reply from Martin Weale, Rebecca Riley, Justin van den Ven)

## Section 2: Use and potential use of tax data by government departments: Office for National Statistics

43. This section reports on the use and potential use of tax data by the Office for National Statistics (ONS). It draws on a series of interviews with representatives from the ONS divisions that are the principal users of HMRC data in the department. Amongst government departments other than HMRC, ONS makes the most extensive use of tax data for statistical purposes.
44. The section first describes the current uses of tax data by ONS. It then goes on to discuss some possibilities for more extensive use of tax data by the department. Finally it considers the legal restrictions on HMRC passing on tax records to third parties, including other government departments, which present a major obstacle to a more extensive use of tax data for statistical purposes.

### 2.1: Current uses of tax data by ONS

45. Tax data are used by ONS for two main purposes. Non-disclosive data on tax receipts and the tax base are used as a data source for statistics compiled and published by the department. And disclosive microdata, with taxpayer names and addresses, are used in compilation of the sampling frame for statistical surveys, mainly of businesses.
46. The main uses of tax records (microdata with taxpayer names and addresses) are:
- PAYE income tax employer and VAT trader records in compiling and updating the **Inter-Departmental Business Register (IDBR)**: and
  - PAYE income tax employee records in selecting a sample for the **Annual Survey of Hours and Earnings (ASHE)**.
47. The main uses of tax base and receipts data are:
- samples of income tax and corporation tax records in compiling estimates for the **national accounts** of non-financial companies trading profits; income from employment, self-employment, pensions, expenses and benefits; and social contributions;
  - samples of income tax records in compiling **regional economic accounts**;
  - receipts data for taxes and national insurance contributions in compiling the **public sector finances** data releases.

## 2.2: Use of tax data in business surveys

48. The **IDBR** is a list of UK businesses maintained by ONS, which is used for selecting samples for surveys of businesses, to produce analyses of business activity and to produce lists of businesses. As well as names and addresses, and information on industry classification and legal status (eg company sole proprietor, partnership), the register also holds information on employment and turnover. The main data sources for the IDBR are HMRC tax records – VAT registrations and PAYE employer records. Matching of these data is carried out by ONS.
49. Other data sources used in compiling and updating the IDBR, where they are matched with the tax data sources, include information on companies from Companies House registrations, and information on the links between companies provided through a contract with Dun and Bradstreet. Companies House registrations data are becoming more important as the level of incorporation of businesses increases, whilst knowledge of the links between businesses helps meet a strong and growing interest in the impact of globalisation. ONS also receives lists of Intrastat traders – those mainly VAT traders who have transactions with other countries of the European Union. These administrative data sources are supplemented by survey data, mainly from the Business Register Survey and from the employment part of the Annual Business Inquiry (ABI/1).
50. One of the main uses of the IDBR within ONS – it is also used outside of the department – is as a sampling frame for ONS business surveys. The Annual Business Inquiry – which has two parts, a survey of employment (ABI/1) and a survey of financial information (ABI/2) – is probably the key business survey, but ONS also conducts a number of monthly and quarterly surveys, including the monthly production (sales) inquiry, the retail sales inquiry and the short-period surveys of the service sector, which are used in compiling, from the production side, early estimates of output and thus of GDP.
51. **ASHE** is an annual sample survey of employers which asks for information on employment (eg hours worked) and earnings in respect of specified employees. The sample of employees is selected on the basis of employee national insurance numbers; PAYE employer records supplied by HMRC are used to identify the employers of the employees in the sample. ASHE was launched in 2004 as the successor to the earlier New Earnings Survey (NES), which collected similar information.

## 2.3: Use of tax data in national and regional accounts

52. Tax data in the **national accounts** are mainly used for measuring incomes. These data come from a number of sources:
- data on profits at industry level from a sample of non-financial companies' profits data taken from tax returns;
  - data on income from employment and social contributions from a 1 per cent sample of annual PAYE records, taken from HMRC's National Insurance Recording System (NIRS);
  - data on self-employment income from income tax self-assessment returns;
  - data on pensions, expenses and benefits from HMRC's survey of personal incomes (HMRC's own sample survey of income tax records)
53. Tax data are a key source for **regional economic accounts**. The main data source is the 1 per cent NIRS sample (see above), which provides information on pay from employment, tax and national insurance contributions by region.
54. The monthly **public sector finances** first release, a joint release by ONS and HM Treasury, provides broad details on the public finances – expenditure, receipts, borrowing – for the previous month. HMRC supplies data for this release on the amounts of tax (and national insurance contributions) collected in the previous month. These data subsequently feed into the quarterly national accounts.

## 2.4: Extending use of tax data: (i) Surveys of business – the ABI and other surveys

55. The main area where ONS is actively looking at the possibility of making more use of tax data as a data source is **business surveys**. ONS has set up a project aimed at exploring possible ways of making use of administrative data to replace some business survey activity. The principal motive driving this project is to reduce compliance costs for business, especially small businesses, by using tax data in place of estimates that currently derive from business surveys.
56. The principal survey in question is the financial part of the Annual Business Inquiry (ABI/2), which collects some basic financial information – including turnover, gross value added (GVA), capital expenditure – from a sample taken from the IDBR. There are also a suite of monthly and quarterly business surveys, which provide data for the short-term indicators of output that feed into the early (production-based) estimates of GDP.

57. Much of the data collected from ABI/2 could be substituted by information on company profit-and-loss accounts from corporation tax (CT) returns, supplemented by similar data, for the self-employed and partnerships, from self-assessment income tax returns. Research by an ONS secondee to HMRC has confirmed that CT data and IT self-assessment data could be a viable data source for some of the financial information currently collected in ABI/2. However, that would require that ONS be given access to the necessary data records – at present they are not allowed access.
58. Another potential use of tax data to reduce business compliance costs surveys is the use of VAT data as a source for some of the estimates currently derived from the monthly and quarterly business surveys. What is needed for the short-term output surveys is an indicator of gross value added (GVA). Very approximately, GVA = turnover less inputs. The IDBR already holds monthly turnover data from VAT returns, although this is a combination of monthly, quarterly and annual returns that depends on the reporting arrangements of the VAT trader. A proxy indicator for GVA could be constructed from VAT data if the IDBR also held data on input tax reclaimed. However ONS does not have a legal gateway for access to other VAT data. Research by an ONS secondee to HMRC is planned to determine whether this approach might be feasible, were ONS to get access to the data in future.
59. There are no plans at present for more extensive use of tax data in **ASHE**.

## 2.5: Extending use of tax data: (ii) regional accounts and the Allsopp programme

60. The Allsopp Review,<sup>10</sup> commissioned by the Chancellor in 2003, reported on 31 March 2004. A major theme of the report was the need for better **regional economic data**. To this end, the review made a number of recommendations, including the development of a production-based measure of regional GVA (measured in real terms). Allsopp saw more use of administrative information, including tax records, as a data source as one of the keys to better regional data, as it might allow more information to be collected without a corresponding increase in business compliance costs.
61. The Allsopp Implementation Programme includes a programme of work to improve the quality of data in the IDBR on the location of the activity of companies; in IDBR terms, better data on ‘sites’ rather than ‘enterprises’. Availability of employment and classification for each site enables the apportionment of the activities of an enterprise to any specified geographical area. However this work does not involve, initially at least, the more extensive use of tax data; neither the VAT nor the PAYE systems hold information for the individual sites of a business. Better data on the IDBR for ‘sites’

<sup>10</sup> Christopher Allsopp, ‘Review of Statistics for Economic Policymaking’, *Final Report to the Chancellor of the Exchequer, the Governor of the Bank of England and the National Statistician*, HM Treasury (2004) [http://www.hm-treasury.gov.uk./consultations\\_and\\_legislation/allsopp\\_review/consult\\_allsopp\\_index.cfm](http://www.hm-treasury.gov.uk./consultations_and_legislation/allsopp_review/consult_allsopp_index.cfm)

rather relies on development of the other IDBR data sources, in particular the supplementary survey sources, where the short-term alignment (now realised) of the Business Register Survey and ABI/1 has been a priority. The intention is to now introduce a new survey, the Business Register and Employment Survey (BRES), to replace the Business Register Survey and the existing ABI/1. Pilots of BRES are planned for 2007 and 2008, with the first full BRES scheduled for 2009.

## 2.6: Extending use of tax data: (iii) national accounts and public sector finances

62. There is limited activity and interest at present in developing tax data sources for **national accounts** at the national, as opposed to the regional, level. Whilst any move away from providing data derived from on a sample of tax records towards data derived from the complete set of records would be welcome from the point of view of data quality (which could be expected to improve), it is not seen as a development priority. The present samples are regarded as of adequate size and generally fit for the purposes to which they are put.
63. Neither is improving tax data sources seen as a major priority for **public sector finances** – administrative data are already used to the full in this area. That said, there is probably potential for improving the early monthly, and quarterly, estimates of tax accruals (as opposed to cash receipts) that are used to construct the key national accounts-based fiscal aggregates (net borrowing, current budget) – HMRC’s primary short-term focus is on cash receipts.

## 2.7: New and potential new uses of tax data

64. The **Wealth and Assets Survey** (referred to in ‘the field’ as the Household Assets Survey) is a new survey of households, which aims to measure households’ wealth and assets. The first (‘wave 1’) survey was launched in July 2006. HMRC is a joint sponsor and HMRC tax records are drawn on to select the sample, where the aim is to focus in particular on – and over-sample from – households with particular characteristics. The tax records used are income tax self-assessment returns and records of PEPs and ISAs. The sample is designed so that high wealth individuals – those with high investment income and/or substantial investments in PEPs and ISAs – have a significantly greater chance of selection than do the rest of the population.
65. Population statistics is an example of an area where there is little or no use of tax information as a data source at present – but where the potential exists for more extensive use. In a 2003 discussion paper, ONS set out some proposals for an **Integrated Population Statistics System**, which would “combine census, survey and administrative data to create a single comprehensive population statistics database”. The first step would involve creation of a linked statistical database, combining administrative and survey data by linking at the individual person and



household level. The administrative data that would be used in such a system would include tax records. Following the 2011 Census, this would be combined with the census database to create a linked population statistics database. The Government Register Office for Scotland (GROS), who administer the Census in Scotland, has also noted the potential use of tax data, alongside other administrative and survey sources, in compiling population statistics (see Section 3, paragraph 99).

66. More recently the December 2006 report of the **Interdepartmental Task Force on Migration Statistics** discussed some possible solutions for addressing the gaps in migration statistics, including obtaining information about migrants from administrative systems. Amongst the administrative systems identified as holding information about migrants, or having the potential to provide it, is the Work and Pensions Longitudinal Study (WPLS), which links DWP benefits data with HMRC tax data. (The WPLS is discussed at greater length in Section 3, paragraphs 75-76). One of the recommendations of the Task Force Report is that “the WPLS should be used to provide information on patterns of employment, children and benefits among migrants.” It also recommends that “work should be started on record linkage [from a number of sources] to improve migration and population statistics”; seven possible sources are listed, including the WPLS.
67. The proposals for population and migration statistics outlined in the preceding paragraphs revolve around a substantial degree of data linking between tax records and other administrative and survey data, which in turn requires data sharing between departments. This raises important legal issues, discussed further below. The success of the WPLS illustrates that inter-departmental data linking is feasible. But here a ‘legal gateway’ exists, permitting HMRC to pass on specified tax records to DWP for data linking with benefits information in the WPLS. There is no ‘legal gateway’ at present that would allow HMRC to pass on detailed tax data to ONS for the purpose of improving population statistics.

## 2.8: The legal position

68. Under existing legislation, HMRC is not allowed to pass on disclosive tax data to other government parties, even for purely statistical purposes, except where primary legislation has been passed which specifically permits this. The provision of specified disclosive microdata to ONS for the IDBR is permitted by the 1969 Finance Act (the employment counts held for PAYE income tax employer records) and the 1994 VAT Act (for VAT registration data and the returned turnover of the VAT trader). The 1969 Finance Act also provides legal cover for HMRC to pass on PAYE data for the purpose of identifying the employee sample for ASHE. The 1987 Finance Act (Number 2) provided additionally for local authorities to have access to the data for planning purposes.

69. The need to provide a legal gateway through new primary legislation is a major obstacle to the more extensive use by ONS of administrative (tax) data as a statistical source. For example, the provision by HMRC to ONS of corporation tax data as an alternative data source to business surveys (discussed in paragraph 57 above) would first require the inclusion of a clause specifically permitting this in a future Finance Act or other primary legislation. New legislation would similarly be required for HMRC to provide ONS with individual tax records for data linking purposes to improve population and migration statistics.
  
70. The Statistics and Registration Service Act 2007 should make the provision of legal cover for new instances of data sharing somewhat easier in future. The Act includes an 'enabling clause', the effect of which is that new instances of data sharing between the newly created Statistics Board (the legal successor to ONS) and other government departments, including HMRC, will no longer require new primary legislation. Instead a legal gateway can be provided by secondary legislation under the new Act, with the agreement of the relevant Ministers (Commissioners as well in the case of HMRC data). Thus legal cover for the provision by HMRC to ONS (or its successor) of corporation tax data to replace business surveys will only require a statutory instrument under the new Act, rather than new primary legislation.

## Section 3: Use and potential use of tax data by other government departments (excluding ONS)

71. This section reports on the use and potential use of tax data by government departments, other than the Office for National Statistics (ONS).

### 3.1: The questionnaire to departments

72. The section draws upon replies to a questionnaire sent to senior statisticians – in most cases heads of profession – in selected government departments (including two devolved administrations) that were known to make some use of HMRC tax data for statistical purposes. It covers the use and potential use of tax data by three UK government departments – the Department for Work and Pensions (DWP), the Department for the Environment, Food and Rural Affairs (DEFRA) and the Department for Business, Enterprise and Regulatory Reform (DBERR) (formerly the Department of Trade and Industry – DTI) – and by the Scottish Executive and the Welsh Assembly Government. Use by DBERR is covered by two separate replies to the questionnaire – one for the former DTI and a separate one for what was then the Small Business Service (SBS), and an Executive Agency of DTI, and is now the Enterprise Directorate of DBERR. The Scottish Executive's reply to the questionnaire also covered use of tax data by the General Register Office for Scotland (GROS).

73. Based on these replies, the section first discusses the current uses of tax data by these departments. It then goes on to look at the various possibilities for more extensive use of tax data in their statistical work mentioned by departments in response to the questionnaire. Overall these amount to quite a long 'wish list'. The section concludes by considering the extent to which legal restrictions on HMRC passing on tax records to third parties currently prevent other government departments from making a more extensive use of tax data for statistical purposes.

### 3.2: Current uses of tax data

74. Departments were asked a number of questions about their current use of tax data or tax records from HMRC:

- Did they use HMRC tax data as a data source for regularly published statistics?
- Did they use HMRC tax data for any other statistical purpose, eg research?
- Were the data used publicly available? If not, what kind of restrictions were placed on their use?
- Were the data generally fit for the purposes they were used for? Were there any changes that might be made to improve their usefulness?

75. Amongst the departments covered in this survey, the most extensive use of tax data for statistical purposes is made by DWP. HMRC tax records are used by DWP as a primary data source for the Work and Pensions Longitudinal Study (WPLS), which links employment data from HMRC PAYE (P45/6 and P14) and self assessment income tax returns with benefit and programme records for DWP's customers. This data matching requires that the tax data provided to DWP for the WPLS are 'disclosive' microdata, with personal identifiers attached. The WPLS is the only instance where HMRC provides disclosive tax data for statistical purposes to a government department other than ONS. The legal gateway that enables HMRC to do this is provided by the 1992 Social Security Administration Act<sup>11</sup> and the 2002 Employment Act.<sup>12</sup>
76. Tax information held in the WPLS is used by DWP as the basis for the Job Outcome Target (JOT) that is used to assess the performance of Jobcentre Plus, by analysing employment starts for Jobcentre Plus customers. Tax data are also used in the compilation of published statistics on the destinations of leavers from the Government's New Deal programmes. The WPLS is also used for a wide range of statistical and research purposes within the department; these include provision of statistics, management information and research on Jobcentre Plus, and understanding the links between savings held and the benefits system in retirement. DWP publishes and updates regularly a full list of uses to which WPLS data are currently put.
77. As well as the use of tax records in the WPLS, DWP also uses (non-disclosive) tax data from two other HMRC sources – the Survey of Personal Incomes (SPI) and the 1 per cent sample of National Insurance Records (NIRS). The SPI is used in construction of the Households Below Average Income (HBAI) dataset. The main data source for HBAI is the Family Resources Survey (FRS); SPI data are used to adjust the FRS sample at the top of income distribution to correct for volatility in the highest incomes. The NIRS sample (also used by ONS – see Part 1, paragraph 48) is a data source for certain published information on pension contributions.
78. Current use of tax data by departments other than DWP (and ONS) is rather more limited. SPI data are used by DEFRA and by the former SBS. DEFRA uses income data from the SPI in compiling estimates and analyses of farmers' incomes that are published annually; and SPI data on earnings by local authority in calculation of the indicator for their PSA target for reducing the productivity gap between the poorest rural areas and the English median. SBS uses SPI data to help estimate unregistered businesses for their publication *Small and Medium-sized Enterprise (SME) Statistics*. The Welsh Assembly Government also makes use of the SPI in much the same way as the SBS, in production of business structure estimates. Welsh regional and sub-regional analyses from the SPI are also reproduced on their website.

11 Social Security Administration Act, 1992 [http://www.opsi.gov.uk/ACTS/acts1992/ukpga\\_19920005\\_en\\_1](http://www.opsi.gov.uk/ACTS/acts1992/ukpga_19920005_en_1)

12 Employment Act 2002 <http://www.opsi.gov.uk/acts/acts2002/20020022.htm>

79. There is also some indirect use of tax data by departments through ONS. SBS uses data on PAYE and VAT records obtained through the Inter Departmental Business Register (IDBR) in compiling regularly published statistics for three publications – *SME Statistics*, *VAT Registrations and Deregistrations*, and *VAT Business Survival Rates*. IDBR data have also been used to measure performance against certain DTI targets for SMEs. The Welsh Assembly Government also makes considerable use of the IDBR, as well as being a substantial user of ONS regional accounts outputs (regional Gross Value added and Gross Household Disposable Income), for which HMRC data are a major data source.
80. Three rather more specialist uses of tax data by government departments are:
- (by DBERR) Monthly hydrocarbon oils duty data for clearances of road fuels are used generally as a check on deliveries data collected from the oil companies and specifically as a data source for liquid petroleum gases used as road fuels, and consumption of bio-fuels;
  - (by Scottish Executive) The annual publication *Government Expenditure and Revenues in Scotland (GERS)* makes use of a wide range of tax data, including the SPI, in compiling estimates of the Scottish share of UK tax revenues;
  - (by GROS) Council tax data are used for producing household estimates (eg. numbers and size of households) for Scotland by local authority. [GROS is in the process of collecting data at a more disaggregated level – data zone in most cases – so that household estimates can be produced for small areas.]
81. In most cases the tax data used by departments for statistical purposes (other than for the WPLS) are publicly available; this includes the SPI, which is made available on the UK Data Archive. There are a couple of exceptions. The IDBR information used by SBS is provided in an unrounded (and therefore unpublished) form; in order to protect against inadvertent disclosure, the estimates produced from these data are rounded for publication. The NIRS sample data used by DWP (and ONS) are not publicly available.
82. Overall the departments replying to the survey were content with the tax data provided to them; the data were considered to be generally fit for the purposes they were used for. Comments from users of the SPI suggested that a larger sample might be helpful; DEFRA noted that the small sample size limited the reliability of analysis with the data, particularly more detailed distributional analysis, whilst the Welsh Assembly Government observed that a larger sample would provide more detailed local area data. Improved timeliness would also increase usefulness.
83. The Welsh Assembly Government also indicated that their main concern was with the quality of inputs to the (ONS) regional accounts process, which relied heavily on tax data to apportion GVA over regions. Any improvements here – a larger sample or quality assurance of the address information – would be beneficial. [ONS regional accounts mainly rely on the NIRS sample, rather than the SPI.]

### 3.3: Extending the use of tax data

84. Departments were asked about the possibilities for their making more extensive use of tax data for statistical purposes:
- Is there potential for greater use of HMRC tax data or tax records as a data source, or for other statistical purposes?
  - Is there any development work in progress involving the use of tax data or tax records as a statistical source?

#### **DWP developments – extending the WPLS**

85. DWP reported that a wide range of developments are planned or under way. Most of these involved further development of the WPLS, including:
- the integration of Child Tax Credit (CTC) and Working Tax Credit (WTC) data into the WPLS; and
  - the enhancement of self-assessed tax returns into the WPLS.

DWP is also looking to develop a 100 per cent feed of NIRS data in place of the current 1 per cent sample. They hope that these developments will lead to an overall improvement in the quality of their National Statistics outputs.

86. DWP aims to expand the use by the department of the WPLS and the tax data in it – operationally, in predictive analysis, and in research into the effectiveness of the department's interventions. Two examples given are:
- matching individual CTC and WTC records with individual records from the FRS, in order to better measure child poverty, through adjusting for under-recording of CTC and WTC in the FRS; and
  - developing a measure of the take-up of formal child care using the child care element of CTCs.

#### **Other uses of tax credit data**

87. The Scottish Executive and the Welsh Assembly Government are looking to use CTC data in the calculation of the Index of Multiple Deprivation for Scotland and Wales respectively. In both cases the need is for (non-disclosive) small area data, rather than the individual tax records that are required for the WPLS. The Scottish Executive is also interested in using small area tax credit data on their Scottish Neighbourhood Statistics site, and to supplement benefits data in building up a picture of low income families in Scotland.

## **Tax data and business surveys**

88. Development work in ONS aimed at reducing the burden of business through replacing some surveys of business by a more extensive use of tax data is discussed in Section 2, paragraphs 55-58. The Scottish Executive makes extensive use of ONS business survey data and is keen to see this approach to reducing the burden of business surveys progress.

## **Survey of Personal Incomes**

89. DEFRA would like to have more SPI information at a 'suitable geographical level' for their work on rural policy. This means ward level data – which may require a larger sample for the SPI than the current 10 per cent. DEFRA would also like to see the SPI 'combining individual tax records at the household level', to allow for analysis of farm household incomes. They note that the use of these data for distributional analysis would require that the SPI be based on a larger sample, or even a full sample.
90. The Welsh Assembly Government has also indicated that a larger sample for the SPI would be helpful (see paragraph 82). The Scottish Executive, who use the SPI for calculating the Scottish share of income taxes for their GERS publication, would like to see a Scottish sample from the SPI.

## **3.4: New uses of tax data – the 'wish list'**

91. Collectively the departments surveyed put together quite a long 'wish list' for potential new uses of tax data. In a number of instances departments would like more detailed information than is currently available to them, but are not seeking disclosive data from individual tax records (though in some cases the level of detail sought might come close to effective disclosure of data about individual taxpayers). Examples are DBERR's demand for tax data 'showing breakdowns by industry or size of business for the take up of tax reliefs and incentives', for use in policy analysis, and a number of items on the SBS 'wish list':
- data on the number of people registering for NICs Class 2 contributions (newly self-employed) – to monitor targets for enterprise creation at a local level;
  - data from self-assessment tax returns for individual self-employed, partnerships and companies – to improve methodology for SME statistics;
  - data on the reasons for de-registration for VAT – to understand how many are due to business closures;
  - data on self-employed/partnership profits by industry sector and employment size band and 'other distinguishing firm characteristics' – to enhance tax benefit modelling.

92. With regard to the last item in the above list, SBS outlines an ‘ideal scenario’ where HMRC would provide a ‘representative sample of each subset of the SME sector (industrial sector by employment size band) with information that would allow us to mimic the tax computation for the company’. SBS goes on to provide a full list of the data they would like to have – some 15 items for limited companies and a similar number for sole traders and partnerships. However they acknowledge that some of the data they would like may not be collected by HMRC as part of the tax assessment process.
93. The Scottish Executive registers a similar demand for more detailed, but non-disclosive, tax data in respect of their GERS publication (see paragraph 80). Basically they would like specific information, tax-by-tax, on the amounts of tax collected in Scotland, or paid by Scottish residents, eg income tax received from taxpayers resident in Scotland, corporation tax paid by establishments located in Scotland, capital gains tax paid by establishments located in Scotland. For many taxes, data are not presently available broken down in the way required for the GERS calculations, and the Scottish share of tax revenues has to be estimated from the best available proxy information.
94. Most of the other items on departmental wish lists for tax data would require the release to departments of ‘disclosive’ data – In effect departments are looking to access individual tax records. DEFRA would find it useful to have tax records that identify the extent to which farmers are making use of the various tax schemes that are available to farmers (they list six tax reliefs that are specifically targeted at farmers and agriculture). DEFRA believes that this information would help them develop a better understanding of farm finances and how profit translates into actual post-tax income.
95. DBERR lists three areas where they would like to have access to tax records for statistical purposes:
- company level clearance information for dutiable oils – DBERR would use this to verify the individual company data that they collect in the process of putting together published energy statistics. The information would also be useful from a policy perspective;
  - VAT records, which they would use to construct microdata matrices of user-supplier relationships between firms – this would be helpful for evaluation research, eg statistical analysis of spillovers from DBERR interventions;
  - (income) tax microdata, to use as a sampling frame for surveys of indebtedness and other financial issues – this should help improve the efficiency of survey designs. It might also reduce survey burdens if it means that DBERR does not need to ask again for information that HMRC already holds in the tax returns.



96. The Scottish Executive would also find access to anonymised individual 'access to income tax data', eg tax returns for people, extremely useful. They see this as a valuable asset in development of income estimates at the level of 'smaller geographies' – for which there is a high demand. The Scottish Executive would also like to have access to vacant properties information from council tax returns – this could be used to improve the sampling frame for various sample surveys.
97. The Scottish Executive would also like to have access to anonymised individual WPLS data, which combines DWP and HMRC data, for research and analysis purposes. However they have so far been refused on the grounds of absence of a legal gateway.

### 3.5: Tax data and population statistics

98. ONS proposals for the possible use of tax data in an integrated population statistics system are discussed in Section 2, paragraph 65. The ONS proposals were referred to by GROS, who compile population statistics and administer the decennial Census for Scotland.
99. GROS sets out a number of potential uses for tax data – alongside other administrative and survey sources – in estimation of the size and the socio-demographic characteristics of the population in Scotland:
  - as a primary source of census information – to replace census returns where there are particular gaps in the data collected;
  - as a supplementary source in census enumeration – to help assess the quality of information collected in the census and the coverage achieved, and to develop statistical adjustments where needed;
  - to provide additional data not collected on the census return – eg to meet user needs for income data at a small area level;
  - in compiling population estimates between censuses – to help estimate numbers of migrant and their socio-economic characteristics;
  - for council tax data, to improve the quality of the address register used for the census and related surveys.

### 3.6: Legal restrictions as an obstacle to more extensive use of tax data

100. The legal constraints on HMRC passing on tax data to ONS are discussed in Section 2, paragraph 68; they are the same in respect of other government departments. HMRC cannot pass on disclosive tax information, except where specifically allowed for in primary legislation. The tax records provided to DWP for data linking with benefits data in the WPLS are the only instance at present where there is legal cover for HMRC to share disclosive tax information for statistical purposes with a department other than ONS.
101. Departments were asked, in the questionnaire, whether legal restrictions on data sharing by HMRC presented a major barrier to realising the full potential of use of tax data for statistical purposes. All responding departments reported that legal restrictions were a barrier. Even though DWP has a legal gateway for the data they use for the WPLS, they observed that the statutory gateway places a number of restrictions on the use of the data and on the acquisition and use of data from other government departments. For example, there are restrictions around merging of (tax) data with other sources that restrict some of the analysis possible within the department.
102. All the other departments reported examples where legal restrictions had been cited as preventing their access to tax data. DBERR has encountered legal obstacles to the release to them of oil duty information. DEFRA noted that legal restrictions prevent the statistical analysis of self-assessment tax levels at the household level. SBS has found existing legislation to be a stumbling block to HMRC sharing data with them. And both the Welsh Assembly Government and the Scottish Executive have encountered problems in gaining access to data at the small area level they require for their respective Indices of Multiple Deprivation. For Wales, data are required 'at the most accurate level', ie unrounded, but some data are only available rounded. For Scotland, the Scottish Executive has previously seconded an employee to DWP to combine DWP and HMRC data in order to create the indicators that are required for the Scottish Index of Multiple Deprivation (SIMD).
103. The provisions of the Statistics and Registration Service Act 2007 regarding data sharing between other government departments, including HMRC, and the new Statistics Board are outlined in Section 2, paragraph 70. The main effect is to replace the requirement for primary legislation in order to establish a legal gateway for new instances of data sharing with secondary legislation under the new Act.

104. The provisions of the Act in respect of data sharing cover both data that are passed by government departments to the Statistics Board and data that are passed by the Statistics Board to government departments – but not data sharing between government departments. However it might be possible to use the relevant clauses to provide legal cover for new instances of data sharing between HMRC and government departments other than ONS, as long as that data sharing takes place through the Statistics Board as intermediary. So whilst not fully removing the legal obstacles to more extensive use of tax data by government departments, the new Act might at least make setting up of a legal gateway for new instances of data sharing a bit easier.

# Annex A

## Tax data – what is available: an inventory of tax statistics

1. This Annex seeks to identify and list the statistics currently compiled and published from tax records by HMRC (and any statistics published by other tax collecting departments), and the further published analyses compiled from these data. It also seeks to identify and list any statistics on taxes published by departments other than HMRC and other tax collectors.
2. Statistics on taxes or derived from tax data can be categorised into three general types:
  - **Statistics on tax receipts**, including details of repayments and of tax credits.
  - **Statistics relating to the tax base** for the different taxes, including data on taxpayer numbers and/or numbers of transactions of a particular type, and statistics derived from returns claiming various tax reliefs. The latter include estimates of the cost of tax reliefs ('tax expenditures').
  - **Statistical analyses based on tax data** that go rather further than just extracting details regarding the tax base. Examples include analyses of personal income distribution derived from income tax records, and analysis of vehicle excise duty fraud.
3. This Annex starts by considering the availability of data and analyses of tax receipts in aggregate. It then goes on to look at the detailed statistics and analyses that are available from tax on individual taxes.

## Statistics on aggregate tax receipts

4. ONS publishes a number of analyses of aggregate tax receipts, in which receipts are broken down into five categories based on national accounts – 'taxes on production', 'taxes on income and wealth', 'taxes on capital', 'other taxes' and 'compulsory social contributions'. Data are available on the ONS website and are published in :
  - the monthly National Statistics First Release Public Sector Finances (released jointly by ONS and HMT), which includes data for central government current tax receipts. [This covers the bulk of tax receipts but excludes capital taxes (mainly inheritance tax) and local authority taxes (mainly council tax).] As well as totals for the categories listed above (except taxes on capital), there are also sub-totals for 'VAT' (part of taxes on production) and 'income and capital gains tax' (part of taxes on income and wealth). Data are monthly and annual (financial year).

- *Economic Trends*, in the 'selected Financial Statistics' tables. Table 5.4 includes quarterly and annual (calendar year) data for total general government tax receipts (including capital taxes and council tax) by the five national accounts categories.
  - *Financial Statistics*. Tables 2.3C and 2.3D, public sector transactions by sub-sector, include quarterly and annual data for tax receipts broken down by the five national accounts categories. The tables also show central government and local authority receipts separately.
5. Most taxes are collected by HMRC; the main exceptions are vehicle excise duty (collected by DVLA), business rates (collected by local government) and council tax (collected and retained by local government). HMRC publishes on their website statistics for total taxes collected by HMRC, broken down by individual tax. For some taxes, there are minor definitional differences from the equivalent ONS data; the HMRC data show cash receipts as measured in producing HMRC's accounts, whereas ONS data reflect payments into the Consolidated Fund.
6. The main tables published by HMRC cover:
- HMRC annual receipts (financial year) by tax, including forecasts for the year ahead.
  - Monthly and quarterly receipts, by tax, of former Inland Revenue taxes.
  - Monthly and quarterly payments, by tax, of receipts of former Customs & Excise taxes into the Consolidated Fund [so these data are not fully in line with those in the HMRC annual receipts table].
7. The breakdown of taxes in the annual receipts table is shown below, with an indication as to which individual taxes are former Inland Revenue (IR) and which are former Customs & Excise (CE). 2006-07 receipts are shown as a guide to relative magnitudes. The breakdown in the quarterly tables is similar, except that income tax and capital gains tax are combined.

<b>Tax receipts of HMRC by tax</b>	
	2006-07 receipts (£ bn)
Income tax (IR)	143.3
National insurance contributions (IR)	87.3
Capital gains tax (IR)	3.8
Value added tax (CE)	77.4
Corporation tax (IR)	44.3
Petroleum revenue tax (IR)	2.2
Fuel duties (CE)	23.6
Tobacco duties (CE)	8.1
Spirit duty (CE)	2.3
Beer duties (CE)	3.1
Wine duties (CE)	2.4
Cider & perry duties (CE)	0.2
Stamp duties (IR)	13.4
Betting & gaming duties (CE)	1.4
Customs duties & levies (CE)	2.3
Air passenger duty (CE)	1.0
Insurance premium tax (CE)	2.3
Landfill tax (CE)	0.8
Climate change levy (CE)	0.7
Aggregates levy (CE)	0.3
Inheritance tax (IR)	3.5
<b>Total HMRC tax receipts</b>	<b>423.7</b>

8. Treasury publishes, in the Budget and Pre-Budget Reports, annual financial year data for tax receipts by tax for the most recent year and forecasts for the current year and year ahead. This is the only place in which data on a tax-by-tax basis for non-HMRC taxes (vehicle excise duty, council tax, business rates, etc) are published alongside data for HMRC taxes. Data for 2006-07<sup>13</sup> shown below illustrate the relative magnitudes of HMRC and non-HMRC taxes.

<b>Tax receipts of HMRC and other tax collectors</b>	
	£ bn
Total HMRC tax receipts	423.7
Vehicle excise duties	5.1
Business rates	21.0
Council tax	22.2
Other taxes & royalties	13.9
<b>Total taxes and NICs</b>	<b>485.9</b>

13 Estimates for 2006-07 from 2007 Pre-Budget Report, published in October 2007.

## Other across-the-board analyses from tax data

9. As well as breakdowns of total tax receipts, HMRC also publishes a number of analyses from tax data that cover a range of taxes rather than individual taxes. These include:
- estimates of the **cost of the main ‘tax expenditures’ and structural reliefs**, for the most recent year and the current year (projected) only. The estimates cover the cost of reliefs for income tax, corporation tax, national insurance contributions, value added tax (VAT), capital gains tax, petroleum revenue tax, inheritance tax and stamp duties;
  - statistics of **numbers of taxpayers** for income tax, corporation tax, capital gains tax and inheritance tax, and of **numbers of registered traders** for VAT, air passenger duty, landfill tax, climate change levy and aggregates levy. All data are annual, tax year;
  - statistics for the **costs of tax reliefs for charities**, by category of relief. Data coverage includes repayments of income tax under Gift Aid, tax relief to individuals on inheritance tax and income tax, and tax relief to charities on business rates, VAT and stamp duty land tax. (VAT relief for charities is estimated.) Data are annual tax year;
  - statistics for **Government revenues from UK oil and gas production**, by tax. Currently these cover petroleum revenue tax and corporation tax receipts from North Sea companies, including the supplementary charge levied since 2002. Royalties were included up until 2002-03;
  - statistics on **Government revenues from environmental taxes**, by tax. Environmental taxes are split into three groups – taxes on energy (hydrocarbon oils duty (including VAT on that duty), climate change levy and its predecessors (fossil fuel levy, gas levy)); taxes on road vehicles (vehicle excise duty); other environmental taxes (air passenger duty, landfill tax, aggregates levy).

## Statistics on individual taxes – HMRC taxes

10. Although the former Inland Revenue and Customs & Excise departments are now combined in a single department, the presentation of the statistical outputs of the two departments is not as yet very joined up. This goes as far as having separate, and apparently unlinked, websites for the two sets of outputs. Ex-IR statistics are presented in the National Statistics section of the main HMRC website, [www.hmrc.gov.uk/stats](http://www.hmrc.gov.uk/stats). However, although the HMRC/Stats site also contains some analyses of aggregate tax receipts – see above – the detailed statistics and analyses for individual C&E taxes are published on a completely separate website, [www.uktradeinfo.com](http://www.uktradeinfo.com). (This is the main website for the statistical outputs of the former HM Customs & Excise; its name reflects the fact that the majority of these outputs are trade statistics.)

## Ex-Inland Revenue taxes

11. All statistics and statistical analyses published by HMRC that are derived from ex-Inland Revenue tax data are published in a series of tables on the HMRC website (Stats section). In effect, these tables are the electronic successor to the annual publication *Inland Revenue Statistics*, last published in hard copy format in 2000. The focus is, for the most part, on annual data. (This is in contrast to presentation on the ex-C&E uktradeinfo website, where the receipts series are usually monthly in frequency.)
12. Statistics derived from data on individual taxes fall into three main categories:
  - **further details of tax receipts**, including separate identification of gross receipts, repayments and net receipts, and, for certain taxes, details of tax accruals;
  - **statistics on the tax base** for that tax, of two main kinds:
    - taxpayer numbers and their distribution with respect to some specific aspect of the tax base (eg. income band, industry groups);
    - statistics derived from data provided on returns claiming particular tax reliefs.
  - **other analyses from tax data**. It can be difficult to draw a line sometimes, but the following are examples of statistical analyses published by HMRC that appear to go further than just recycling tax base information:
    - statistics on the distribution of income, derived from income tax data;
    - estimates of the distribution of wealth, from inheritance tax data;
    - analyses of property transactions, derived from returns for the purpose of assessing liability for stamp duties.

## Income tax statistics

13. HMRC publishes a wide range of statistics derived from data on income tax. All data are annual tax year, though for some series statistics are updated twice a year, rather than annually. HMRC statistics derived from income tax data fall into all three of the main categories of statistics from tax data identified in the previous paragraph.
14. **Further statistics on receipts, including repayments**. The figures for income tax receipts in the various presentations of aggregate tax receipts are for receipts net of repayments and tax credits. Further statistics published on the HMRC website include an analysis of receipts by 'type' – three categories of gross receipts (PAYE, tax deducted at source, self-assessment receipts) plus repayments. There is also a more detailed breakdown of repayments and tax credits.



15. **Statistics from tax base data.** These include further statistics on numbers of taxpayers, including a breakdown by region; data on taxable benefits in kind by main kind of benefit, with further analysis for company cars; and some detailed analyses of personal tax credits (child tax credit, working tax credit and their predecessors), including geographical analysis and take-up rates. Statistics on employer-provided benefits are based on data from HMRC's Employer Compliance System, a database of information captured from end of year returns provided by employers.
16. HMRC also publishes a variety of statistics derived from data on tax reliefs, including:
- statistics on savings and investments in various schemes offering tax relief, including:
    - employee share schemes (SAYE Share Option schemes, Company Share Option Plan, Share Incentive Plan, Enterprise Management Incentives, and earlier schemes now closed);
    - other incentives to invest in shares (Enterprise Investment Scheme, Venture Capital Trust scheme);
    - tax-free savings accounts (Individual Saving Accounts (ISAs) and earlier schemes – TESSAs and PEPs);
  - statistics on tax relief given on 'approved pension schemes', including occupational pensions, and on contributions made to personal and stakeholder pensions, including FSAVCs.
17. HMRC publishes estimates of the cost of the main income tax reliefs, as a major part of the overall analysis of main tax expenditures and structural reliefs (see paragraph 9, first bullet).
18. **Other analyses of tax data.** Income tax data provides a rich source for statistics and analyses of the distribution of income, both by income bands and geographically. Most statistics are derived from the Survey of Personal Incomes (SPI), which is an annual sample survey taken by HMRC statisticians from the individual income tax records held by the department. The SPI is the data source for a number of analyses of the distribution of tax liabilities and of taxpayers, and for a wide variety of statistics on personal incomes and income distribution.
19. Analyses from the SPI include:
- tax liabilities by income range and by taxpayer's marginal rate;
  - shares of total income and income tax liability by percentile groups;
  - percentage of earnings paid in income tax (specimen taxpayers);

20. The SPI is also the data source for a number of published analyses of personal incomes by tax year, covering:
- key parameters of the income distribution (percentile points), median incomes by age group and gender);
  - distribution of taxpayers over income ranges;
  - details of income, deductions and tax paid, by income range;
  - geographical analyses – income and tax by region and county, income by local authority and Parliamentary Constituency.

## Corporation tax statistics

21. The range of statistics produced by HMRC from corporation tax data is rather smaller than that produced from income tax, but nevertheless still quite extensive. The published statistics cover both tax receipts and accruals, and various aspects of the tax base, including taxpayer numbers and the various tax reliefs.
22. **Corporation tax receipts and accruals.** The published statistics break down onshore companies corporation tax receipts (net of ACT) into five industry groups – manufacturing, distribution, other industrial and commercial, financial (excluding life assurance) and life assurance. Net receipts totals are also shown for North Sea companies (net of ACT) and for Advance Corporation Tax (ACT). Data are for financial years. (ACT was abolished in 1999, but some repayments are still being made.)
23. The statistics for corporation tax accruals break down total net receipts by four sectors – industrial and commercial companies, financial companies excluding life assurance companies, life assurance companies and North Sea companies. Further analyses, for industrial and commercial companies and for financial companies excluding life assurance only, set out broad details of the computation of tax liability at an aggregate sectoral level, identifying among other things gross taxable trading profits, capital allowances set against profits, trading losses from previous years offset against the current year's profits, the corporation tax charge, and various set-offs against that charge (including ACT set-off – now largely used up – and double taxation relief). Data are annual calendar year
24. **Statistics from tax base data.** These include statistics relating to:
- tax liability and taxpayer numbers by industry group;
  - certain tax reliefs;
  - corporation tax on capital gains.

25. Statistics on corporation taxpayer numbers and liability show, by industry group (12 industries plus 'overseas activities'), the number of taxpaying companies in each group, total tax payable and broad details of the computation of tax liability. Data relate to the tax liability for a financial year; only data for the most recent four years (two years for computation details) are available on the website.
26. Statistics available on the website that relate to corporation tax reliefs include:
- capital allowances due, by type of asset and industry group. The analysis by industry group is only available for the latest four years;
  - research and development (R&D) tax credits – cost of R&D tax credits on an accruals and a receipts basis, and numbers of claims;
  - tax relief under the Corporate Venturing Scheme – numbers of companies raising funds, totals raised, and number of companies investing in them;
  - claims and amounts invested.
27. Companies are charged corporation tax on capital gains on disposal of assets. HMRC publishes data on the distribution on net gains on disposals by asset type, and for length of ownership, broken down into financial and non-financial assets. Similar figures are produced for CGT payments by individuals – see the following paragraph.

## Capital gains tax statistics

28. **Capital gains tax** (CGT) is charged on disposal of assets by individuals, personal representatives and trustees. HMRC publishes annual financial year data on various aspects of the tax base – numbers of taxpayers, amounts of gains and tax paid (accruals) on them, the distribution of taxpayers by size of gain and by taxable income (the latter for individuals only). Data for the most recent three years are provisional; final data are not available for several years afterwards. HMRC also publishes annual financial year data for numbers and value of disposals, by eight asset types (three financial assets, five non-financial assets); a further analysis, by length of period on ownership, is produced broken down into financial and non-financial assets.

## Statistics from inheritance tax data

29. HMRC publishes some limited analysis of **inheritance tax** (IHT) receipts, separating out IHT on lifetime transfers from IHT on transfers of death (the latter accounts for the vast bulk of receipts). But the majority of the statistics derived from IHT data provide various details of the tax base, ie the estates on which IHT is levied. These statistics cover:

- For estates above the IHT threshold, details of the tax calculation at an aggregate level, starting with the value of the net estate, continuing with details of the various reliefs and exemptions, and concluding with the net chargeable value, and the relative proportions of estates charged at a zero and at a non-zero rate. All series are shown both in terms of numbers of estates and of the aggregate amounts involved.
  - For the wider category of estates notified for probate, statistics on numbers taxed and untaxed; on assets and liabilities of the estates analysed by type of asset and by the value of estate; and on gender, age and marital status of the deceased.
  - For (IHT) taxpayers only, statistics on gender, age and marital status of the deceased.
  - Numbers and amounts of assets held in discretionary trusts, by range of net chargeable value.
  - The distribution of the value of bequests, by gender and marital status of the deceased.
30. IHT data are used in the production of estimates of the distribution of personal wealth, published by HMRC on their website alongside the other data (receipts, tax base) derived from ex-Inland Revenue taxes. The main tables show 'identified personal wealth' analysed by wealth range (measured by value of net estate) and by type of asset and liability; there are further analyses by gender and broad age band (18-44, 45-64, 65+). ['Identified personal wealth' refers to the wealth of that part of the population whose estates are sufficiently large that they require a report is made to HMRC on their death – around 40 per cent of the adult population.] Estimates are compiled from IHT returns by treating the estates of those who die in a year as a sample of the estates of the total population, stratified by age, gender and marital status. Wealth data from IHT assessments are grossed up by a series of multipliers, supplied by ONS and based on mortality rates by age/gender/marital status, etc.
31. HMRC also publishes reconciliations of 'identified wealth' with household wealth from the ONS balance sheet, and with wealth on a national accounts basis. These reconciliations include estimates of the value of the property and of non-property assets excluded from 'identified wealth', as well estimates of the value of funded pensions.

## Stamp duties

32. **Stamp duties** are charged on transfers of land and buildings, including leases, on new leases of land and buildings, and on transactions in stocks and shares. There is a threshold, and a progressive rate structure, for land and property transactions. For stocks and shares there is no threshold, a main rate (which accounts for over 95 per cent of the yield) and an 'entry charge'.

33. HMRC publishes a number of different analyses of tax receipts (all data are annual financial year):
- receipts by type of transaction – a breakdown of receipts into land and property, and stock and shares. Further detail (eg a breakdown of receipts from land and property into land and buildings, other property and leases) is available up to 2003-04, but changes to the tax system mean the same breakdown will not be available from 2004-05 onwards;
  - estimated receipts from property transactions – broken down into residential and non-residential, and by ‘consideration band’ (value of transaction – <£250k, £250k-500k, >£500k);
  - estimated receipts from residential property analysed by region (with Scotland and Northern Ireland combined).
34. HMRC produces a number of analyses of property transactions derived from stamp duty data. Annual statistics are published on numbers of property transactions by price band and region, and also by type of seller and type of buyer (eg. individuals, property companies). (Statistics for the number of property transactions by type of seller are not available from 2004-05 onwards.) Statistics are based on HMRC’s Survey of Property Transactions, which is an annual sample taken from the forms that have to be completed on any transfer of property or granting of a notifiable lease, and on which assessment of stamp duty liability is based. The HMRC forms and Survey only cover England and Wales; for property transactions by price band and region, where statistics are presented on a Great Britain basis, data from the Survey are supplemented by similar data for Scotland. (Statistics for property transactions by type of buyer and seller are for England and Wales only.)
35. HMRC also publishes quarterly and monthly data for aggregate property transactions, based on the forms returned for stamp duty assessment (the full set – not a sample). The published data cover Northern Ireland, as well as England and Wales, but not Scotland.

## Ex-Customs & Excise taxes

36. Statistics derived from tax data for the taxes previously collected by HM Customs & Excise are presented rather differently from those previously collected by Inland Revenue. For each individual tax, HMRC publishes both a Statistical Bulletin, with monthly data on receipts and (in most cases) the tax base, and a Factsheet, where the focus is on annual (and in some cases quarterly) data. Statistical Bulletins are published monthly; the Factsheets are updated less regularly, typically annually. As well as longer time series for the Statistical Bulletin series, the Factsheets generally include some further data on that tax, eg numbers of registered traders/operators, and different presentations of tax data, eg growth rates for receipts, receipts at

constant prices, tax as a proportion of selling price for typical taxed products. The Factsheets also include contextual information relevant to analysis of that tax, including data series for household expenditure on relevant goods and services, and EU and international comparisons of tax rates and consumption levels of the relevant goods and services.

37. Access to these Statistical Bulletins and Factsheets is not through the statistics pages of the main HMRC website, but through a completely different web address, [www.tradeinfo.com](http://www.tradeinfo.com). This is the website for all the statistics produced by the former HM Customs & Excise; the majority of the statistical data on the website relates to UK trade flows, hence the website name. There does not appear to be any link to the tradeinfo site from the statistics pages of the main HMRC website, which, coupled with a site address which makes no mention of tax (nor of HMRC) and isn't even a gov.uk address, makes finding these publications difficult for anyone who doesn't already know where to look.

## Value Added Tax (VAT)

38. The Statistical Bulletin for VAT has monthly data for:

- VAT receipts, split into home and import VAT, with home VAT net receipts further split into gross receipts and repayments;
- the VAT register – new registrations, de-registrations and the 'live trader population'.

The monthly receipts series are also presented on a seasonally adjusted basis; this is the only case of a time series for receipts from an individual tax that is presented on a seasonally adjusted basis.

39. The Factsheet for VAT presents annual statistics for a number of further data series derived from tax data. These include:

- net tax receipts by trade group (some 60 groups in all);
- VAT receipts as a proportion of household expenditure;
- the proportions of expenditure on goods and services subject to the different rates of VAT.
- the distribution of VAT traders by taxable turnover;
- estimates of the costs of the various structural reliefs for VAT.

## Ex-C&E environmental taxes

40. Monthly data in the Statistical Bulletins for the four ex-C&E environmental taxes – aggregates levy, air passenger duty, climate change levy and landfill tax – follow a similar format. There are data for:
- tax receipts;
  - tax liabilities declared on trader tax returns (the difference between receipts and liabilities declared reflects timing);
  - the tax base for that tax (except climate change levy):
    - tonnages (taxable/relieved/exempt) for aggregates levy;
    - passenger numbers (EEA/non-EEA) for APD;
    - waste tonnages (standard rate/lower rate/exempt) for landfill tax.

For climate change levy, there are no monthly tax base data, but tax liabilities are shown disaggregated into levy on gas, on electricity and on solid and other fuels.

41. Annual statistics presented in the Factsheets for these taxes include, for each tax:
- receipts at constant prices;
  - numbers of registered operators (also numbers of sites for landfill tax).

## Insurance Premium Tax (IPT)

42. Presentation of statistics for IPT is similar to that for the environmental taxes. The Statistical Bulletin has monthly data for tax receipts and for tax liabilities declared on tax returns (split by the two tax bands). As with the environmental taxes, the difference between liabilities declared and receipts reflects timing. However there is no monthly tax base information in the Statistical Bulletin for IPT.
43. Information on the tax base for IPT ('estimated taxable premiums') is instead presented in the Factsheet. These data are quarterly, reflecting the quarterly assessment and payment of the tax. The Factsheet also includes annual statistics on IPT registered traders and quarterly data on 'revenue declared and received', which serve to illustrate payment lags.

## Alcohol and tobacco duties

44. There are five Statistical Bulletins covering the various excise duties on alcohol and tobacco, covering beer and cider, wine (from fresh grapes), made wine, spirits and tobacco. Monthly data in each of these Bulletins follows a similar format, with data for:
- tax receipts (a single series in most cases, but with separate series for beer and cider and a five-way breakdown of tobacco receipts);
  - the tax base, in the form of data for the volume of 'clearances' or 'quantities released for consumption' (there is usually some disaggregation of these tax base volume indicators, even in cases where there is no equivalent disaggregation of tax receipts).

For beer and for spirits only, the Bulletins also include monthly volume data on UK production.

45. There are also five Factsheets for alcohol and tobacco duties, covering beer and cider, wine (including 'made wine'), spirits, total alcohol and tobacco. These Factsheets present a number of additional data series and analyses derived from tax data, including:
- tax receipts at constant prices;
  - tax as a percentage of price for a typical taxed product (eg tax per pint of beer);
  - (for the alcohol duties) clearances and clearances per adult, in terms of volume of pure alcohol;
  - (for spirits and tobacco only) 'revenue evaded and avoided' (estimates of duty losses from cross-border shopping and from evasion (fraud));
  - (spirits only) whisky stock in warehouse;
  - (tobacco only) 'smoothed' receipts, adjusted for pre-Budget forestalling behaviour.

## Hydrocarbon oils duties

46. The Statistical Bulletin for hydrocarbon oils duties presents monthly data for tax receipts and for the tax base (quantities released for consumption). The level of disaggregation for presentation of these data is greater than for most other ex-C&E taxes, with both receipts and quantities released first disaggregated into duties on petrol, on diesel and on other fuels, and then split into a further four or five categories within each of these main areas.



47. The Factsheet has annual statistics on tax receipts at constant prices. Otherwise the additional information in the Factsheet is mostly of the contextual kind, with information on retail deliveries of petrol and diesel, on oil prices and on vehicle stock and traffic flows.

## Betting, gaming and lottery duties

48. The Statistical Bulletin for betting, gaming and lottery duties follows a similar format to those for other duties, with monthly data for tax receipts and the tax base. This heading covers six different taxes – general betting duty, pool betting duty, gaming duty, amusement machine licences, bingo duty and National Lottery duty. Monthly data are published in the Bulletin for:
- tax receipts for each of the six taxes;
  - taxed stakes (the tax base) for general betting duty, pool betting duty, bingo duty and National Lottery duty.
49. The Factsheet includes annual statistics for total receipts from betting, gaming and Lottery duties at constant prices.

## National insurance contributions

50. National insurance contributions (NICs) are collected by HMRC, for the most part alongside PAYE income tax. But, unlike the ex-Inland Revenue and ex-Customs & Excise taxes discussed in the foregoing paragraphs, receipts of NICs are not paid over into the Consolidated Fund. Instead receipts of NICs are paid into the separate National Insurance Fund.
51. NICs are classified as a tax in national accounts, and Treasury follows this treatment in presentation of the public finances in the Budget and Pre-Budget Reports. However until recently NICs receipts were excluded from HMRC analyses of aggregate tax receipts. This has now changed; data on NICs receipts are now included in the published tables for HMRC annual receipts by tax, and for monthly and quarterly receipts of former IR taxes. But no further details of NICs receipts – even a breakdown of the total into contributions by employers and by employees – appear to be publicly available. Neither does there appear to be any published information on the tax base for NICs.

## Non-HMRC taxes

52. The Treasury's 2007 Pre-Budget Report shows non-HMRC tax revenues of £62 bn in 2006-07 – this is around 13 per cent of total tax revenues. The bulk (78 per cent) of these non-HMRC tax revenues are accounted for by three taxes – vehicle excise duty, business rates and council tax. The remaining 22 per cent (£14 bn) include rates in Northern Ireland, VAT refunded to central government departments and local authorities (and therefore netted out from the HMRC VAT receipts total), Lottery revenues paid into the National Lottery Distribution Fund (classed as a compulsory levy and thus a tax in national accounts), and certain licence fees.

## Vehicle excise duty

53. Vehicle excise duty (VED) is collected in Great Britain by the Driver and Vehicle Licensing Authority (DVLA), formerly an executive agency of the Department for Transport (DfT) and now a 'trading fund', and in Northern Ireland by Driver and Vehicle Licensing Northern Ireland (DVLNI), acting as Agent for the Secretary of State for Transport. A financial year figure for aggregate VED receipts, including Northern Ireland, is included in the 'Vehicle Excise Duty Account', published in DVLA's *Annual Report & Accounts*. (VED is accounted for separately from the rest of DVLA activities.) DVLA does not publish any regular statistical series for VED receipts, either in aggregate or broken down in some way, or for the VED tax base.
54. DVLA supplies data from the VED Account to ONS (and to DfT at the same time) as an input to the national accounts tables published in *Economic Trends* and *Financial Statistics*. However neither of these tables shows VED receipts separately. The Treasury does separately identify VED receipts in the Budget Report and Pre-Budget Report analyses of the public finances, but these show only two or three years data, including forecasts.
55. DfT publishes regular annual reports on VED evasion, covering the whole of the UK, which include estimates of the loss to the Exchequer arising from fraud. These estimates do not use actual tax receipts, but are instead based on average licence values applied to survey observations and licensing data.

## Business rates

56. Business (non-domestic) rates are collected by local authorities, but the tax rate (the 'national non-domestic rate') to be applied to each property's rateable value is set centrally – in England by the Department for Communities and Local Government (DCLG), in Scotland by the Scottish Executive, and in Wales by the Welsh Assembly Government. Within each country, business rates revenues are paid in to a 'central pool', which is then re-distributed to local authorities by formula. (Northern Ireland does not have a single 'national non-domestic rate' but instead retains the old pre-

1990 rates system, with the tax rate determined both by the Northern Ireland Assembly (the regional rate element) and by the local council. Northern Ireland rates are not included in the business rates total.)

57. There are annual data for total UK business rates receipts in the Treasury's Budget and Pre-Budget Reports, but the tables generally show only two or three years data, of which at least one will be a forecast. In the national accounts tables published by ONS in *Economic Trends* and *Financial Statistics*, which have longer time series, business rates are included in 'taxes on production' and not shown separately.
58. At the country level, statistics on business rates receipts for England are published by DCLG in *Local Government Financial Statistics England*. Similar data for Scotland are published by the *Scottish Executive in Scottish Local Government Financial Statistics*, and for Wales by the Welsh Assembly Government in *Welsh Local Government Financial Statistics*. All three administrations publish aggregate annual data for tax collected by local authorities, net of reliefs. For Scotland and Wales, receipts are disaggregated by individual local authority. For England, receipts (net yield) are disaggregated by region (receipts by local authority are not published in *Local Government Financial Statistics England*, but are available on the DCLG Local Government Finance website). All data are annual financial year; there are time series going back ten years or more for total receipts in England and Wales (also by local authority for Wales), but data for Scotland in *Scottish Local Government Financial Statistics* relate to the most recent outturn year only (plus best estimates for the latest year).
59. Further data on receipts of business rates published by DCLG in *Local Government Financial Statistics England* include details of the total amounts granted under the various reliefs, and figures for (a) the total contribution to the pool from the local authorities list; (b) the total contribution to the pool from the 'central list' (the relatively small amount of tax that is collected directly by central government); and (c) the 'distributable amount' (ie the total amount distributed from the pool to local authorities). All data are annual financial year. For (a) and (b) a time series going back around 15 years is published, but for the other variables mentioned above the available time series go back no more than two or three years. The Welsh Assembly Government also publishes figures for the total contribution from local authorities to the pool, and for the total amount distributed; unlike England these are also broken by individual local authority. Data go back up to ten years. Further details on the amounts granted under the various reliefs, by individual authority and going back ten years, are available on the Welsh Assembly Government's StatsWales website.
60. There is a substantial amount of information available on the tax base for business rates. Data are financial year, and in most cases only relate to a single year (the most recent one). In England and Wales, the property valuations on which the tax is levied (the rateable values) are set by the Valuation Office Agency, an executive agency of HMRC. The VOA website has full details of the latest (2005) rating lists, with a facility

to find the rateable value of any property. HMRC has published on their website various analyses of the most recent rating lists (2005) for England and Wales, showing property numbers and total rateable values analysed by property type (shops, offices, warehouses, etc) and by region, and the distribution of properties over rateable value bands. HMRC has also published analysis of the distribution of changes in rateable values from 2000 to 2005, by property type and by region. DCLG publishes summary analyses of the tax base for business rates for England in *Local Government Financial Statistics England*; these cover rateable value by region, by property type and by rateable value bands.

61. Rateable values in Scotland are set by assessors; the Scottish Assessors Association website contains full details of the current rating lists, together with a facility to find the rateable value of any property. The Scottish Executive publication *Scottish Local Government Financial Statistics* includes an analysis of numbers of properties and rateable values by property type, based on the latest 'valuation roll'.

## Council tax

62. Council tax is set, collected and retained by local authorities. As with business rates, policy responsibility for local government finance is devolved and statistics for England, Scotland and Wales are published separately, by DCLG for England and by the respective devolved administrations for Scotland and Wales. (There is no council tax in Northern Ireland, which retains a system of domestic rates, set partly by the Northern Ireland Assembly and partly by the local council.)
63. There are annual data for total GB council tax receipts in the Treasury's Budget and Pre-Budget Reports, but the tables generally show only two or three years data, of which at least one will be a forecast. In the national accounts tables published by ONS in *Economic Trends* and *Financial Statistics*, which have longer time series, council tax is included in 'other current taxes' and not shown separately.
64. Further data on council tax receipts published by the Scottish Executive in *Scottish Local Government Financial Statistics* show council tax income net of benefits for Scotland in total and by local authority. Data are for the most recent financial year. Similar outturn data are published by DCLG in *Local Government Financial Statistics England* for England in total (time series going back to 1998-99). Data by individual local authority (latest two years only) are available on the DCLG Local Government Finance website. Outturn data for Wales (latest outturn year only) are for the country in total are published in a Welsh Assembly Government statistical release on council tax collection rates and in *Welsh Local Government Financial Statistics*. Outturns by individual local authority, going back 10 years, are available on the Welsh Assembly Government's StatsWales website.

65. There is a substantial amount of information available on the tax base for council tax. Many of these data relate to the current and/or most recent outturn year only, though in a few cases there are time series of up to ten years. DCLG for England and the respective devolved administrations for Scotland and Wales each publishes on their website analyses of numbers of chargeable dwellings by council tax rate and individual local authority; DCLG also publishes the same analysis grouped by region. All three countries publish information on council tax rates in the form of the Band D rate for each local authority, together with national (also regional in England) averages. For England and Wales, data are also provided in the form of average rates per dwelling, for individual local authorities and for (English) regions. For Scotland, the equivalent information is provided in the form of numbers of Band D equivalents, by local authority. All three countries publish annual statistical releases on collection rates for council tax.
66. Full details of the current council tax valuation lists for England and Wales, showing the tax band for each individual property, are available on the Valuation Office Agency website, which includes a facility to find the council tax band of any property. The equivalent information for Scotland is available on the Scottish Assessors Association, together with a similar facility to find the tax band of any property.

## Micro data sets underlying published statistics

67. The Survey of Personal Incomes (SPI) is an annual sample survey taken by HMRC from its own administrative records of income tax collected. HMRC uses the SPI data set to derive a variety of published statistics on personal incomes and income distribution, and analyses of the distribution of tax liabilities and taxpayers (see paragraphs 18-20 above).
68. HMRC releases a public use version of each year's SPI on the UK Data Archive at the University of Essex. The main difference between the public use version of the SPI and the original data set used by HMRC for their published statistics and analyses is that the data in the public use version have been deliberately anonymised so that it is not possible to identify any individual from the information held in the data set. Techniques used to anonymise data include combining some individual records (eg for high earners).
69. Data codes for individual records on the public use SPI include income group, industry group of employer, and geographic location (Government Office Region).
70. Currently there are no other publicly available micro data sets.

# Annex B

## Legal restrictions on dissemination of tax data records by HMRC

### Introduction

This Annex sets out the legal position regarding the dissemination of tax data records held by HM Revenue and Customs to third parties, including other government departments. It takes the form of replies provided by HMRC to six questions from the Statistics Commission.

### Legal restrictions on dissemination of tax records data

**Q1. What are the operative legal restrictions which apply to HMRC in relation to the dissemination of individual tax data records to third parties? Which Acts matter?**

Section 18 Commissioners for Revenue and Customs Act 2005 (CRCA) provides a statutory bar on the disclosure of any information received or held by HMRC for any of its functions.

However we do provide other Government Departments with disclosive taxpayer data where a strong business case has been made and primary legislation enacted.

Where a disclosure of information is made, it must be made in accordance with the conditions set out within this section.

This excludes for example researchers working for HMRC under contract. These people sign HMRC conditions of employment and thus for the purposes of the work fully covered by HMRC disclosure policies.

The Statistics and Registration Service Act 2007 has clauses which explicitly cover:

- 1 The production of anonymised and non-disclosive information; and
- 2 An enabling clause which allows the sharing of disclosive information via secondary legislation.

**Q2. What exactly do the Acts in question prevent HMRC from doing in respect of passing on tax data records to third parties? Is any distinction made between making data records publicly available, and (a) supplying data to other government departments for (pre-defined) statistical purposes?, (b) making data available for research purposes under controlled access conditions? (ie a data lab on the ONS model)**

Section 18 of the CRCA firstly establishes HMRC's obligation not to disclose any of its information to any person outside of the department, thereafter the Act prescribes certain exemptions to that overriding principle. A disclosure of information may therefore only be made with the lawful authority of Sub-section 18(2) and 18(3) and Section 20 of the Act. For example we may disclose information with the taxpayer's consent, by order of the court or by an enactment of parliament. The CRCA makes no distinction between the nature of the information being sought (i.e. anonymous or aggregate data), the purpose to which the information will be put or the third party making the request, but in practice:

For non-disclosive data this is provided under the scope of National Statistics. The current practice is reinforced by clauses in the Statistics and Registration Act 2007.

For potentially disclosive data the disclosures are made under 'statutory gateway' for other Government Departments or by HMRC contracts for researchers as covered in 1 above.

**Q3. Does the Data Protection Act place any further restrictions on HMRC's ability to pass on tax data to third parties, or are the restrictions it imposes already effectively covered by existing tax legislation?**

HMRC is a 'data controller' for DPA purposes, but the CRCA provides a level of protection greater than that afforded by the DPA alone. DPA remains a consideration though because even if a disclosure is lawful under CRCA, consideration must still be given to the disclosure's DPA and HRA compliance.

**Q4. Are the legal restrictions that govern the passing on of tax data for companies any different from those that govern tax data for individuals?**

The restrictions on the sharing of information are the same for all HMRC customers and all of our functions: the Act makes no distinction between companies and individuals.

In practice it is sometimes possible to make data on individuals non-disclosive eg the SPI public use tape, but for corporate data this is considerably more difficult and we have not done this.

## Passing data to other government departments for statistical purposes

**Q5. What kind of legal restrictions apply to HMRC passing on tax data to other government departments, where the data are to be used strictly for statistical purposes (eg compilation of other statistics)?**

The same legal restrictions as noted above will always apply to information supplied to other government departments for statistical purposes, that is, information may only be shared in accordance with Section 18 CRCA.

## Legal position of an HMRC data lab

**Q6. What is the legal position in regard to an HMRC data lab? Would HMRC need to make any adjustments to raw data records before they could be made available for research purposes under data lab conditions? What kind of restrictions would need to be applied to dissemination of research results?**

The legal position for information within a data lab would be treated in the same way as any other HMRC information. Those wishing to access those records would need to be working on behalf of HMRC and so have signed a declaration of confidentiality (under S.3 CRCA). Alternatively, if the purpose of the research provides a tangible and direct benefit to HMRC, then the disclosure of our information can be made for our functions under S.18(2)(a), providing suitable safeguards are in place. Researchers who obtain direct access to information by working on behalf of HMRC would only be able to use the information for previously agreed purposes and would not be able to make further use or onwardly disclose any information without HMRC prior consent.

In principle the data within the datalab would be potentially disclosive, but in practice datalabs usually have names and sometimes identifiers removed.

The key principle of datalabs are that the researchers have access to the potentially disclosive data, but results taken away must be non-disclosive and not linkable to other material to make them potentially disclosive. The researchers are still covered by the confidentiality rules after they leave the datalab and so would be subject to the same sanctions as HMRC staff.





