HM Revenue & Customs

Sharing and publishing data for public benefit

Consultation document
Publication date: 17 July 2013
Closing date for comments: 24 Sept 2013

Subject of this consultation:	Proposals for data sharing and publication by HM Revenue & Customs (HMRC) to deliver wider public benefits.
Scope of this consultation:	This consultation invites views on proposals for HMRC to increase the scope for the department to share non-identifying information (that is, information that does not relate to identifiable individuals or legal entities), and on proposed safeguards. Options cover general and aggregate data as well as anonymised data sets.
HMRC also seeks views on the potential benefits, costs / risks and necessary safeguards for proposals to share VAT registration data, either publicly or under controlled conditions for specified purposes, for example, credit rating. Views are welcome on the principles underlying these proposals and on the suggested approaches to implementation.
Who should read this:	We would like to hear views from those with an interest in public data, including civil society representatives, trade or professional bodies and businesses – in particular VAT traders and individuals.
Duration:	The consultation period runs from 17 July 2013 to 24 September 2013.
Lead officials:	Cindy Bell, HM Revenue & Customs (chapter 2)
Simon Woodside, HM Revenue & Customs (chapters 3&4)
How to respond or enquire about this consultation:	Please email mailbox.datapublicbenefit@hmrc.gsi.gov.uk

Alternatively, you can send written responses to:
Data Sharing Consultation,
Room 1C/22
100 Parliament Street
London
SW1A 2BQ |
| **Additional ways to be involved:** | HMRC will be consulting key stakeholders about these proposals, including our Joint VAT Consultative Committee. If you would like to be included in a consultative meeting please let us know via the email address above. |
| **After the consultation:** | The Government will publish a response at Autumn Statement. This will include a summary of responses to this consultation and of the outcomes of our project to help assess the impact of a controlled release of VAT registration data (annex D).
Should the Government decide to take forward proposals in this area legislation will be required. The Government's programme for the current parliamentary session has already been announced. Therefore the earliest that a legislative response could be taken forward would be in 2014 -15. |
| **Getting to this stage:** | The Government published its response to the Shakespeare Review of Public Sector Information on 14 June 2013, including a commitment to this consultation. |
| **Previous engagement:** | This is the first consultation on these proposals. |

Contents

On request this document can be produced in Welsh and alternate formats including large print, audio and Braille formats

Ministerial Foreword

The data held by the public sector is among the most useful and valuable anywhere. This is why the UK Government is at the forefront in making a step change in the availability of data held by the public sector, with the potential to deliver significant public benefits.

Last month, the Government took two major steps towards realising these benefits. First, the Government published its response to the Shakespeare Review of Public Sector Information, setting out a framework for pursuing this agenda in the public sector. Secondly, the UK helped secure the G8's Open Data Charter, which presumes that the data held by Governments will be publicly available unless there is good reason to withhold it.

It is important that HMRC plays a full part. HMRC`s relationship with businesses and individuals is unique, and this is reflected in the scope and depth of the information HMRC collects, creates and protects on behalf of taxpayers.

HMRC operates under a strict legislative framework that limits data sharing. This provides all of us with the assurance that our information will be protected, essential to the effective operation of the tax system.

There are, however, potential uses of HMRC data which could generate public benefits without compromising the core principle of taxpayer confidentiality. This consultation brings forward three options:

- wider sharing of aggregated and anonymised tax data, for example, for the purposes of research or policy development;

- release of basic non-financial VAT registration data as public data; and

- sharing more detailed VAT registration data on a more restricted and controlled basis for specific purposes, such as credit referencing.

There are potentially significant benefits, including: improving policy making across government; supporting the Government's transparency and growth objectives by helping improve access to credit for business; and making it easier to protect against fraud. VAT registration data could also provide a foundation for private sector business registers. There are likely to be further positive uses which emerge only once the data is available.

But it is paramount that any data release has appropriate safeguards, essential to maintaining taxpayer confidence and protecting HMRC's reputation. None of the proposals introduced in this consultation proposes release of individual taxpayers' financial or tax payment data.

The Government will reflect on responses to this consultation when deciding whether - and, if so, how - to take forward the proposals. Should the Government do so, legislation would be required, with further opportunity for public and Parliamentary scrutiny.

This consultation explores important issues of how publicly-held data should be used and how HMRC should undertake its public functions. I welcome it and would encourage all interested parties to participate.

David Gauke MP

Exchequer Secretary to the Treasury

1. Introduction – Exploiting HMRC's data to deliver wider public benefits

Overview of this consultation

1.1 The nature and extent of HMRC`s relationships with businesses and individuals are unique, and this is reflected in the scope and depth of the information HMRC collects and creates. The data sets HMRC holds are among the largest and richest held by government.

1.2 The Government wishes to explore whether and how the use of this information can be optimised to achieve benefits above and beyond the vital but specific business requirements of HMRC. HMRC is keen to explore new opportunities to support the Government's Open Data policy[1] and work more effectively with other public sector bodies to improve transparency and promote economic growth.

1.3 This consultation seeks feedback on areas of possible data release that could make a real difference in HMRC's ability to contribute to wider public goals, while safeguarding the core principle of taxpayer confidentiality. The proposals reflect the Government's strategic view, which identifies a spectrum of ways in which public sector data will be treated according to its sensitivity.

1.4 **This chapter** provides an overview of the context for this consultation, including the broad context of 'open data' and HMRC's current position.

1.5 **Chapter 2** sets out proposals and consultation questions about giving HMRC greater flexibility to share and publish data that does not identify individuals or businesses. It considers the benefits, risks and what safeguards may be required. It covers general and aggregate information, then looks separately at anonymised individual level information.

1.6 **Chapters 3 and 4** set out proposals and consultation questions about sharing certain VAT registration data, including potential benefits, risks and safeguards. chapter 3 looks at the possible public release of a limited set of data, and variations on this proposal, while chapter 4 considers sharing a wider data set with specific parties (e.g. with credit reference agencies), under strict conditions for specified purposes.

1.7 **Chapter 5** lists the full set of consultation questions for ease of reference, and **chapter 6** sets out the consultation process.

---

[1] Open data white paper: Unleashing the Potential – Improving the transparency and accountability of government and its services.
Open Data White Paper: Unleashing the Potential - Publications - Inside Government - GOV.UK

Open data – recent developments

> *Data allows us to adapt and improve public services and businesses and enhance our whole way of life, bringing economic growth, wide-ranging social benefits and improvements in how government works.*
>
> *The richest source of data is government, which accounts for the largest proportion of organised human activity…*
>
> Shakespeare Review pp.5&7[2]

1.8 In 2012 the Government invited Stephan Shakespeare, Chair of the Data Strategy Board, to conduct an independent review of public sector information and to explore new ways to use information held by the public sector. The report, published in May this year, identifies the need to move from a policy of transparency, with primarily a focus on accountability, to a data strategy that supports economic growth.

1.9 The Government response to the Shakespeare report[3] was published on 14 June 2013 and recognised that the report's recommendations and core strategic approach provided a good basis for it to develop a refreshed data strategy.

> *The Market Assessment Report published by Deloitte as part of the Shakespeare Review of Public Sector Information provides some case study examples of public sector information generating value, including:*
>
> - *Using NHS prescribing data to identify efficiency savings of up to £1.4 billion per annum from switching from branded to generic drugs*
>
> - *Using public sector information on road works to better co-ordinate utilities work and journey planning – leading to estimated benefits of around £25 million per annum for local authorities and road users in terms of efficiency savings and reduced congestion*

---

[2] 'Shakespeare Review – an Independent Review of Public Sector Information', May 2013
Shakespeare review of public sector information - Publications - Inside Government - GOV.UK
[3] The government's response to the Shakespeare report which explores new ways to use public sector information
Government response to Shakespeare review - Publications - Inside Government - GOV.UK

1.10 This year's UK Presidency of the G8 has embraced this theme. The G8 Open Data Charter[4] sets out principles on the use of open data, the first of which is 'open data by default'. This presumes that data will be released unless there are strong reasons not to do so, e.g. legal, commercial or security concerns. The fifth principle, 'releasing data for innovation', recognises that the greatest social and economic value can be unlocked by opening data to the widest range of people and organisations. This can only be achieved by improving the access to, and use of, data.

1.11 The Shakespeare Review identifies the need for people to feel confident that the benefits of open data can be achieved while safeguarding privacy – ensuring that data is made as useful as possible but not misused. This balance is core to our approach in this consultation, which considers broadening our objectives for opening up HMRC's data while maintaining strict controls.

1.12 This consultation also broaches the question of how any data release proposals that are taken forward should be funded and what, if any, charges should be made, or benefits sought, for use of the HMRC information that might be released. The Shakespeare Review recognises that there may be cases where greater benefits accrue to the public with an appropriate charge, and that a default position of free access to data is not necessarily the only option.

HMRC's current position

1.13 HMRC was created by the Commissioners for Revenue and Customs Act 2005 (CRCA). This legislation provides strong protection for the information that HMRC holds. HMRC officials are prohibited from sharing information except in the limited circumstances set out in legislation (see annex A for a summary of the main exceptions).

1.14 The prohibition on sharing applies to all of HMRC's information and includes non-identifying (aggregate or anonymised) information as well as information on identifiable individuals or legal entities. This is the case even where the same information is already in the public domain.

1.15 Maintaining confidence in the core principle of 'taxpayer confidentiality' and the related safeguards is essential to the effective operation of the tax system, because it supports compliance and willing cooperation. This consultation does not seek to disturb that core principle. However, the existing legislation arguably goes further than necessary by extending strong protection to non-identifying information, and to information (such as name and address) which is often widely available from other sources (although it is recognised that safeguards would be needed for certain vulnerable people).

---

[4] G8 Open Data Charter – Open Data Charter - Publications - Inside Government - GOV.UK

1.16 Within the framework of its existing legislation, HMRC already makes a lot of useful data available for broader use.

- For decades, HMRC and its predecessor departments have produced a wide range of National and Official Statistics relating to the tax system. These are kept under regular review, with cost-effective enhancements incorporated as appropriate based on consultation with users. The department also publishes performance and management information on a quarterly basis[5].

- Since 2010, HMRC has published key transparency data sets on www.data.gov.uk. These include details of departmental spending and procurement, information on HMRC's organisation and information on meetings between senior officials and external organisations.

- In May 2011, HMRC launched its Datalab. This initiative allows researchers to have access to anonymised data under strictly controlled conditions to undertake projects that contribute to the department's objectives, such as measuring and improving tax compliance.

- In June 2012, HMRC published its Open Data Strategy. This set out a further expansion in the range of data the department publishes, including an upgrade of the uktradeinfo website to improve public access to trade statistics data.

- In December 2012, as part of its Open Data initiative[6], HMRC created a Tax Transparency Sector Board, which advises on its open data work plan and serves as a focal point for requests from external organisations for publication of new data sets.

- Over the next few months, as part of the Open Data initiative to build a 'national information infrastructure', HMRC will be working with the Tax Transparency Sector Board, Cabinet Office Transparency Team and other stakeholders to identify the most important data sets it holds. It will also consider opportunities to publish more data while safeguarding taxpayer confidentiality. The outcomes of this process will feed into the department's updated Open Data Strategy in October.

What is HMRC consulting on and why now?

1.17 HMRC's existing legal framework takes as its starting point that HMRC's information should not be shared with third parties except in specified circumstances set out in legislation.

---

[5] HM Revenue & Customs:National and Official Statistics
   HM Revenue & Customs:Performance and Management Information
[6] Tax Transparency Sector Board - Policy advisory groups - Inside Government - GOV.UK

1.18 While HMRC currently shares and publishes a lot of useful information, the blanket nature of the prohibition on disclosing its information has prevented the department from contributing fully to initiatives designed to increase transparency and data sharing and to deliver better services across the public sector. It has also limited HMRC's ability to share anonymised individual level data for the purposes of research and analysis.

1.19 The case for relaxing current restrictions on sharing non-identifying information is covered in chapter 2, which sets out the potential for public benefits if the legislation can be made more flexible without compromising the protection for private or commercially sensitive information. Combining HMRC's non-identifying data with other government data sets would enhance government's capacity and capability to deliver value through improving public services and promoting economic growth. This would be to the benefit of citizens and businesses.

1.20 There are also some specific HMRC data sets that, while they could be classed as identifying information, could potentially be shared more widely without causing material harm and with the delivery of significant public benefits as a result. The VAT registration data set (excluding financial data) may be a good example of this, particularly given the fact that much of the data is already available from other sources. Chapters 3 and 4 set out proposals for release of some VAT registration data with appropriate controls and safeguards.

1.21 Together, these proposals form a significant development in HMRC`s Open Data Strategy[7], which sets out a strong commitment on transparency and the exploration of new opportunities to optimise the use of HMRC information. The key issue is how to maximise benefits while safeguarding confidentiality and security. This is why the consultation covers the principles of these proposals, as well as their possible implementation.

1.22 Primary legislation would be required to bring these proposals into effect. The Government will consider the responses to this consultation and, if a decision is made to proceed, an announcement will be made in the autumn, setting out the measures, process and timetable. The legislative process would enable a further period of consultation as well as Parliamentary scrutiny.

Longer Term

1.23 HMRC will also use the feedback provided during the consultation to inform its consideration of other areas where there may be scope to exploit its data more effectively to support key departmental and wider government objectives while safeguarding taxpayer confidentiality. Any specific proposals developed as a result of these considerations that would require legislative change will be published for consultation in due course.

---

[7] HMRC Open Data Strategy - HMRC Open Data Strategy | data.gov.uk

2. Sharing non-identifying information

What types of non-identifying information does HMRC have?

2.1   In broad terms, HMRC currently has:

- **general information**. Much of the information that HMRC holds is general information on, for example, our policies and processes;
- **aggregate information**. This is created by combining information on individuals and legal entities into groups. One example is average tax paid by individuals within a certain income range; and
- **anonymised individual level information**. This is created by removing any information (e.g. name or address) that could lead to direct identification of an individual or legal entity. Additional safeguards may be needed to prevent identities being deduced from anonymised information (see also 2.19 – 2.21)

What are the benefits of HMRC being able to share non-identifying information more widely?

2.2   As set out in chapter 1, HMRC collects and holds a significant amount of data as a result of its unique relationship with individuals and businesses. Officials are forbidden by law from sharing *any* of this information or data except in the limited circumstances set out in legislation. The main exceptions to this general prohibition are where information is disclosed for the purpose of the department's own functions, through legislation that expressly allows the disclosure, or when consent is given. There is an additional protection for identifying information in that unlawful disclosure of such information is a criminal offence.

2.3   There is a case for arguing that the current disclosure restrictions afford more protection to non-identifying information than is needed. This type of information could be readily and widely shared to help deliver more effective and efficient public services without compromising customers' confidentiality. Currently, specific legislative gateways are considered and created on a case-by-case basis to enable HMRC to share non-identifying information with third parties to generate wider public benefits. This is a time- and resource-intensive process.

2.4   For example, the Department of Health needs up-to-date and accurate information on doctors' and dentists' incomes in order to inform a wide range of policy decisions. Undertaking a survey of doctors and dentists to collect this information would be very expensive, and would be subject to sampling error, measurement error and bias. HMRC already holds incomes data from tax returns of doctors and dentists or their employers, which it can aggregate in different ways to ensure that identities cannot be deduced. So better quality information is available at far lower cost from HMRC. However, in order for HMRC to make this aggregated data available to the NHS, it was necessary to create a legal gateway. This took around 18 months to put in place from start to finish.

2.5 Having a general block on sharing this type of information also sits oddly in the context of Freedom of Information (FoI) obligations. As things stand, HMRC could be legally required to provide non-identifying information to a citizen in response to an FoI request, which it would be unlawful to provide in response to a request from, say, another government department. The general block on disclosure also runs counter to the Government's presumption that data should be 'open by default'.

2.6 The current legal restrictions have prevented HMRC from contributing fully to a range of policies and research areas, in circumstances where there has been no risk of disclosing anyone's identity. These areas include increasing transparency, promoting UK economic growth, increasing social mobility and promoting health. Two specific examples are:

- an approach was made to HMRC to supply anonymised data to help the Chief Medical Officer for Wales to carry out research into the factors underlying excess winter mortality. Considerable work was undertaken to identify whether and how the information could be disclosed. Obtaining customers' consent to disclose anonymised data was not practical, there was no link to HMRC's functions and legislating a specific gateway would have been too time consuming and would, in any case, have been too late to inform the research. The conclusion was therefore reached that it would not be possible for HMRC to supply the requested information; and

- HMRC was approached by the Department for Education (DfE), to provide information supporting the national evaluation of the Special Educational Needs and Disability (SEND) Pathfinder programme. DfE were seeking to bring together some of these services to reduce the complexities for families who were taking part in this programme. Again, HMRC was unable to supply the information requested in time.

2.7 Even if a legal gateway had been created in each of these examples, it would only have addressed the specific issue being considered. In contrast, relaxing the current restrictions on sharing non-identifying information in more general terms would enable HMRC to share this type of information quickly and easily while maintaining appropriate protection for customers' sensitive and personal information. This would enable HMRC to deliver its own and wider government objectives more effectively and contribute to a more efficient and joined up use of data across government.

Proposals for sharing non-identifying information

2.8 The proposals are to relax the current legal constraints on disclosing non-identifying information and specifically to allow HMRC to:

A) share general and aggregate information more widely; in respect of both (i) making non-identifying information generally available by publishing and (ii) sharing non-identifying information with specific third parties to deliver public benefits wider than HMRC functions (see example at para 2.4); and

B) improve access to anonymised data sets for research or statistical purposes where there are no direct benefits to HMRC's functions but there are wider public benefits.

2.9 This would allow more effective use of the department's data to help deliver better public services and promote economic growth.

A: Removing legal constraints on sharing general and aggregate information

2.10 HMRC already publishes a wide range of general and aggregate information on its priorities, policies, procedures, performance and activities. The department is able to do this within its current legal framework as part of its functions and to fulfill its obligations under the Freedom of Information Act to maintain a publication scheme.

2.11 HMRC also, however, holds an extensive amount of general and aggregate information that, if shared more widely, would contribute to the more efficient and effective delivery of services and benefits beyond HMRC's immediate functions. In addition, this information could be exploited more effectively for the general benefit of UK plc. This would be consistent with the Government's presumption that data should be 'open by default', the recommendations of the Shakespeare Review of Public Sector Information[8] and the Administrative Data Taskforce report[9] on improving access to government data for research purposes.

---

[8] 'Shakespeare Review – an Independent Review of Public Sector Information', May 2013
Shakespeare review of public sector information - Publications - Inside Government - GOV.UK
[9] Administrative Data Taskforce report: government response - Publications - Inside Government - GOV.UK

Risks and safeguards

2.12 If HMRC were given the power to disclose general and aggregate information to third parties on a wider basis than at present, this would remain subject to the same robust principles and processes currently applied to HMRC Official and National Statistics.

- There is a small element of risk associated with publishing aggregate data, for instance by linking to other data sources, or simply through local knowledge of a small local area, especially if one taxpayer makes up a large proportion of economic activity in that area. These risks can be addressed through statistical disclosure control policies.

- HMRC's current policy on statistical disclosure has been developed and quality assured by methodology experts from the Office for National Statistics and the techniques and parameters the department uses are recognised as best practice.

- Legislation would be 'permissive', meaning that HMRC would be able, but not obliged, to share information in each instance. This would enable case by case assessment of the risks and benefits, and on how to disclose e.g. the appropriate format, again in line with the policy on statistical disclosure.

- Finally, the criminal sanction would remain a strong safeguard against unlawful disclosure of identifying information.

2.13 In addition, HMRC would continue to ensure that in taking disclosure decisions, appropriate consideration would be given to the impact on HMRC's resources and the delivery of its core business objectives alongside the total package of benefits.

How would this work in practice?

2.14 The ways that HMRC would use this power in practice would include, for example:

- sharing with other government departments where HMRC currently does not have legal gateways to do so; and

- supplementing HMRC's continuing commitment to comply with its obligations under the FOI Act with more flexible options that would allow HMRC to focus additional resource on disclosures where there is the greatest public benefit.

Consultation Questions

Q1 Do you agree that the legal constraints on sharing general and aggregate information should be removed, on a permissive basis, in respect of:

    (i)   **HMRC making information generally available through publishing?**
    (ii)   **HMRC sharing information with specific third parties to deliver public benefits wider than HMRC's functions?**

Please give reasons for your answers.

Q2 Do you agree with the proposed safeguards on the proposal to share general and aggregate data? Should any further controls be considered on what can be shared, with whom or how?

B: Relaxing HMRC's legal constraints on sharing anonymised individual-level data with accredited third parties for the purposes of research and analysis
.

2.15 Extending the relaxation of legal constraints to include anonymised information, with the additional safeguard of controlled secure conditions, has significant potential to deliver wide public benefits through research and other use.

2.16 Currently, research projects that require access to information held by HMRC must demonstrate a benefit to HMRC's functions, in order for data to be shared. This provides a valid legal basis ('gateway') that allows disclosure of information so that research can go ahead – e.g. with academics or social research companies. But projects that would deliver wider public benefit, but not benefit HMRC directly, must often be turned down because HMRC is unable legally to provide researchers with access to the necessary anonymised data sets.

2.17 Relaxing the current legal constraints to enable HMRC to disclose anonymised individual level information for the purpose of general research and analysis would mean that HMRC would be able to share such data where the benefit was for wider public objectives, rather than the current need for HMRC to benefit directly.

2.18 This would enable HMRC to support important projects similar to those summarised in paragraph 2.6 without needing to create specific legislative gateways that provide access to the relevant anonymised data sets. Annex B provides further detail on the benefits of sharing anonymised tax data for non-tax purposes.

Additional Risks and safeguards

2.19 HMRC recognises that anonymised individual level data presents a slightly greater risk of a customer's identity being deduced than aggregate data. For instance, even where names, trade and address are removed from a dataset, certain prominent individuals or businesses might still be identifiable within a broader industrial or regional classification, especially where HMRC data can be combined with other information in the public domain or privately known about those customers. For this reason, further safeguards are required to reduce the risk of indirect disclosure. HMRC already makes anonymised customer data available for research purposes that benefits HMRC's functions, in the safe and secure environment of its 'Datalab'. Here, accredited researchers are allowed access to anonymised data for research and statistical purposes. Access takes place under strictly controlled conditions and with tight governance.

2.20 Before allowing access, HMRC requires researchers to undertake a rigorous accreditation process. Statistical disclosure controls are carried out on any research outputs before they leave the Datalab. In addition, researchers are subject to the same confidentiality provisions as HMRC staff, including the criminal sanction in the case of unlawful disclosure of identifying information. The principles under which the Datalab operates are consistent with the recommendations in the recent Administrative Data Taskforce report for safe sharing of anonymised data for research and statistical purposes.

2.21 The proposal to relax legal constraints on the sharing of anonymised data with accredited researchers for the purpose of research and statistical purposes would not compromise the current best practice safeguards in place, but it would open up the potential for research using HMRC data that does not benefit HMRC directly but is in the wider public interest (for further details see annex B). Naturally, HMRC would want to ensure in considering any data request that its business and priorities would not be compromised, and this would need to be balanced with the overall benefits that the project was intended to generate.

Consultation Questions

Q3 **Do you agree that HMRC should be able to share anonymised individual-level data for the purposes of research and analysis to deliver public benefits wider than HMRC's own functions? Please give reasons for your answer.**

Q4 **Do you agree with the proposed safeguards on the proposal to share anonymised individual level data? Should any further controls be considered on what can be shared, with whom or how?**

Funding the use of HMRC data

2.22 HMRC currently requires other government departments to meet the costs of generation, collection, retention and dissemination of HMRC information that they request for their own purposes. This ensures that costs fall on the user and that HMRC's core business is not compromised. This is in line with Government Accounting requirements. HMRC does not propose changing this practice where its data is requested by other Government departments.

2.23 Sharing some of HMRC's information with accredited researchers to generate public benefits might also secure benefits for HMRC in undertaking its core functions of managing and collecting taxes. The controlled data release proposal in chapter 4 is an example of this: HMRC would provide VAT registration data to credit reference agencies to help improve access to credit; the credit reference agencies might be able to enrich or supplement the VAT registration data for HMRC to use in compliance risking and enforcement activities.

2.24 One of the principles of Open Data is a presumption that public data should be freely available, in the absence of good reasons to the contrary. But the Shakespeare Review, for instance, also recognises that there are circumstances where charging for information is appropriate. HMRC is considering how these principles should apply to its own data strategy. For the purpose of this consultation the immediate consideration is how these particular data-sharing proposals should be funded, and what, if any, charges should be made, or benefits sought, for use of the HMRC information which might be released.

Consultation Question

Q5 How should the generation and release of anonymised or aggregated data be funded? Please give reasons for your answer.

Implementation

2.25 As set out in the introduction, relaxing the current constraints on disclosing non-identifying information would require a change in legislation. Any draft legislation would be published for further public consultation and Parliamentary scrutiny.

2.26 Presently, no equality impact is anticipated as a result of these proposals. Any impacts that are identified as a result of this consultation will be set out as part of the next round of consultation where they would form part of any proposed change to existing legislation.

2.27 On a practical level, it is likely that with the removal of legal restrictions on sharing non-identifying data, there would be an increase in requests for HMRC data. Providing HMRC with the discretion to consider and prioritise these requests is key to managing any resulting impacts. In dealing with data requests from any third party, there will often be trade-offs between timeliness, completeness and quality of the data. Wherever possible, HMRC would seek to discuss what options were available and come to a pragmatic agreement depending on user needs.

3. Publication of VAT registration data

Why consider disclosure of VAT registration data?

3.1 While the proposals in chapter 2 deal with release of aggregate and anonymised data, the proposals in this chapter deal with the release of data which identify a group of business taxpayers (without disclosing any financial or tax payment data).

3.2 An underlying principle in developing the proposals in this chapter is brought out in the Shakespeare Review[10]. Data belong to citizens and the presumption of government should be towards openness, unless this causes harm. It is not for government to dictate the nature of the opportunity. The corollary is that the Government will not always be aware of the range or scale of potential benefits, as the quotation below shows – this consultation will help to establish these.

The benefits will be many including: transparency, accountability, improved efficiency, increased data quality, creation of social value, increased participation, increased economic value, improved communication, open innovation, and data linkage.

Shakespeare Review, page 7.

3.3 There has been public interest for some time, for example from credit reference agencies (CRAs), in the publication of VAT registration data as a resource to generate benefits. Key elements of the VAT registration data set are already and deliberately placed in the public domain, for example in business directories or on invoices, but are not easily or systematically available as a full set. There are a range of possible uses, including in combination with other publicly available data, such as electoral or company records. The proposals below consider how best to unlock this potential, whilst ensuring appropriate safeguards are in place.

What are VAT registration data?

3.4 VAT registration data are the identifying data provided by traders on registration and kept updated on an ongoing basis. The full data set in scope for these proposals is set out at annex C, and is grouped in the following categories:

- VAT registration number (VRN);
- names of the business and/or its proprietors;
- contact details (eg, address, telephone, email, website); and
- status (eg, legal structure, industry classification, details / dates of registration, incorporation or insolvency).

---

[10] 'Shakespeare Review – an Independent Review of Public Sector Information', May 2013
https://www.gov.uk/government/publications/shakespeare-review-of-public-sector-information

3.5 Very little financial data is provided in the context of VAT registration, but where it is, e.g. estimated turnover, this is excluded from the scope of these proposals. Other financial data about businesses or their tax affairs, e.g. about VAT payments or receipts, are also excluded.

3.6 The VAT registration data set provides a comprehensive picture of UK businesses above the turnover threshold (currently £79,000 a year) for mandatory registration, as well as those businesses with turnover below the threshold which have chosen to register voluntarily because it is in their interests to do so. At the start of 2012, there were some 1.9 million live registrations on the system.

3.7 Of all VAT registered businesses, around 800,000 (varying over time) are not incorporated as companies, and there is a particular gap in the public record for information on these businesses, which release of VAT registration data would help to fill.

3.8 The VAT registered population is very diverse, ranging from the largest FTSE-listed businesses, to home-based sole traders. It does not include approximately 2.9 million non-VAT registered businesses, among them the smallest companies, self-employed sole traders and partnerships. HMRC holds data on these businesses through the Self Assessment system, but this class of businesses is not included in the disclosure proposals for consultation, because this would raise additional privacy issues and the benefits are less clear.

Why might release be valuable?

3.9 Disclosure of all or part of the VAT registration data set has the potential to benefit:
- those who provide services which seek to identify, verify or assist businesses, e.g. by improving the accuracy of such services;
- businesses and individuals acting for example as traders, customers, suppliers, lenders or employees, e.g. by making it easier to get independent, accurate verification about aspects of a business such as its credit rating; and
- the economy as a whole, by improving the flow of credit and trade.

3.10 Commercial business information providers, such as credit reference agencies, are less effective in the absence of reliable information, of which there is relatively little about unincorporated and newly incorporated businesses. Depending on which data items were disclosed and to whom, there could be improvements in credit scoring (and thus in availability and pricing of credit), in the effectiveness of checks to combat commercial fraud, and in the ease of complying with anti-money laundering requirements.

3.11 Release of a VAT registration data set might also provide the foundation for private sector development of a UK business register or multiple registers. A number of countries (including France, Germany, and Spain) have instituted national business registers with differing objectives, levels of coverage and cost structures. As an existing resource, VAT registration data might enable some of the benefits of a stand-alone business register at lower cost.

3.12 These are only examples of benefits which can be identified at this stage. There are likely to be further examples which are not immediately apparent. As the Shakespeare Review makes clear, information disclosure can generate benefits in unforeseen ways.

Proposals for public release of VAT registration data

3.13 There is a range of options for VAT registration data release. This runs along a spectrum in terms of how much data might be released, to whom (e.g. to the public or to selected parties), and under what conditions (e.g. with or without the consent of individuals). The Government wishes to find the option or options which can deliver the most benefits, whilst ensuring that appropriate safeguards are in place.

3.14 It is already theoretically possible to release VAT registration data with the consent of those affected, without legislation. HMRC is not putting this forward as an option, however, because it would be expected to generate fewer benefits (by offering an incomplete base data set), and would cost more to implement, including imposing an additional administrative burden on VAT registered traders.

3.15 Whatever option or options might be chosen, primary legislation would therefore be required for implementation, providing the assurance of full public and Parliamentary scrutiny.

Option A – Core public release proposal

3.16 The core proposal for public release by HMRC is a very selective extract covering just three data fields - VAT registration number (VRN), trading name, and Standard Industry Code (SIC) classification number. This would be made available to the public digitally, and regularly updated.

3.17 This would deliver a comprehensive core data set, in the sense that it would provide full coverage of the registered population. It would not of course be as useful as provision of the full data set, but as a reliable set of the existing registered population could be linked with other data, serving as a foundation for other uses. It might thereby realise a significant part of the potential benefit, while minimising any risks.

3.18 With regard to the impact of publication on privacy, much registration data (particularly but not limited to incorporated businesses) is already public. In addition to Companies House filings, most businesses seek to publicise their services and contact details, for instance through advertising or directory listings, and include their VAT number on invoices or similar documentation.

3.19 Some traders operate in a more private space, with details of their services known only to a handful of clients. Some businesses operate from the proprietor's home and use their own name as the trading name. These proprietors may be sensitive to disclosure of personal contact details or even the existence of their business.

3.20 It could be argued that identifying data supplied to HMRC under current rules which preclude its publication should remain confidential. On the other hand, it can also be argued that business activity takes place in a more public arena than private life. A principled case can be made (and indeed underpins the many national business registers in other countries) that businesses offering services should be readily identifiable, and that the public has a legitimate interest in this data.

3.21 The greater the number of data fields released, the greater the potential impact on privacy (this is discussed in more detail under Option C below). In considering how much data might be released, the Government might well wish to err on the side of caution in minimising potential risks to privacy. Restricting publication to the core proposal provides a credible minimum disclosure, which could still generate a significant proportion of the potential benefits.

Ensuring appropriate safeguards for release

3.22 At the same time, any risks arising from publication of the limited VAT registration data under the core proposal need to be identified and appropriately managed.

3.23 It is important that publication does not increase the potential for VAT fraud against HMRC or for fraud against VAT registered traders, for example by impersonation or identify theft.

3.24 Publication of the core proposal data would provide an additional source of identifying data (for example trader names and VRNs) which are used in both of these types of frauds. However, this data is already widely, if variably, available through sources such as invoices and the internet.

3.25 In addition, some businesses, particularly smaller, non-incorporated businesses without a current public presence (e.g. website or directory listing), would become more visible to potential fraudsters. However, it is not clear that publication would necessarily make fraud easier or increase its incidence, as it would not increase the availability of the private information required to perpetrate a successful fraud. Significantly more private information, including previous VAT returns and personal identification numbers, is required to perpetrate VAT fraud or to attack HMRC systems. The success of fraud against traders is determined more by other factors, such as plausibility or system weakness, than by the availability of trader name and VRN.

3.26 More detailed risk assessment will be carried out if these proposals are taken forward.

Further options for public release of VAT registration data

3.27 There are two variants for modifying the core proposal, which need to be considered.

Option B – Exceptions to or opting-out of the core proposal

3.28 It would be possible, within any legislation implementing the core proposal above, to provide for exceptions from publication. Relevant categories of businesses could be identified, or an opt-out option could be provided, which could be exercised by traders wishing to be excluded from a public disclosure. The opt-out could be framed as a free choice, or permitted if the trader wishing to opt out met specific criteria (e.g. demonstrating a risk of harm arising from identification).

3.29 Permitting exceptions to, or opt outs from, publication could provide greater scope to protect privacy (although privacy is substantially protected by setting the core proposal at the minimum end of the disclosure range).

3.30 On the other hand, exceptions or opt-outs would potentially reduce the value of the published data set, as users could no longer rely on its being fully comprehensive. The loss of value is not easy to estimate, but is likely to depend in part on the numbers excluded from publication, probably greatest in the case of a free choice opt-out.

3.31 In addition, permitting exceptions to, or opt outs from, publication would substantially increase the administrative costs of gathering and confirming exclusions, and of managing and publishing the right selection of data.

Option C – Volunteering additional disclosures

3.32 It would be possible to give businesses an option to volunteer to disclose additional VAT registration data fields beyond the core proposal. This could cover all or some of the remaining data fields.

3.33 There is no simple answer to the question of what incremental benefits this might generate over the core proposal. These data fields are likely to have different values to different users. The incremental administrative costs of applying this option would also need to be taken into consideration.

3.34 The most valuable data field (after the core proposal data fields) could be the trader's address (i.e. principal place of business) and contact details, which users could seek to match for example with Electoral Register, Companies House and individual business data to improve accuracy. Legal entity status, and registration/deregistration, incorporation and insolvency details could be cross matched to company and court records. Historical records showing changes in data fields (e.g. principal place of business) may provide evidence of continuity and/or correlate with other relevant business characteristics.

3.35 Equally, offering a choice would allow traders to make the level of additional disclosure which best fitted with their own desires for visibility or privacy, and their perception and valuation of the benefits of these alternatives.

Consultation questions

Q6 **What potential uses and benefits do you think would arise from publishing VAT registration data?**

Q7 **Do you believe that the proposed safeguards are sufficient to mitigate any risks arising from publication? What, if any, additional impacts may arise and what further safeguards should be considered?**

Q8 **Do you agree with the core proposal (Option A) to publish the VAT registration number, trading name and industry classification? Please give reasons for your answer.**

Q9 **What are your views on the proposal to allow exceptions or opt-outs to publication (Option B)?**

Q10 **What are your views on the proposal to publish additional data fields on a voluntary basis (Option C)?**

4. Controlled release of VAT registration data

Why consider a controlled release of VAT registration data?

4.1 While chapter 3 proposes a limited publication of VAT registration data, maximising the economic benefits depends on access to the full data set at annex C. Controlled release of the full data set, i.e. release to specific parties for limited uses, could have greater benefits – for example release to credit reference agencies (CRAs) could have an impact on economic growth by materially increased access to credit, especially for non-incorporated SMEs, as a result of improved credit scoring. A controlled release of the full data set could potentially realise these benefits without incurring the risks that might arise from general publication.

4.2 A controlled release would be consistent with the framework set out in the Shakespeare Review, which acknowledges 'a spectrum of uses and degrees of openness', including sharing with 'approved legitimate parties whose use can be tracked and against whom penalties for misuse can be applied' (Shakespeare Review p.8-9).

Proposal for controlled release

4.3 A controlled release could cover the full set of identifying (but non-financial) data provided by VAT-registered businesses listed at annex C. This would be updated regularly, increasing the richness of the data over time.

4.4 The data would only be available within a strict framework:

- to specific, qualifying parties;
- for nominated purposes only (e.g. credit scoring, anti-fraud checking and identity checking for anti-money laundering purposes); and
- subject to rigorous controls, including data security and privacy standards, with penalties for misuse.

4.5 Such a controlled release moves away from the equal access offered by open publication, providing privileged access for any qualifying parties. It is therefore legitimate to ask whether HMRC should seek to generate additional benefits for taxpayers from any controlled data release. This could take the form of charging for access to the data, or requiring qualifying users to provide data in return (e.g. an enrichment or supplement to VAT registration data) which would assist HMRC in its objectives of the collection and management of taxes. This is discussed further in the following section on potential benefits of controlled release.

Potential benefits of controlled release

4.6 Controlled release of the full VAT registration data set (at annex C) could deliver substantial benefits. This full data set, including a historical record of changes, would provide a much richer picture of a business than the core proposal discussed in chapter 3, and one which would produce more value when cross-referenced to existing databases. It would help business information providers such as CRAs verify identity and status more accurately, and make more reliable judgments about a business and its proprietors, location, activities and age. While use by CRAs has already been identified as an option, controlled release could potentially be justified to any parties who could make a compelling benefits case for a proposed use, and could credibly demonstrate their ability and willingness to abide by the relevant controls.

4.7 If used by CRAs, this information could well have a significant effect on credit scores and thus on the availability and pricing of credit. This effect is likely to be most pronounced for unincorporated or newly incorporated businesses which have no credit footprint, and thus are potentially denied access to trade credit. The VAT registration data set could be sufficient to unlock access to trade credit for this group, opening up new opportunities and reducing dependence on more costly sources of finance such as credit cards.

4.8 It is important to understand any potential negative impact in benefit distribution. Availability of additional data usually has a positive impact on credit scores for individuals and businesses. Where there is an apparently negative (downward) impact, this is likely to reflect a more accurate understanding of creditworthiness, ensuring that credit is not offered inappropriately.

4.9 It is possible in theory that higher credit scores for some businesses could reallocate credit away from those whose scores remain static – for example, benefitting VAT-registered businesses over others. In practice, however, an increase in the volume of credit is a more likely outcome than reallocation. Any negative impact identified for businesses, such as any reduction in credit availability or pricing, will be taken into consideration in assessing the benefits of the proposal.

4.10 HMRC would also wish to understand if the distribution of potential benefits might impact business behaviour. For example, benefits such as improved credit scores might encourage more firms to voluntarily register for VAT. While this would be likely to increase VAT receipts, it would also increase the costs of administrating VAT.

4.11 Other potential benefits to businesses, over and above credit improvements, from better verification of identity and status, include improvements in checks to combat commercial fraud. This would make it easier to frustrate fraudsters' attempts to impersonate legitimate businesses, for example fraudulent purchase of goods on account. In addition, faster and more reliable business (e.g. client) identification could make it easier for firms to comply with anti-money laundering obligations.

4.12 As mentioned at paragraph 4.5 above, a further potential benefit from controlled data release could be assistance to HMRC in undertaking its function of the collection and management of taxes. Potential qualifying parties such as CRAs may be able to enrich or supplement VAT registration data from their own data sets. Enriched or supplemented data could be useful to HMRC in better understanding and segmenting its business customer base and in enhancing compliance risking and enforcement activity. This could increase tax revenues by, for example, reducing VAT fraud and improving compliance more generally.

Research project to assess benefits

4.13 Given the importance of the benefits case to any decision on the controlled data release proposal, HMRC will be undertaking, over the summer, a research project with a group of CRAs to help assess potential external uses of VAT registration data, possible benefits, and issues which arise from such uses.

4.14 In undertaking the research project and handling HMRC's VAT registration data, the CRAs will be bound by the same strict statutory privacy and data protection rules which govern HMRC's activities and which prevent any misuse or unauthorised release. The VAT registration data provided will be held securely, only used for the purposes of the research project, and destroyed when the project is finished. The results will inform Ministerial consideration of the outcome of this consultation and a summary of the research outcomes will be published alongside the Government's summary of responses to this consultation in the autumn.

4.15 Annex D provides more detail on the project.

Ensuring appropriate safeguards for controlled release

4.16 There would need to be appropriate safeguards for a controlled data release, as summarised at paragraph 4.4. The mechanism for implementation would be a legislative gateway, as well as contractual arrangements with any qualifying parties, which together would specify the terms under which data was transferred, held, and used and the necessary security and privacy requirements. A gateway would require primary legislation, permitting public and Parliamentary scrutiny of proposed uses and safeguards. Such gateways are used regularly and extensively to govern existing transfers of HMRC data to other public bodies, but it would be novel to use such a gateway to transfer HMRC data to private sector parties.

4.17 Compliance with safeguards would be reinforced by penalties, incentivising qualifying users to avoid breaching data use requirements. Breach could incur both contractual and legal penalties, as well as causing serious and potentially costly reputational damage to the relevant qualifying party.

4.18 As set out above, greater availability of data could make it easier to produce effective anti-fraud checks and measures, frustrating commercial fraud attempts in some circumstances. As with the proposals in chapter 3, HMRC would wish to be satisfied that, as well as potentially reducing general fraud risks, controlled release would not increase the risk of VAT fraud against HMRC, or of fraud against VAT registered traders (such as identity fraud). Given the security standards and safeguards that would be expected of qualifying users, and the fact that data use would be limited to qualifying purposes, fraud risk against HMRC or VAT registered traders is not expected to increase if controlled release proposals are implemented.

Consultation questions

Q11 Do you agree with the proposal for controlled release of the full set of VAT registration data to qualifying parties? Please give reasons for your answer.

Q12 What uses do you see for VAT registration data (see full list at Annex C) that would generate sufficient public benefit to justify controlled release? Please describe the uses and benefits in as much detail as possible.

Q13 What factors should be considered in determining the uses and parties that would qualify for release of the full VAT registration data set?

Q14 Do you believe that the proposed limits and safeguards are sufficient to mitigate any risks arising from controlled release? What, if any, additional impacts may arise and what further safeguards should be considered?

Q15 How should a controlled data release by funded? Please give reasons for your answer.

5. Summary of consultation questions

Chapter 2 Sharing non-identifying information

Q1 Do you agree that the legal constraints on sharing general and aggregate information should be removed, on a permissive basis, in respect of:

(i)   HMRC making information generally available through publishing?
(ii)   HMRC sharing information with specific third parties to deliver public benefits wider than HMRC's functions?

Please give reasons for your answers.

Q2 Do you agree with the proposed safeguards on the proposal to share general and aggregate data? Should any further controls be considered on what can be shared, with whom or how?

Q3 Do you agree that HMRC should be able to share anonymised individual level data for the purposes of research and analysis to deliver public benefits wider than HMRC's own functions? Please give reasons for your answer.

Q4 Do you agree with the proposed safeguards on the proposal to share anonymised individual level data? Should any further controls be considered on what can be shared, with whom or how?

Q5 How should the generation and release of anonymised or aggregated data be funded? Please give reasons for your answer.

Chapter 3 Publication of VAT registration data

Q6 What potential uses and benefits do you think would arise from publishing VAT registration data?

Q7 Do you believe that the proposed safeguards are sufficient to mitigate any risks arising from publication? What, if any, additional impacts may arise and what further safeguards should be considered?

Q8 Do you agree with the core proposal (Option A) to publish the VAT registration number, trading name and industry classification? Please give reasons for your answer.

Q9 What are your views on the proposal to allow exceptions or opt-outs to publication (Option B)?

Q10 What are your views on the proposal to publish additional data fields on a voluntary basis (Option C)?

Chapter 4 Controlled release of VAT registration data

Q11 Do you agree with the proposal for controlled release of the full set of VAT registration data to qualifying parties? Please give reasons for your answer.

Q12 What uses do you see for VAT registration data (see full list at Annex C) that would generate sufficient public benefit to justify controlled release? Please describe the uses and benefits in as much detail as possible.

Q13 What factors should be considered in determining the uses and parties that would qualify for release of the full VAT registration data set?

Q14 Do you believe that the proposed limits and safeguards are sufficient to mitigate any risks arising from controlled release? What, if any, additional impacts may arise and what further safeguards should be considered?

Q15 How should a controlled data release by funded? Please give reasons for your answer.

6. The consultation process

This consultation is being conducted in line with the Tax Consultation Framework. There are 5 stages to tax policy development:

Stage 1   Setting out objectives and identifying options.
Stage 2   Determining the best option and developing a framework for implementation including detailed policy design.
Stage 3   Drafting legislation to effect the proposed change.
Stage 4   Implementing and monitoring the change.
Stage 5   Reviewing and evaluating the change.

This consultation is taking place during stage 1 of the process. The purpose of the consultation is to seek views on the policy design and any suitable possible alternatives, before consulting later on implementation / draft legislation.

How to respond

A summary of the questions in this consultation is included at chapter 5.

Responses should be sent by 24 September 2013, by e-mail to mailbox.datapublicbenefit@hmrc.gsi.gov.uk, or by post to:

Data sharing consultation
Room 1C/22
100 Parliament Street
London
SW1A 2BQ

Paper copies of this document or copies in Welsh and alternative formats (large print, audio and Braille) may be obtained free of charge from the above address. This document can be accessed from HMRC Inside Government. All responses will be acknowledged, but it will not be possible to give substantive replies to individual representations.

When responding please say if you are a business, individual or representative body. In the case of representative bodies please provide information on the number and nature of people you represent.

Confidentiality

Information provided in response to this consultation, including personal information, may be published or disclosed in accordance with the access to information regimes. These are primarily the Freedom of Information Act 2000 (FOIA), the Data Protection Act 1998 (DPA) and the Environmental Information Regulations 2004.

If you want the information that you provide to be treated as confidential, please be aware that, under the FOIA, there is a statutory Code of Practice with which public authorities must comply and which deals with, amongst other things, obligations of confidence. In view of this it would be helpful if you could explain to us why you regard the information you have provided as confidential. If we receive a request for disclosure of the information we will take full account of your explanation, but we cannot give an assurance that confidentially can be maintained in all circumstances. An automatic confidentiality disclaimer generated by your IT system will not, of itself, be regarded as binding on HM Revenue and Customs (HMRC).

HMRC will process your personal data in accordance with the DPA and in the majority of circumstances this will mean that your personal data will not be disclosed to third parties.

Consultation Principles

This consultation is being run in accordance with the Government's Consultation Principles.

The Consultation Principles are available on the Cabinet Office website: http://www.cabinetoffice.gov.uk/resource-library/consultation-principles-guidance

If you have any comments or complaints about the consultation process please contact:

Amy Burgess, Consultation Coordinator, Budget Team, HM Revenue & Customs, 100 Parliament Street, London, SW1A 2BQ.

Email: hmrc-consultation.co-ordinator@hmrc.gsi.gov.uk

Please do not send responses to the consultation to this address.

Annex A: Relevant (current) legislation

Commissioners for Revenue and Customs Act 2005

HMRC was created by the Commissioners for Revenue and Customs Act 2005 (CRCA) and its functions are governed by that legislation. HMRC has no common law powers. HMRC's functions include managing and collecting tax and paying benefits, and matters ancillary to those functions such as employing its staff.

CRCA prohibits the disclosure of information held by HMRC in connection with its functions except in limited circumstances set out in legislation. This prohibition applies to all information held by HMRC in connection with its functions and reflects the importance placed on 'taxpayer confidentiality' by Parliament when the department was created. There is additional protection for information that relates to an individual or legal entity whose identity is specified in the disclosure or can be deduced from it ('identifying information'), in the form of a criminal sanction for unlawful disclosure.

Information gateways

The prohibition on disclosure does not apply where the exceptions ('gateways') set out in CRCA apply. The main gateways are:
- where there is UK or EU legislation that permits disclosure ('legislative gateway');
- where the disclosure is made for the purpose of an HMRC function ('functions gateway'); or
- with the consent of the subject(s) of the information.

Legislative gateways

HMRC shares information through legislative gateways with a range of public bodies including other government departments, agencies, the Devolved Administrations, Local Authorities and tax authorities in other countries.

The terms of each information gateway are specific as to the type of information that can be disclosed and the purposes to which it can be put.

Functions gateway

The prohibition on disclosure does not apply to a disclosure that:
- is made for the purposes of a function of HMRC; and
- does not contravene any restriction imposed by the Commissioners for HM Revenue and Customs.

At present, no restriction has been imposed by the Commissioners.

Consent

HMRC will only use the consent gateway if the following are evidenced:

- there is a positive indication of the wishes of the data subject;
- the consent is freely given; and
- the consent is fully informed and specific.

Data Protection and Human Rights Acts

HMRC is subject to the Data Protection Act 1998 (DPA) and the Human Rights Act 1998 (HRA). Any disclosure of data must be in accordance with those principles, including the level of disclosure being 'proportionate' to the ends which the release is intended to achieve

CRCA and the Freedom of Information Act (FOIA)

A disclosure by HMRC in compliance with FOIA is a lawful disclosure under CRCA, but information relating to identifiable individuals or legal entities is exempt from disclosure under FOIA.

Annex B: Research and statistical benefits of sharing anonymised tax data for non-tax purposes

Background

The Administrative Data Taskforce (ADT) Report *Improving Access for Research and Policy (December 2012)* made recommendations to help harness the potential of administrative data to improve the evidence base for policy, and to facilitate research for the wider benefit of society.

A key recommendation was to enact a generic legal gateway for research and statistical purposes, enabling efficient access to, and linkage between, administrative data held by different government departments, agencies and other statutory bodies. The Report recognised that government departments, local authorities, agencies and other public bodies face different legal restrictions on how they are able to provide access (for research) to data they control, and pointed out that creating legal gateways for specific projects was both cumbersome and inefficient.

The HMRC Datalab

The HMRC Datalab was launched in May 2011 as a new Research Data Centre (RDC) allowing access for approved academic researchers to anonymised administrative and survey data. The aim of this initiative is to produce high quality research that benefits both the Department and the wider academic community, in the form of a wider evidence base to support knowledge sharing and policymaking.
Within existing legislation, only research that serves HMRC functions under CRCA 2005 is permitted. The Datalab has already hosted 18 projects that investigate tax related matters. Examples of approved projects can be found on the HMRC Datalab website at:

http://www.hmrc.gov.uk/datalab/research-topics.htm

Only non-disclosive statistics, tables, graphs, charts etc may be taken out of the Datalab.

Benefits of relaxing legal constraints on sharing anonymised data

Relaxing current legal constraints would make it easier for HMRC to share anonymised information for statistical and research purposes to deliver wider public benefit, rather than needing to benefit HMRC directly. Information could be shared with other government departments and approved researchers, through the Datalab, Administrative Data Research Centres (ADRC) and other channels as appropriate. If legislation were introduced to permit research for non-tax purposes, this would create opportunities to advance the evidence base in a range of areas related to, for instance, national and regional economic policy, social mobility, health and environment.

The ADT Report points to several areas where access to income data would significantly enhance policy research. These include:

- addressing social mobility – by linking data on education, training, employment, unemployment, **incomes** and benefits;
- researching causal pathways over the life course – linking data on education, health, employment, **incomes** and wealth;
- informing policies designed to tackle poverty – linking data on housing conditions, health **incomes** and benefits; and
- exploring the role of poor mental health - linking data on (re)offending behaviour, **incomes**, benefits and health.

In addition to current Datalab projects, HMRC has received requests via the Social Mobility Transparency Board and other routes to undertake research that requires access to income data for matching with education records. This could significantly improve the evidence base on how individuals' educational opportunities and choices influence short term and long term earnings potential, and improve the information available to individuals before they make those difficult decisions. HMRC have been able to satisfy some of those requests where the research also serves one of HMRC's functions or are covered by an existing legal gateway, but there is a very wide range of other research or statistical analysis that would be facilitated by a change to legislation to permit wider data sharing in the public interest.

Similarly, HMRC data on corporate tax and trade has the potential to enable better answers to research questions such as the drivers of productivity growth and the links between trade, productivity and growth. Again we have been able to meet some of the requests for data access in this area where they also serve HMRC's functions, but there are many more that would be facilitated by a wider legal gateway.

If HMRC were given a general power to share anonymised individual level data for research and statistical purposes, this would remove the need for the creation of specific legal gateways for each new requirement – a process that can be resource intensive and difficult to schedule in an already heavily subscribed parliamentary/legislative calendar.

In undertaking any wider sharing of its data, HMRC would need to ensure that any risks of identification of personal or confidential data were minimised, as is already achieved in the HMRC Datalab. The ADT report makes a number of recommendations about how to achieve that, including anonymisation of data, use of trusted third parties to link the data, use of state of the art safe environments with levels of security commensurate with the sensitivity of the data, accreditation of researchers using the data, and disclosure control on research and statistical outputs. The Administrative Data Research Network and associated governance framework would enable HMRC to ensure that decisions on wider sharing of its data were made in a transparent fashion and that the data sharing and access were accomplished safely.

Annex C: VAT registration data set

This table shows the set of VAT registration data (containing identifying data supplied by businesses at registration and subsequent updates) within scope for this consultation and proposed for controlled release to qualifying parties (chapter 4).

The fields proposed for open data publication (the core proposal in chapter 3) are VAT registration number, full trading name and current trade classification (SIC).

Historical data e.g. dates when fields were changed and previous contents are also held. There are some fields in the VAT registration database that are excluded from the scope of these proposals – in particular, financial data such as business estimates of turnover at registration are excluded because of their sensitivity.

	Types of data	Data fields
1	**Registration number**	VAT Registration Number
2	**Name**	Full Organisational Name (i.e. name of proprietor or entity) Full Trading Name
3	**Contact details**	Current Address (i.e. principal place of business) Telephone Number Mobile Number Email Address Fax Number Web Address
4	**Status details**	Current Trade Classification SIC Code 40 character (maximum) Trade Classification SIC Code description Legal Entity Status (e.g. partnership, sole proprietor, company, public corporation, local authority, other) Incorporation Number (if company) Incorporation Date (if company) Effective Date of Registration Date of Insolvency (if insolvent) Date of Deregistration (if deregistered) Date of Transfer of Going Concern

Annex D: VAT registration data release research project

HMRC will be undertaking, over the summer, a research project to help assess potential external uses of VAT registration data, possible benefits, and issues which arise from such uses. This will assist HMRC policy development and enhance HMRC capability to undertake compliance functions. The research project will inform any Ministerial decision on release of VAT registration data. The research project will be governed by the full suite of existing controls on data release (see below).

To make these assessments, HMRC will work with selected credit reference agencies (CRAs) which hold information on businesses (and which are part of the Business Information Providers Association).

HMRC will provide the CRAs with a sample of approximately 70,000 trader records, including a randomly drawn representative sample and a geographically-based sample. The CRAs will determine the potential of the VAT registration data to improve services such as calculation of credit scores/limits and fraud risk assessment, with HMRC validating the approach.

HMRC and the CRAs will work together to determine what impact improved services could have on market outcomes, for example, changes in trade credit availability and uptake, and their associated economic effects. The CRAs will seek to identify if VAT registration data could be supplemented or enriched by data held by the CRAs, assisting HMRC in its collection and management of taxes.

As required by HMRC's legislation, the project will be subject to strict controls. In undertaking the research project and handling HMRC's VAT registration data, the CRAs will be bound by the same strict statutory privacy and data protection rules which govern HMRC's activities and which prevent any misuse or unauthorised release. The VAT registration data provided will need to be held securely, only be used for the purposes of the research project and will be destroyed when the project is finished. Individuals will not be identifiable in the project outputs (which will be aggregated assessments), nor will the CRAs provide data on individuals back to HMRC.

The research project will be concluded in time to feed into Ministerial consideration of the outcome of this consultation. A summary of the outcomes will be included in the Government's summary of responses to this consultation in the autumn.