

# ODUG Response to HMRC Data Sharing Consultation

## Introduction

The Open Data User Group (ODUG) is an independent advisory body appointed by the Chair of the Group and the Cabinet Office Minister. It receives some administrative support from staff in the Cabinet Office, but speaks independently on behalf of the data community. The views of ODUG should not be construed as representing the position of Her Majesty's Government.

Individual members of ODUG, or their employing organisations, may make separate responses.

The Open Data User Group (ODUG) welcomes the opportunity to respond to the HMRC Data Sharing Consultation which closes on September 24, 2013. Two members of the Open Data User Group, Paul Malyon and Dominique Lazanski, have been in close contact with the HMRC team leading on this consultation. We are grateful to the HMRC team for meeting with us throughout the process so that we could discuss the consultation ahead of our submission. We would welcome any further questions or discussion about our consultation response in person.

In December of 2012 ODUG made the case for the release of the VAT register. (See <http://data.gov.uk/benefits-of-releasing-an-open-vat-register> and full text in Appendix 2) In it we put forth the case the Business Information Providers Association makes; opening up of this data could add £50 billion to the UK economy. In this consultation response we are expanding on that business case to widen our evidence.

Additionally, as part of our further research, we realise that there are over twenty unreleased HMRC owned and potentially VAT related data sets on data.gov.uk ranging from VAT register information to VAT payment information to VAT mainframe data. For the purposes of this response the Open Data User Group refers primarily to the VAT register information as discussed in the consultation documents but generally supports the premise that opening up further data will create value for HMRC and the economy.

The Open Data User Group advocates the release of government held data on behalf of users. In order to support our assertion that data should be made available for free reuse under an open government license (OGL), we provide case studies and supporting documentation. Our response to HMRC's Data Sharing Consultation will be as follows: first, we will answer the questions below that we deem relevant to the work we do; then we will provide a list of case studies; finally, we will close with a conclusion about next steps.

***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?***

Currently, HMRC data is not made available due to the legal position they find themselves in (namely that data can only be collected and used for the purposes of tax collection). This prevents academia, public bodies and businesses from reaping the benefits of various statistics and datasets.

It is important that any non-personal data is made available on an Open, free basis wherever possible to ensure that access is fair to all. While there may be a cost to HMRC over and above current data hosting and maintenance fees, the return on investment of releasing data for free re-use could be significant.

It is of critical importance that HMRC fully understands the legal framework under which its IT suppliers host HMRC data (or provide infrastructure) to ensure that these contracts do not prevent the release of data or add additional cost for regular extraction / use.

***(ii) HMRC sharing information with specific third parties to deliver public benefits wider than HMRC's functions? Please give reasons for your answers.***

This opens up opportunities for more data sharing but HMRC must carefully consider the privacy of any individual featured in the datasets and have relevant safeguards in place for data shared with specific third parties.

The requirements of the Data Protection Act (and other requirements linked to financial data), may preclude some organisations from accessing the data. Putting in place the security and other technology & business processes required to manage personal / financial data could be costly for start-ups or for business cases with a low ROI. Therefore, the sharing of aggregate and general information should be opened up for free access, taking into account the type of data released, e.g., non-financial data, and the fact that HMRC will adhere to data protection and privacy safeguards.

While the aim to release data to as wide an audience as possible is good, the risks associated must be understood by all. It is also important to clarify what 'public benefit' would be as this may include economic benefits as well as social benefits.

***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?***

There is always a risk of aggregate data being linked to other datasets to enhance the detail of the information available. In fact, in many cases this is the whole point of a data product so is beneficial. However, having clear and concise guidelines in place is a sensible approach.

Having a process to 'vet' data requests is also a sensible option but could become resource-hungry if many requests are lodged for multiple datasets. If data is released under the OGL though, this will be unnecessary. It's also worth noting that the existing ODUG process for requesting data releases could be used to do much of the work before requiring any HMRC resource.

In terms of further controls, it really depends on the data to be released and level of detail contained within. Any dataset containing personal information should be made available only with the relevant DPA safeguards in place and only to those organisations who can demonstrate strong security procedures and a data use that is complementary to the specified purposes defined by HMRC. Although, whilst HMRC should take reasonable steps to protect personal data it must acknowledge that it is, ultimately, the responsibility of the recipient to ensure that they do not contravene DPA requirements.

For any data at a postcode level of granularity or above, existing public sector data sets already made open give a good indication that minimal safeguards are required (some good examples are the ONS NSPD and OS Code-Point Open).

Finally, attention should be paid to the VIES system, a European Commission service which allows an EU VAT number to be verified, on-line, for free. As public access to VAT registration information is already made possible through this portal, this sets a precedent for the release of address-level information. Also, though cumbersome, the VIES service could be used to 'fill in the gaps' if HMRC choose to hold back address information from any publication of VAT data.

***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.***

The HMRC already make some aggregate data available to what they deem as 'trusted parties' within their Datalab. These trusted parties can only use the data for academic research which supports the HMRC. Their results must also be shared publicly.

In 2 years of operation, only 2 papers have been published via this method from the 24 studies either completed, underway or about to begin. There may be a number of reasons for this relatively low output, for example the tenure of research projects, the requirement for papers to be peer reviewed, changes in PhD focus and so on.

ODUG recommends that, in order to increase the volume of studies and the social & economic benefit of these, the Datalab should be opened to any interested party willing to fund the activities and share their results publicly (in a form that does not endanger personal privacy or commercial sensitivity). By charging a modest fee to businesses accessing the Datalab, HMRC could re-invest in further data releases.

By limiting permitted research to the functions of HMRC, the value of the public information held by the organisation is being severely limited. HMRC should widen the pursuit of benefits beyond their own functions to take account of wider economic and social benefits and allow more value to be derived from the data.

Beyond the Datalab, HMRC should also look to release sample data for any purpose to help data businesses test ideas before engaging more formally for access to the lab or full data sets. By limiting research to the functions of HMRC, the value of the public information held by the organisation is being severely limited.

***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?***

If data is anonymised then the risk of fraud or personal data issues are minimised. However, something to be aware of is that personal data may be implicit in some of the anonymised data. For example, small businesses may well be named after their owners (for example, John Smith plumbers), with the business address being the residential address of that individual. For a registered Limited Company this information is already public at Companies House. For Non-Limited businesses and VAT registered Sole Traders, a release of VAT Register data could be the first time that information on these businesses was made 'public'.

The introduction of an opt-out mechanism, perhaps on the new online tax form or at the point of registration for new VAT numbers, would be the most practical way to provide the non-limited and sole trader businesses mentioned above the option to have their business information removed from the published dataset. The challenge is to ensure that all registrants, both current and new, have the opportunity to opt-out and that this option is available to all.

If data is anonymised, then there should be no limitations imposed on re-use under the Open Government License and the data should be available at zero cost.

Careful consideration should be given to what 'anonymous' means in the context of an address. The VIES system already allows a VAT number to be used to find an address. It's not clear what opt-out controls are in place today, if any. We would recommend a review of the VIES process alongside the HMRC process of releasing VAT-related data.

***Q5 How should the generation and release of anonymised or aggregated data be funded?***

***Please give reasons for your answer.***

Anonymisation of data should be an extremely simple process (e.g. remove the Name fields from the data) with aggregation easily possible by removing address elements below Postcode or street level.

Many public and private bodies already do this so advice will not be hard to find. It is also probable that much of the HMRC data has already been anonymised during the work of the HMRC Datalab. In fact, assisting with this could be made a condition of accessing data within the lab.

In short, no extra funding should be required for anonymous Open Data as the costs are minimal and the work may already have been done. For any personal level data that is not open, a small charge to cover costs may be appropriate but this should be discussed with organisations that may benefit from the data directly, to gauge the appropriate level.

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

Much of this had been covered in the ODUG business case (Appendix 2). There may be additional advantages outside of the business finance arena such as using the VAT number as a validation tool when a business registers for a consumer-facing service such as Rated People or trade credit from an organisation such as Screwfix (see Appendix 1). It may also allow consumers to validate that a business exists when they receive a quote or invoice (and to use this data linked to other business information to choose businesses that are better rated / lower risk).

The increase in transparency may also result in a fall in VAT fraud if simple validation tools are made available to businesses and consumers.

Other countries have already published VAT (or similar) data to enhance trust in businesses, improve tax collection (and cut tax fraud) and increase international trade. Further examination of the benefits realised in these Countries may help identify additional UK benefits (see Appendix 1). As we have noted above, the EU VIES system allows searches of VAT numbers to find company details. While this is not quite the same as an open bulk dataset, VAT register data is already publicly available for UK businesses for purposes which may be outside of pure tax collection.

***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?***

Suggestions for an opt-out are well intentioned but will (as the HMRC document acknowledges) result in a dataset of reduced completeness, and therefore, of reduced value. However, much like the electoral roll (where benefits include simpler access to credit), being in the public domain for citizens, the VAT register would provide similar benefits to businesses. If this is clearly explained, opt-outs would be minimal and the value of the data would be maintained.

The proposals for businesses to volunteer enhanced data are also well intentioned but would create higher costs for HMRC, in order to manage these responses. A simple opt-out would be the favourable method and a method which could take place, as mentioned

above, at the point of tax return filings or new registration.

With any release of data, there is obviously some risk. However, the availability of the data allows identification services to be created and improved to include many more businesses than before, improving onward access to funding, company mobile phone contracts, fleet vehicle access, 'trade' accounts and other business services.

***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.***

Yes and No. Publishing the VAT number, trading name and industry classification is good but without an address, the value is limited as this information will be key for matching the new data to previously available information.

Also, for those wishing to contact a business, the address and phone number is very important. Having no address in the data makes this impossible nullifying many of the benefits that data release could bring.

It is a fair point that some business addresses could be the same as the owner's residential address. Therefore, an 'address opt-out' should be offered (i.e. a modified Option B) or address details could be limited to the postcode or street level.

Regardless of any opt-outs the European Commission VIES system could be used via the currently available SOAP API interface to take a VAT number and source all addresses. The VIES system sets a precedent for this data to be open and available. Therefore, it should be published openly to avoid web developers over-using the VIES system.

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

Allowing an opt-out where a business name or address includes personal information (or the individual has concerns about this) is perfectly acceptable for an Open Data release. Clear guidance should be supplied to business owners about what data is held on them, what will be made available and the expected/possible impacts of opting out.

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

This could add an extra level of management (and cost) for HMRC. A self-serve portal for business owners could make this possible but would need to be set up and maintained.

Private sector businesses offer similar 'update your data' services today with varying success and whilst an HMRC-owned solution where the business owner can tick which fields they want published would make sense it would require training and promotion to users and

would also lead to a patchy dataset.

***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.***

The Open Data User Group supports the release of non-personal information to all in an open format with zero license fee or use limitations.

For additional sensitive information, some control of the data could be sensible to ensure that those using the data do so in alignment with the DPA. An outline of the 'sensitive' data to be included should be shared with stakeholders to examine what limitations would be applicable. A review of the data fields described in Annex C does not present an overly risk-laden dataset over and above some of the contact details (for example, person names could feature in organisation names and a mobile phone could also be for personal use).

Careful consideration of what constitutes a 'qualifying party' should also be undertaken. Assumptions would indicate that those involved in business lending decisions, who already hold consumer data for these purposes, would be included. However, which other types of organisations could meet the requirements? Are they in the public and / or private sector? Examples could include organisations like Open Corporates and Rated People.

***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.***

The Open Data User Group support the open release of data. With this expectation in place, there would be public benefit from much of the data. Specific fields of interest include:

- Registration Number - an excellent unique identifier of a business that would help de-duplicate databases to avoid wastage & create a single business view across databases, systems and services.
- Name - a valuable field to understand the businesses that are located in an area.
- Contact details - Address, phone number and website are paramount for making data available to products and services such as business listings and comparison sites.
- Status details - certainly of use in metadata for providers of business information products with the SIC code and company description probably the most useful elements. Some other elements may also be useful for identity verification but this may mean that making them openly available might decrease the effectiveness of this.

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

Provided that an opt-out, privacy mechanism is in place and that the HMRC is meeting the legal requirements as discussed above, the release of this data should be made available free under the Open Government License (OGL). We would recommend HMRC continues to discuss data protection concerns with the ICO.

**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?**

We recognised that there are a number of issues regarding personal data safeguards. We take privacy seriously, however, we propose that a privacy mechanism be put into place as discussed as opposed to a controlled release with only a limited number of trusted and approved organisations.

A controlled release not only has to be managed by HMRC, and could be seen as offering competitive benefits to some organisations over others, but also reduces the wider benefits that are available from the full release of this data.

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

In 2003, the European Commission adopted the Directive on the re-use of public sector information. In the re-use of public sector information documents held by the commission. It says:

“Article 7 Principles governing charging: 1. The re-use of documents shall in principle be free of charge. 2. In specific cases, marginal costs incurred for the reproduction and dissemination of documents may be recovered.”

In 2008 the OECD produced recommendations for public data based on a number of working groups and workshop meetings. The result, entitled *The OECD Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information*, makes the case clearly for best practices and enhanced use of Public Sector Information. Many of the issues that arise in this consultation are clearly outlined in the document with recommendations covering privacy, security, quality, integrity and copyright best practices and principles.

In particular, the OECD Recommendation notes the idea of pricing and says:

**“Pricing.** *When public sector information is not provided free of charge, pricing public sector information transparently and consistently within and, as far as possible, across different public sector organisations so that it facilitates access and re-use and ensures competition. Where possible, costs charged to any user should not exceed marginal costs of maintenance and distribution, and in special*



*cases extra costs for example of digitisation. Basing any higher pricing on clearly expressed policy grounds.”*

It is clear that there is a precedent for the not charging for the release of the data and that, ideally, HMRC should bear the cost of the release.

If the costs to HMRC are prohibitive then there are several options available to help avoid commercial licensing. Firstly, the HMRC Datalab already contains mechanisms for anonymising, storing and publishing data. Re-tooling this infrastructure could provide a cost effective short cut for full release.

Additionally, organisations wishing to access open HMRC data could be asked to contribute hosting or other infrastructure in-kind or to cover the initial set up costs and offer a number of consultancy days each year in return for continued access.

## **Conclusion**

The Open Data User Group has previously made the case for the release of VAT registration data. While we are aware of the importance of the privacy of individuals and the ability to opt-out, we recognise the importance that this data and other HMRC data assets will have for all who wish to use it, for free and under the Open Government License (OGL). Furthermore, we would welcome the release of additional data so that the name, contact details (with address and postcode), VAT registration number and status are available. This information should be published once privacy checks are in place. Additionally, the release of anonymised, aggregate data should happen as quickly as is feasible as put forth by the HMRC in the consultation.

The release of data to all, and not just to individual, approved, entities, should be a priority. Once the risks are mitigated, as mentioned above, the benefits of the data will be achieved for both HMRC and the wider economy / society.

## **Appendix 1 - Case Studies**

### **Rated People**

Rated People is a website that links homeowners requiring a builder, plumber or other tradesperson (there are 32 categories) with local businesses who are rated on quality & price by the users of the site. The idea of this is to encourage improvements within the industry and to avoid the 'cowboy builder' label unfairly given to many tradespeople.

Members of the public post their requirements on the site with individual tradespeople paying a basic fee of £15 per month to be able to 'bid' for the job. Further fees are charged to find the latest 'leads' with SMS alerts also provided.

This social approach to sourcing tradespeople has proved successful with over 1 million jobs posted by the public with over 300,000 ratings given.

During research for this response, ODUG contacted Rated People to discuss whether the VAT register and other PSI from HMRC would be of use to their business. Their response covered the following main points:

- With 80% of their members being sole traders, it has proven difficult to verify business identity while using personal identity for a tradesperson is not always suitable. For example, a driving license number cannot be linked back to the person and may not be relevant where a company is not a sole trader.
- Being able to confirm whether a tradesperson is VAT registered may offer further reassurance to a customer that this person is not a 'cowboy builder'. This then increases the quality of service received leading to more business for Rated People, their members and the public.
- Rated People stated that it was difficult to estimate the benefit to their business but would hope for at least a 5-10% revenue improvement on top of an improvement in the quality of the listings provided to members of the public.
- Rated People also noted that other organisations could benefit from a VAT number validation service such as Screwfix and Wickes. Being able to validate a small business customer will prevent the risk of VAT fraud and ensure that the benefits of a business account are given to stable businesses more likely to be credit worthy.
- It was also noted that sole traders are not always required to be VAT registered. Rated People would also benefit from the release of additional information from HMRC or other public bodies that would help match a sole trader business address or other unique data element to their business. Perhaps National Insurance Number or self-employment status would help.
- Rated People supported the release of the VAT register in an open, free format and also supported the calls for an opt-out system similar to that employed in the Electoral Register.

## Examples from other countries:

A number of other countries around the world release VAT equivalent data to different extents. We have researched those countries and have come to the conclusion that the effects on releasing data has little impact on fraud. The following countries are just a few of the many examples around the world.

Belgium:

<http://data.gov.be/dataset/vat-registered-businesses>

Belgium releases the number and growth of VAT registered individuals and business by region. This is historical data available from 2007.

Australia:

<http://abr.business.gov.au/Webservices.aspx>

Australia provides their equivalent of VAT called ABN. Fields provided include Name of trader or business, status, entity or business type, Goods and Services Tax registration and postcode.

Denmark:

<http://www.skat.dk/SKAT.aspx?old=69073>

Denmark has published a data set of companies, their registration number, entity type and tax paid along with the profit or loss for that company over the last tax year.

Slovenia:

[http://www.durs.gov.si/si/storitve/seznami\\_davcnih\\_zavezancev/](http://www.durs.gov.si/si/storitve/seznami_davcnih_zavezancev/)

Slovenia has published a list of taxable individuals, companies, and VAT registrants. In the case of Slovenia, the company and/or individual number is the same as the VAT registration number. Other fields include date of registration and activity code.

## **Appendix 2 - Original ODUG Business Case**

### **Description & Request Overview**

HMRC hold the dataset known as the VAT Register which contains information on businesses such as; VAT number, Business status (i.e. Sole proprietor, partnership, corporate, etc.), Name and business address, Contact information – telephone, email, website, Business activities, Other business associations in the past 2 years, New registration or ceasing registration, Turnover, Date of Establishment, Applicants Name & Address.

Releasing some or all of this data will enable banks to make better lending decisions to increase the availability of trade credit.

### **Data Release Rationale**

BIPA (Business Information Providers Association) have estimated that the VAT register could benefit UK PLC to the tune of around £50 billion. This is based upon an increase in credit requests and associated credit approvals powered by the positive data from the VAT Register.

The VAT Register will be positioned as an additional source of information to those provided by Companies House and member of BIPA. It will also be used by organisations such as the members of BIPA to add value to existing data products and services, further increasing the positive economic benefits for their customers.

This enhanced data could be returned to HMRC to improve the efficiency of tax collections to provide greater revenue and / or cost savings.

Businesses (including SMEs and lenders) will be able to comply more easily with anti-money laundering (AML) requirements. This creates positive business efficiencies and cost savings.

Commercial fraud may be reduced through the enhanced identification services enabled by the VAT Register. This has obvious economic & social benefits. In 2011, CIFAS members reported 250,000 instances of commercial fraud (a 9% year on year increase).

While the data should be released openly, some elements could be offered as an enhanced, paid for product by HMRC (similar to the Companies House model).

There are European Countries that already have a full business register including both incorporated and unincorporated businesses e.g. France and some that in addition use the VAT number as the basis of business identification e.g. Belgium.

### **Benefit Description**

BIPA (Business Information Providers Association) have estimated that the VAT register could benefit UK PLC to the tune of around £50 billion through improved access to funding for small businesses.

This is based upon an increase in credit requests and associated credit approvals powered by the positive data from the VAT Register.

The VAT Register will be positioned as an additional source of information to those provided by Companies House and members of BIPA.

It will also be used by organisations such as the members of BIPA to add value to existing data products and services, further increasing the positive economic benefits for their customers.

This enhanced data could be returned to HMRC to improve the efficiency of tax collections to provide greater revenue and / or cost savings.

Businesses (including SMEs and lenders) will be able to comply more easily with anti-money laundering (AML) requirements. This creates positive business efficiencies and cost savings.

*Please identify where further case studies or quantifiable evidence to support the release of this dataset?*

### **Benefits Analysis**

Release of the VAT register as Open Data will:

Enable the release of business credit leading to growth of the UK Economy by up to £50 billion (BIPA estimate).

Improve access to business funding for 750,000 businesses.

Reduce the instances of commercial fraud (250,000 cases in 2011).

Enable the creation of new services such as 'trusted business' apps with the VAT Register acting as the unique data identifier.

Create internal government efficiencies through better data sharing and tax gathering.

*Can you identify further areas where this dataset release will create opportunities for innovation and new business?*

*Can you help identify sectors, businesses and organisations that will benefit from the release of these data?*

### **Barriers and Requirements for Release**

It is assumed that the VAT Register data contains no information of a personal or sensitive nature.

Any sensitive data would need to be removed before release based upon advice from the ICO.

HMRC would need to make the data available on a regular basis (weekly or more regular is recommended).

HMRC would need to make the data available as a download or API.

While one of the main aims of the request is to reduce commercial fraud, risks posed by the data (in terms of enabling fraud) will need to be examined by Government with partners such as BIPA.

An 'Opt-Out' scheme for Unincorporated business owners (similar to Edited Electoral Roll system) could prevent any negative perception of the data.

Consideration should be given to how the data will be used (e.g. marketing, credit or other).

*Can you identify other barriers to this data release, or solution to those listed?*

### **Recommendation**

The business case for releasing the VAT Register is clear. Today, small business credit scores are approximately 40% lower than is expected due to a lack of positive data available to credit reference agencies (and in turn, lenders). This makes access to trade credit lower than it should feasibly be.

By providing the VAT number and additional details; business information providers will be able to create linked, up to date data assets to improve the chances of successful, well managed small businesses having access to the kinds of funding required to help them grow.

Without access to the VAT Register, business lending will remain sluggish and the chances of success for many small enterprises will be endangered.

The Open Data User Group therefore call on HMRC to release the VAT Register as an Open Database under the Open Government License.