

“Making Open Data Real: A Public Consultation”

Demographics User Group’s response

(8 October 2011)

1. Introduction

The Demographics User Group (DUG)¹ has the objective “To work with government on behalf of commercial users to ensure that the right data is made available, in the right way, at the right time”, and represents the views of 15 major commercial companies – Barclays, Boots, Camelot, Co-operative Group, E.ON, Everything Everywhere, GlaxoSmithKline, John Lewis, Marks & Spencer, Nationwide, Sainsbury’s, Serco, Tesco, The Children’s Mutual, and Whitbread.

We would emphasise that these are the tip of the iceberg of 2.3 million companies in the UK which can benefit from using information gathered by government in carrying out its public tasks. All too often discussion of Public Sector Information and “the private sector” focuses on private sector resellers, not the much larger number of potential users.

Many commercial companies already make extensive use of geographical and statistical information to target local markets and consumers, and we strongly believe that the government’s policy of Open Data will further promote greater efficiency and growth.

2. Our summary reaction to the proposals

We are very supportive of the main themes of the consultation document:

“The best way to tap into the UK’s tradition of creativity and invention is to give that data away”

“Our proposed approach is, fundamentally, about creating both “pull” (a right to data) and “push” (a presumption of publication)”

“Where data about public services is held outside the public sector, we will work with the service providers to free up that data”

“We must, of course, ensure that privacy is preserved and that personal data is protected”

“We will publish data of lower quality in preference to holding it back, while seeking over time to drive up the quality of that data”

¹ <http://www.demographicsusergroup.co.uk/>

Turning to the question of “how far we should go”, we believe that more radical options for shifting the Free / Charged dividing line than those set out in the accompanying Public Data Corporation consultation should be considered.

3. The context of our response as users of government data (referring to the issues raised in Sections 3-7)

3.1 Government is already a key source of information about people and places in the UK. Commercial companies already use many datasets which are available from government in the form of 1) statistics for small areas (e.g. the Census); 2) geographical information (e.g. Ordnance Survey maps); and 3) lists of properties or people (e.g. Postcode Address File).

3.2 DUG's priorities for further access to government data:

Category	Datasets / Topics	Source
Statistics		
Current Neighbourhood Statistics to be recreated at Output Area level (c.f. the current higher / less valuable Super OA level)	The existing wide range of Neighbourhood Statistics	<ul style="list-style-type: none"> ONS and government departments
New statistics at small area level, using administrative databases.	Start with the topics identified by the Beyond 2011 project	<ul style="list-style-type: none"> ONS and government departments
Government's existing sample surveys (e.g. those held at the University of Essex Data Archive)	The Living Costs and Food Survey is of particular interest to commercial companies	<ul style="list-style-type: none"> ONS / ESRC
Geographical data	<ul style="list-style-type: none"> All the mapping needed by government, and provided for in the recent Public Services Mapping Agreement Infrastructure developments & plans National Statistics Postcode Directory – omitted fields, e.g. delivery points Flood maps 	<ul style="list-style-type: none"> Ordnance Survey LAs / CLG / OS ONS Environment Agency
Lists – individual records	<ul style="list-style-type: none"> The new National Address Gazetteer Council Tax bands for domestic properties House sales & their prices House building completions Planning applications – domestic and business properties Valuation lists for business properties Company information County Court Judgments for debt – personal & corporate Electoral register (if not opted out) Addresses of premises (schools, hospitals, surgeries, clinics, etc.) 	<ul style="list-style-type: none"> Geoplace VOA Land Registry LAs / CLG LAs / CLG VOA Companies House MoJ & Registry Trust LAs (various)

UNCLASSIFIED

UNCLASSIFIED

3.3 Current barriers, and potential solutions

- 1) Charging, and charging policies. These can create major barriers. For example, charges for the new National Address Gazetteer for any organisation or person outside the public sector range from £24K p.a. for one user to £189K p.a. for a big organisation. We have no doubt that much lower charges would attract many more customers.
- 2) Complex license agreements. These deter many potential users, who simply walk away, rather face the prospect of a long dialogue with company lawyers. Government initiatives in this area are very welcome.
- 3) Existing statistics – unfriendly formats. These have caused difficulties over the years, but we are optimistic that organisations such as ONS now understand the need for popular formats such as csv.
- 4) Finding existing datasets. Data.gov.uk is very welcome, but users would benefit considerably if the system could help to narrow down choices by identifying the geographical level of each dataset, and the topics covered.
- 5) 3rd Party intermediaries. Sometimes (e.g. in the Health sector) a third party can be successful in positioning themselves between government and data user. These agencies perform some value added tasks, such as formatting and distribution, but seem to charge a disproportionate fee for doing so. The value they capture in this process could be shared more equally between government and other users.
- 6) Re-creating existing datasets to Output Area level. Very few datasets have been created at OA level since the 2001 Census, yet this is the level which provides most value. Simply aggregating administrative records files to OA level, rather than going straight to higher geographies, would be a quick win.
- 7) Creating new statistics at OA level from administrative files. This presents huge opportunities (e.g. HMRC statistics on local incomes), and should benefit from ONS's current Quality Assurance comparisons for the 2011 Census, and its assessment of such sources for Beyond 2011.

3.4 Minimising the need for government to charge

We believe that the scope of the accompanying Public Data Corporation consultation is insufficiently radical, and should also consider:

- Starting with government's needs for information (the public task), rather than the interests of the existing supplier trading funds and agencies
- What would be the most efficient solution if starting afresh, which might lead to the idea of merging existing suppliers which have apparent commonalities (e.g. Land Registry, Valuation Office, and Ordnance Survey).

- Market testing / outsourcing, to see how much cost can be cut.
- Setting the Free / Charged boundaries on the basis of raw data being free, (to encourage use by end-users, and developers / resellers), with charging for some value-added products.

Further comments are given in our response to the PDC consultation.

DUG's responses to the consultation's questions

Questions for consultation (page 6)

1. Do the definitions of the key terms (on page 5) go far enough or too far?

They are a good starting point – best not get too bogged down in definitions.

2. Where a decision is being taken about whether to make a dataset open, what tests should be applied?

The extent of potential use, and hence value – a presumption of publication, unless there are reasons against.

3. If the costs to publish or release data are not judged to represent value for money, to what extent should the requestor be required to pay for public services data, and under what circumstances?

If it is a special ad-hoc request, unlikely to be of interest to many others, the cost of production should be charged. One example is the 2001 Census commissioned table service.

Another example might be for some if some value added task has been completed on the data prior to release. However, most users would prefer quick “raw” data vs. slower processed data.

4. How do we get the right balance in relation to the range of organisations (providers of public services) our policy proposals apply to? What threshold would be appropriate to determine the range of public services in scope and what key criteria should inform this?

On the demand side, we suggest the interest of the particular topic to a mass audience (e.g. crime, health, income, poverty, etc.). On the supply side, a key issue is whether national coverage is achievable, either directly (e.g. DWP), or indirectly (e.g. co-ordinating Local Authorities): local patches of data are much less useful.

5. What would be appropriate mechanisms to encourage or ensure publication of data by public service providers?

Instruction and pressure by ministers, backed up by the Information Commissioner with appropriate powers.

Policy Challenge Questions (Section 8, page 22 onwards)

An Enhanced Right to Data

1. How would we establish a stronger presumption in favour of publication than that which currently exists?

The use of diktats by ministers, alongside appropriate powers of an Information Commissioner.

2. Is providing an independent body, such as the Information Commissioner, with enhanced powers and scope the most effective option for safeguarding a right to access and a right to data?

Yes, some form of regulator is necessary.

3. Are existing safeguards to protect personal data and privacy measures adequate to regulate the Open Data agenda?

Yes. Indeed, government officials have been too cautious during the last decade, often focusing on very small / obscure risks of disclosure, whilst giving much less weight to utility / value.

4. What might the resource implications of an enhanced right to data be for those bodies within its scope? How do we ensure that any additional burden is proportionate to this aim?

There are many possible quick wins: a) data that is already charged for (e.g. by trading funds) is available now; b) producing small area statistics at Output Area level needs requires the same amount of processing as for higher levels; c) ONS's investigation of many administrative sources for 2011 Census Quality Assurance, and the Beyond 2011 project will produce area statistics as products.

5. How will we ensure that Open Data standards are embedded in new ICT contracts?

Make sure that government lawyers are alert to this!

Setting Open Data standards

1. What is the best way to achieve compliance on high and common standards to allow usability and interoperability?

A Code of Practice based on the Public Data Principles would be valuable (& akin to that of the UK Statistics Authority).

The Government Statistical Service and the UK Statistics Authority should play a leading role in spreading good practice.

We support the publication of imperfect data, but it is vital that it is accompanied by explanatory metadata.

2. Is there a role for government to establish consistent standards for collecting user experience across public services?

We imagine that this refers to things like collecting the levels of user satisfaction with public services – in which case standards will be beneficial. Turning to the use of non-personal datasets, seeking user engagement and feedback is not easy, but needs to be pursued.

3. Should we consider a scheme for accreditation of information intermediaries, and if so how might that best work?

No! It would divert resources from the top priority of getting data out, and there's a risk of resellers creating barriers between suppliers and end-users. The role if information companies should be to take public data and compile value-added datasets from it and from other information.

Corporate and personal responsibility

1. How would we ensure that public service providers in their day to day decision-making honour a commitment to Open Data, while respecting privacy and security considerations.

This must come from the top: providers must be made aware that the minister & top management want it, again backed by appropriate powers for the Information Commissioner.

2. What could personal responsibility at Board-level do to ensure the right to data is being met include? Should the same person be responsible for ensuring that personal data is properly protected and that privacy issues are met?

The Caldicott approach of having one senior person responsibility for both confidentiality and data sharing sounds much better than split (& conflicting) responsibility.

3. Would we need to have a sanctions framework to enforce a right to data?

Yes, a regulator must have power.

4. What other sectors would benefit from having a dedicated Sector Transparency Board?

This sounds worth pushing further, but must be driven by users' needs, rather than the supplier organisations. Broad categories such as Land & Property (addresses, maps, etc.), and Social Statistics (Census, social change, etc.) spring to mind.

Meaningful Open Data

1. How should public services make use of data inventories? What is the optimal way to develop and operate this?

We favour public service data suppliers creating inventories, but these must be designed to help end-users (including occasional / new users) get started, rather than being faced at the outset with overwhelming detail.

2. How should data be prioritised for inclusion in an inventory? How is value to be established?

Again, the extent of potential use, and hence value. This isn't perfect, but it should be possible to roughly grade topics / datasets from very popular through to small minority interest.

3. In what areas would you expect government to collect and publish data routinely?

Please see the table on page 2 above.

4. What data is collected “unnecessarily”? How should these datasets be identified? Should collection be stopped?

Experience in the GSS shows that this isn't easy, but monitoring usage / downloads through data.gov.uk could help form a view.

5. Should the data that government releases always be of high quality? How do we define quality? To what extent should public service providers “polish” the data they publish, if at all?

This is a vital issue. Members of DUG (and, we are sure, most commercial companies) would rather have reasonable data now, rather than wait months for marginal improvements. The corollary is that such datasets should be accompanied by informative metadata. Datasets have different value according to the project in hand, and we are very doubtful about designation, or attributing kite marks.

“Quality” covers different things e.g. whether the data definition is appropriate for proposed use, whether coverage is sufficiently comprehensive, extent of errors, whether format is right for ease of use. A particular dataset may have more than one potential use – it may be ideal for some uses but also useful with reservations for others. Hence need for appropriate metadata – this, and upgrading formats, is where effort should be directed. Note that providing metadata that is helpful, succinct and readable is not a trivial task and one that will probably need some training.

Government sets the example

1. How should government approach the release of existing data for policy and research purposes: should this be held in a central portal or held on departmental portals?

Again, it's important to look at this through the eyes of external users, most of whom look at "government" as an undifferentiated source of data. It's really helpful to have one single starting point for information about government data, but this should then guide users to other sites where they can actually download the data.

2. What factors should inform prioritisation of datasets for publication, at national, local or sector level?

Again, the extent of potential use, and hence value, and also having data at the lowest possible level (e.g. Output Areas for statistics in most cases, or even postcode for non-sensitive data). Much potential value is destroyed if information is only available at Local Authority level or above.

3. Which is more important: for government to prioritise publishing a broader set of data, or existing data at a more detailed level?

The two top priorities must be to publish existing data at a more detailed level, and to make free some of the data that is currently charged for – both are quick wins. A change of mind-set would also be accelerated by publishing new datasets.

Innovation with Open Data

1. Is there a role for government to stimulate innovation in the use of Open Data? If so, what is the best way to achieve this?

"The best way to tap into the UK's tradition of creativity and invention is to give that data away" – and to accompany it with accurate metadata.

The second way is to have and publicise a single easily accessed starting point for information about government data (which may then guide users to other sites where they can download the data). www.data.gov.uk is a good start, but needs further development.

Keith Dugmore

Cabinet Office – Open Data Consultation – DUG views – Final

8 October 2011