

“Making Open Data Real: A Public Consultation”

Comments from Simon Briscoe and Toby White, Timetric, London

Introduction

Timetric* is an internet startup company. We build economic research systems by aggregating the world's interesting public and proprietary data - and then let anyone visualise, analyse, and share it, all through a web browser. The data, graphs, and analysis we produce are distributed worldwide; we work closely with the media, and graphs produced using Timetric appear in major international newspapers. We are a company that would not exist were it not for the developments in data access and technology of the last decade. It follows that we support the principle of open data and are confident that it can more than deliver the benefits set out in the consultation document.

We entirely endorse the sentiment in para 4.3, namely that open data might be the most powerful lever of 21st century public policy. The scope to enhance efficiency, improve accountability and boost the democratic process is almost limitless. Such support is dependent on datasets being provided in a way that they can be used and provided free to the user.

One step that could be taken by government that would do so much to enhance the use of data and engender a positive attitude towards it is to publish data that contributes to or lies behind policy decisions. A high profile success story already doing this is the Office for Budget Responsibility which publishes the data used in its economic forecasts in a way that the Treasury never did. It is important to show that this information can be for anyone and everyone.

The open data initiative in government must encourage the Government Statistical Service under the National Statistician and the UK Statistics Authority to be actively involved. Most government statisticians have shown little interest in openness (either in terms of revealing their activities or the publication of data) which is a shame as they have many of the skills that would allow them to make a significant contribution to the nation's well-being in the decades ahead.

Detailed response

The following sets out responses to the questions for consultation.

1. Glossary of key terms

1. Do the definitions of the key terms go far enough or too far?

Dataset definition should include words such as “transaction logs” and “counts”, in effect any electronic record held in a public authority. “Statistics” should be explicitly mentioned under “information”. “Open data” cannot include data which is not free – there should be a stronger assumption of free publication.

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

The presumption should be that the dataset is open. The only limitation is the need to maintain statistical confidentiality. The UKSA has already established this concept for government statistics and it can be extended to data more generally.

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?

The fear is that government departments that are unwilling to embrace open data will use cost as a barrier. This is a well-worn route for refusing data release already, under Fol or more generally. We have offered to advise departments on how to release data cheaply and present it nicely – and even offered to do the work for no cost – but have been turned down. There are some imaginative and committed individuals in the civil service but a culture change is required in those government departments with sleepy IT divisions. Sometimes, so-called innovations (such as the new DNS website launched in August 2011) actually represent a backwards step for users. Government departments alone cannot be allowed to judge in secret what is value for money – there needs to be a system to challenge any such opinions. Departments should set out their plans showing how long it will take to make all data freely open to facilitate debate. Once all data is open, there will be no more requests for data, delivering a cost saving in the long run.

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?

All public sector organisations are in the range unless there is reason for them to be excluded. This would include quangos. There are already a number of relevant lists eg that showing organisations covered by Fol. If there is data, it should be open regardless of service.

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

The starting point must be the publication of a catalogue of data existing in the department. Once that is open, other players can be involved. Departments will learn quite quickly which bits of their records are of interest. Checks and balances will then play out. The role played by ministers will be crucial, giving instruction or applying pressure accordingly. Parliamentary or other scrutiny will provide a means of encouraging those who are or are not responding effectively, and ultimately, sanction will need to be in place, with the Information Commissioner (or others) being provided with appropriate powers.

Under 5.1, to clarify, “non-personal” is taken to mean anonymised. Many of the data sets that need to be open are “personal”, for example tax records, but need to be anonymised, by removing names and showing the local authority or postcode of an individual not their actual address, prior to release. We repeat our thoughts from (1) above about widening the scope here. An additional section on government surveys should be added – if a survey is conducted (in a statistics team or not) the full results should be published not just the headlines that the government chooses to publish. (This is mentioned in 6.1)

Para 6.2 defines the coverage too narrowly – it cannot be just “public services” but must include data in the public sector too.

Para 6.11 is endorsed.

Section 7 – Six opportunities. An important opportunity is missing – improved public sector policy. If all the Labour Force Survey data, for example, were released we’d have better informed economic policy (say on the graduate jobs market, or the need for geographic mobility to find work). Linked to this is a boost to the democratic process, more engagement, a better informed media and greater understanding.

An Enhanced Right to Data

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

We would endorse all suggestions under 8.6.

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, and please refer back to our response to item (5) above.

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

Yes, and any attempts to tighten the measures should be investigated carefully.

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?

It is probably more costly to keep data secure than to make it open, so in this respect open data could save money in the long run. Departments should be required to explain why datasets that are, in their view, too expensive to release are in fact too expensive. Their estimates could then be discussed and challenged. An approach might be to ensure that all data produced from now is in appropriate formats, with the re-formatting, where needed, of past data dealt with on a prioritised basis. Prioritisation based on user demand would provide a basis for proportionality. Civil servants should, where appropriate, use the public facing databases and not their own internal systems – that would save on IT spend. If internal procedures mirrored what was published externally, there would be no extra cost to making data open.

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

Those involved in drawing up and agreeing contracts should be made fully aware of the need – is a process of certifying contracts that meet policy requirements possible?

Setting Open Data standards

We endorse the comments in 8.8.

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

Such standards are desirable but not readily and widely achievable. Do not let them delay publication – the producers of data that are popular and widely used will be under pressure to improve quality.

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

This is not pressing.

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.

It is vital that there is top-level political support, that top-level management champion these commitments and that those on lower levels recognise their responsibilities.

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

Given a right exists then a sanctions framework is a necessary requirement in support of that right.

Meaningful Open Data

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

Inventories will be vital in providing users with a means of understanding the data available. Once published many more people will be excited by the opportunities (that they are currently unaware of). Without them, far less will happen.

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

An inventory is all inclusive so issues of prioritisation do not arise. It may be that data assets should be prioritised for the completeness of their description within the inventory. Prioritisation could take place on grounds of user demand, potential for economic or social impact, or for supporting public scrutiny.

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

If the objectives of Open Data are realised then all data, not otherwise excluded for publication on grounds of disclosure risk, are to be published routinely. The distinction between collection and publication should weaken over time – if data are collected by an organisation and used by it, the data are in a sense “published”, if internally. It is just a small step to “publish” the data externally too.

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

We suspect that the savings from doing away with the data collected “unnecessarily” are modest compared to the “waste” associated with collecting data but not publishing it! The gains from setting free the data that are collected but not published are relatively huge. Departments will be reviewing their spending as part of the spending reviews.

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?

All data should be published. Very poor quality data can be meaningless and misleading, but if the data is used within government then it should be released for use outside government. Quality will be driven up following wider analysis of it. The focus on quality should be ensuring that the data is easy to access and use.

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?

Departments can publish on their own sites – it needs to be well signposted – and on data.gov.uk. Both should be encouraged as potential users can legitimately come from different starting points. It is not hard to publish data twice, especially with hyper linking.

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

Once a catalogue of potential data releases is published, the relevant department can start to publish on its own set of priorities, allowing that to be amended by outside demands.

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

The priorities are several: to publish existing data at a more detailed level, publish some new data and to make free-of-charge some of the data that is currently charged. It will often be different people publishing these sets so there is no obvious conflict for a department's resources.

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 consultation documents explains succinctly: "The best way to tap into the UK's tradition of creativity and invention is to give that data away". It is likely that the data will be better, more swiftly used if there is appropriate metadata and if it is easily findable and usable.

It is important that the government does not view the benefits of open data too narrowly in terms of directly creating a marketplace, new companies or jobs. Open data should be viewed as part of the nation's infrastructure – vital for efficient and effective modern life. We do not value clean water from taps because it creates new products or jobs but because it enables the rest of our activities to be carried out without personal sickness/illness etc. The government does not build roads (just) to create jobs in service stations but to make the economy more efficient and mobile.

* <http://timetric.com/>

Timetric was delighted to be one of the organisations invited to present a visualisation of the £25k government spending data when it was released last November: <http://byline.timetric.com/2010/11/19/central-government-spending-where-your-money-really-goes/>