



Making Open Data Real: A Public Consultation

Response from getstats, the campaign for statistical literacy

Getstats is the Royal Statistical Society's campaign *to promote numeracy*. We want to see the public better able to deal with risk, probability and quantity at work, at home and in our lives together as citizens. The Nuffield Foundation is supporting us. The campaign is addressing the media, employers, government, teachers (in schools and in higher education) and the third sector. Projects include the 'intelligent board' – how to help school governors and charity non-executives better to handle data.

Main point

Getstats agrees new and richer flows of data from organisations 'in the public space' could enrich democracy and might improve effectiveness and efficiency. More public knowledge (one definition of 'transparency') could stimulate debate about services and money, increase vigilance and arm scrutineers.

But more and better data will not in and of itself bring more accountability or improve services. We must not elide *volume* of information with better decision making. Data must become information: it must be grasped and absorbed. Information has then to be applied. Accountability and public satisfaction could move together in a virtuous circle, provided the public understands the data proffered; provided those releasing the data themselves understand it and its potential; provided its quality and accuracy are guaranteed.

Getstats believes Open Data prompts questions about *public capacity*. The government's response to proposed changes in the [school curriculum](#) allowing many more young people aged over 16 to continue studying maths and stats shows the government itself accepts the public need to be better equipped. Open Data abuts the contention that those leaving education have to be better prepared to deal with data and numbers, for their own sake as employees as well as in their lives as citizens and family builders (dealing with energy tariffs, insurance, pensions and

broadband offers). Open Data links with moves to improve the quantitative skills of university graduates, including those in [non-STEM subjects](#)

As important as the volume of data are *presentation* and ‘visualization’, the discipline of making data more intelligible. In the jargon this means paying attention to metadata and data polishing. It puts emphasis on intermediaries to help the public make sense of data.

Responses to the questions in the consultation

1. Glossary of key terms

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

No. Statisticians and academics are fond of the term ‘metadata’. This directs attention to the explanatory material that ought to accompany data release. [Professor Allan Brimicombe](#) says another missing term is *narrative*. What the public want is data to tell a story about ...the performance of schools, crime in their area and so on. Open Data needs to look at who writes and who puts out these stories.

Another key term is *visualization* – covering the many ways in which data, especially quantitative data, can be projected, for example exploiting the graphical resources of the web.

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

Data release should anticipate the sense the public will make of what is presented and how they might use data. Each department and agency should subject itself to a ‘data challenge’: is the information intelligible? The Audit Commission, thinking about local government, [said](#) ‘translating data into information that is fit for public consumption requires good analysis and interpretation, which is lacking in many councils’.

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 question does not capture the dynamism and spirit of opportunity and innovation that ought to accompany data release. Departments and agencies should relish the chance to share their work (knowledge) with the public and make explicit efforts to present it in ways the public can grasp. The value for money of data release has to be denominated in terms of accomplishing the organisation's wider public purpose, and be accommodated in its notional or actual budget for accountability.

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?

The public tend not to distinguish whether a service provider is public, non-profit or private, though they need to know how it is paid for and how it accounts. A rule of thumb for the application of Open Data is the ratio of public support to turnover (including implicit public support): any positive figure would tip the organisation into the category where Open Data applies.

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

We want a culture in which elected representatives and service deliverers feel open data accomplishes their purposes. Open data should not become a stick with public organisations are beaten, by emphasizing the way data might be used to punish or find defects; instead, it should be celebrated as the basis for 'co-producing' services and engaging the public.

An Enhanced Right to Data

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

We need incentives and awards celebrating data release and data sharing. **Getstats** would be happy to join in sponsoring challenges and competitions that recognised and rewarded good practice.

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?

Probably not. Around the Information Commissioner and Data Protection Act has grown a negative, even punitive culture. Instead of a (static) culture of rights, public organisations should make a dynamic commitment to data collection, handling and release. We could draw on past efforts to identify and praise organisations doing well to account for themselves in the broadest sense, including data sharing. Among earlier schemes worth note were the annual public reporting and accountability awards promoted by the Chartered Institute of Public Finance and Accountancy, with the support of PricewaterhouseCoopers.

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

The question assumes there is a problem when there is none. The government should consider mounting -- in collaboration with the research councils, the Office for National Statistics and others (including **getstats**) -- a campaign to counter the scaremongering that goes on about data use by public bodies, especially those concerned with the advancement of knowledge. The public should be encouraged to view two-way sharing data as beneficial (economically and cognitively). Data sharing can save money and lead to better policies. The apparatus of control through the Information Commissioner and the application of the Data Protection Act should be filleted and prevented from blocking for example the re-use of data collected by public organisations and data sharing between public bodies.

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?

Open data can lead to improved organisational performance and stronger relations between the public, as citizens and service consumers, and providing bodies. Therefore any additional costs associated with data release and data sharing should be regarded as investment. The key link is between more openness and more accuracy. As the National Health Service is reorganised, its data flows have to become more reliable: the risks and incentives associated with open data could place a new premium on data accuracy.

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

N/A

Setting Open Data standards

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

The government should find out how the public are using the data already released (for example on local authority spending) and consider establishing a centre of excellence (which might be based at an existing public body) on 'usability'. The Alliance for Useful Evidence that is being put together by Nesta, the Economic and Social Research Council and others could play a part here.

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

Yes. We need to establish thresholds of intelligibility. What do people understand by the language in which official data is couched? What levels of quantitative capacity do they bring to bear?

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

Yes. But the best information intermediaries are public bodies themselves. They should anticipate how data is going to be received and used and tailor presentation accordingly. The value of invigilators of the quality of public data has already been proven. The UK Statistics Authority, Full Fact, Straight Statistics and similar organisations have done good work in identifying the misuse of statistics and discussed both the timing and quality of official releases. Because independence is going to be a valued attribute of any organisation subjecting official releases to scrutiny or criticism, it will best be situated at arm's length from the government. The government might consider endowing a non-profit organisation to do this work.

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.

Open Data should be part and parcel of performance and monitored accordingly. The National Audit Office could play a more active role in assessing the data environments of the organisations which it audits or inspects for value for money.

The abolition of the Audit Commission has created a gap in the local public space. Government's role includes identifying and extolling good practice, which includes data and information handling in the round – i.e. the ways in which information is collected from the public as well as how it is passed out. Such government bodies as the Driver and Vehicle Licensing Agency, the Driving Standards Agency and HM Land Registry – all with active customer interfaces -- have made commendable efforts to open up operations and finance to public view, and already release large quantities of data.

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?

Open Data should not be 'ghettoised', which is a risk if a named executive is given responsibility. Data culture should be a board item, with responsibility diffused among non-executive and executive directors. Non-executives in particular should constantly be putting themselves in the place of the public and assessing the intelligibility of data flows. Too often, however, NEDs do not do a good job at ensuring the data flowing to boards is adequate or intelligible.

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

Open Data should be characteristic of good public management. Its value lies in interaction between public organisation and public and 'rights' could ossify what will be a dynamic and evolving relationship.

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

The danger with 'sectors' is that they would replicate Whitehall's 'silo-based' division of labour, missing the fact that locally services need to be joined up for effective delivery not split into sectors.

Meaningful Open Data

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

Data inventories are probably best put together at a scale bigger than that of the individual organisation, since public organisations a) share common data sets and b) collect similar or the same data from the public.

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

The simple test is: is the data necessary for achieving the organisation's stated public purpose.

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

The UK needs a data strategy. One of the missing ingredients of the Open Data initiative has been that – preparing a comprehensive analysis of what the state (and its various dependencies, including private firms) needs to know. Again, this is a dynamic conception. The state needs to anticipate knowledge needs for future years and conduct studies and data interrogations with the population of the future in mind.

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

The contours of the state and public services change and with them the ‘cognitive’ bases of government. It follows that some data sets will be anachronistic and should be subject to periodical review.

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 critical question. Put the same question but substitute ‘management information’ for data. Would any self-respecting board calmly say we don’t mind if performance data is dubious? The National Audit Office plays a role in certifying the quality of financial information within public bodies and departments; perhaps this could be extended to include information at large.

Data labelling is important. Polishing data costs money and takes time. ‘Quick and dirty’ data may do, on occasion. But it needs to be identified as more or less reliable. It would not be hard to put together a ‘grid’ attesting to the quality of data, formed from the professional opinions of statisticians (the Royal Statistical Society and the Office for National Statistics might play a part), and by the views of those involved in assembling and processing data for government (chief scientific advisers in Whitehall departments, networks of analysts).

The public are entitled to see an assessment of the reliability and accuracy of data presented to them. They deserve, too, some account of the *significance* of data. Low quality data can be significant just as high quality material can be of trivial importance. This returns to the question discussed above: those who release data should be duty bound to comment on its worth – metadata matters as much as data.

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?

The question of departmental vs central portal is less pressing than putting together a data strategy. A starting point is assessing government's knowledge needs. The strategy would also embrace release procedures and archiving (including the policies of the British Library, the National Archives and academic deposits supported by the Economic and Social Research Council). Storage protocols, access and search engines would be part of this. Much data is held and is subject to release by government 'outside Whitehall', in local authorities and arm's length bodies. Their release plans might be autonomous, but they could be required to observe templates written as part of a national data strategy.

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

No general answer can be given. Organisations should be allowed to prioritise datasets according to their business plans.

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

See the answer to 2 above

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 government needs a 'clever centre' for Open Data, staffed in part by people who understand the specifics of departments and their data economies and government away from the centre. A precondition for innovating in Open Data is, to repeat, minimum levels of public understanding, both of the data people share [to] government and [from] government.

Getstats at the Royal Statistical Society
12 Errol Street
London EC1Y 8LX
Contact: David Walker, Director
Debra Hurcomb, Project Manager

UNCLASSIFIED

UNCLASSIFIED