

Open Data Consultation
Transparency Team
Efficiency and Reform Group
Cabinet Office
1 Horse Guards Road
London SW1A 2HQ

02 November 2011

Dear Sir/Madam,

OPEN DATA CONSULTATION

Ipsos MORI sees government collect a huge amount of data about individuals' quality of life, the delivery and standard of public services every year – and yet much of this is only used partially, and barely ever openly released in full.

We strongly agree that there should be a presumption in favour of full, open publication, with a case having to be made for this not to happen. This is important even in cases where the immediate benefits are not obvious, because there are numerous examples of where opening up data has led to unexpected positive outcomes. However a key sign of success will remain professionals (in government and outside) using it to make decisions. We strongly believe that government departments need to work to published common and explicit standards on openness on data and annually self report against these standards.

Below we list our observations in more detail on some of the issues.

An enhanced right to data

1) Data open by default

The advantage is this would ensure open data becomes a part of every contract with a data provider/gatherer and as such measures would be put in place from the start to ensure data can be anonymised quickly and effectively. Certainly routine performance data needs to be available in more disaggregated forms, for further analysis by others to produce much more public value than is possible with hard pressed small central teams in individual departments and services looking at it, by allowing a much larger group of external organisations to use the information easily and effectively.

Of course, we need to be aware of potentially perverse consequences. In particular, when evaluating new programmes, data is collected and this data may show areas for improvement. Currently this data would potentially remain unpublished until the point at which a final decision is made on the future of the programme. If departments are pushed to publish everything too quickly they may instead decide not to commission at all, and this could impede genuine experimentation and innovation (Apple certainly don't provide running commentary on all their experiments and

new products). A balance therefore needs to be struck. Your point 3) on an independent body seeks to address this.

2) Requirement to pro actively publish data

We strongly support this and feel the duty to make data available needs to be strengthened (although there will be some cost implications). Many Departments do publish their data, although often this process is very slow, and does not automatically lead to the data being available in a usable form. In particular, for adhoc data collection, they may typically not publish until a report is finalised, and this can take many months more than a similar report drafted for private sector organisations.

Also, it is vital to demand very clear and specific format in which data should be published. There have been many useful and important advances in open source tools to extract raw data and make them usable, and the increased availability of data will only further this. From a survey research perspective, the labelling and data dictionary approaches are particularly important, as even slight changes to question wordings can give very different answers: having this in an accessible, searchable and systematic format will be key to avoiding misinterpretation. There could be significant improvements to existing approaches to this: for example, Data.gov has lots of data but it is difficult to know where to start when seeking specific information.

We know from public research that people find 3 levels of information helpful: a top-line pictorial summary which enables them to view the data in a comprehensible way (crime map/pie chart/diagrams etc), secondly a verbal description of the key points illustrated in the diagram (for those who relate better to verbal rather than pictorial descriptions), and finally, raw data for those who want to dig deeper.

Alongside the publication of data it is vital to also publish technical notes and advice on how to use the data. For example, the Citizenship survey had been designed to produce findings reliable at a regional (ie Government Office Region) level, but not a local authority level. As such, even if local authority data was appended, the sampling approach means that even what looks like a relatively robust sample which could be used in analysis (ie over 100 interviews in a local authority) the sampling approach necessarily means the data is not reliable at that level.

Obviously a common concern is that such data could then be unknowingly misused to report on local authorities which would be misleading and potentially unhelpful. Therefore, consideration needs to be given to the costs vs benefits of producing data which is analyzable at local levels, and government will have to consider how it would police or sanction misleading use of publicly collected data by outside bodies.

3) An enhanced right of challenge

See point 1

4) Increasing cost limits under FoIA

As a supplier of data to government, if we know data is to be published from the outset we can produce necessary anonymised data sets with no or very minor cost implications.

6) Statutory time limits for internal reviews

20 days is sufficient for most data sets to be prepared for publication in our view, but there will be exceptions for more complex studies.

7) Procurement rules to ensure data collected is stored in systems which minimise the cost and difficulty of publishing data online

We support the introduction of these rules as part of procurement, as long as they are not over-specified. For example, most research data is typically not difficult to extract/put online. There are however tools which can make the data more accessible.

For example, tools such as the GP Patient Survey portal <http://www.gp-patient.co.uk/results/> enable those interested to interrogate the data contained within the survey without having to operate analysis packages. Websites such as <http://unistats.direct.gov.uk/> provide data from a large scale survey (in this case the National Student Survey) which are integrated with other data sources (e.g. number of students at each university) to give users a rounded picture. These bespoke tools were developed by Ipsos MORI and partners in conjunction with the commissioners of the research, in order to make the data accessible.

Of course, it is likely that other tools developed by third parties not involved in the original data collection, will emerge as part of these changes. For example <http://heatmap.egovtoolkit.co.uk/> enables the visualisation of Place Survey data. Evidence suggests most people do not like, and cannot easily interpret, large tables full of data. Solutions such as these are therefore helpful in making data more accessible to the public, although in order to be used people need to know about them. We would also expect organisations like Google to take raw data and package them for the public.

Currently the more standard approach is to upload the original data-sets (to the Data Archive) alongside an interpretative report (on Departments' websites). Major surveys such as the Citizenship and Place Survey tend to have a time lag, but are then frequently used by academics and organisations such as Ipsos MORI to explore further hypotheses, but this whole area of activity needs reviewing. And while the type of webtools outlined above are most helpful for a lay audience, those wanting to do more sophisticated analysis will benefit from the ability to download the entire data-set, with as many low level geographic markers as possible. This enables us to produce reports such as our 'Frontiers' reports <http://www.ipsos-mori.com/researchpublications/publications/publication.aspx?oltemId=1330> which instead of looking at headline data, use statistical analysis to control for background characteristics which councils cannot easily change, and instead focuses on how well they are doing *given* the context within which they are operating. We (and others) would be able to do much more with this type of information if, in particular government departments looked at making lower level geographic data more routinely available than they currently do (subject to our proviso above about intelligent use of such data).

Setting Open Data Standards

1) Achieving compliance on high common standards

Our view is that public bodies should be required to publish to clear and consistent standards, at the most fine grained level possible, consequent on individual privacy. Be aware that there will be no absolutes in this process. We suggest that a published annual report by each Department on the data they have collected, and whether or not they have released it to allow external

bodies/individuals/taxpayers to assess whether they are in fact making their data public, and to allow a like for like Openness Index of Departments to be collated.

2) Role for government establishing consistent standards for collecting user experience across public services?

Government needs to end the laissez-faire approach in this area (eg in policing, and local government) if it wishes to make it possible to make genuine comparisons between services. It may decide that it does not wish to demand this. But if it wants "open data" to be used validly to compare citizen and user experiences it means insisting on the same method of collection (to prevent modal effects), and to ask identical questions (otherwise comparisons will not be valid). There also need to be rules about frequency of collection, and time from collection to public release.

3) Accreditation for information intermediaries?

We could see that a certification system, similar to that operated for data quality and information security in the survey research industry, setting out clear standards particularly over transparency of publication, interpretation of data, and privacy of individuals would be highly desirable.

Some examples of the government surveys that could be much more widely used under the above suggestions are (among many others):

- Labour Force Survey
- Family Resources Survey
- Family Expenditure Survey
- Health Survey for England
- English Housing Survey
- British Crime Survey
- Public Attitudes and Behaviours Towards the Environment
- Witness And Victim Experience Survey
- National Travel Survey
- National Employers Skills Survey
- Taking Part
- Active People
- Integrated Household Survey
- International Passenger Survey

Please contact us if you want more specific feedback. We applaud government's intentions in this area, and hope it leads to a lasting change in policy

Yours sincerely



Ben Page
Chief Executive, Ipsos MORI
