

## Open Data Consultation

### Consultation Questions

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

There is no reference to National Statistics – it would be useful to know how these fit into ‘Open Data’.

‘Datasets’ could be better defined – is this raw data? Or data as it appears within text or charts in reports or other published material? If data is already published in a table within a report, does the data also have to be published as a flat file dataset?

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

Main tests should be around confidentiality and ICT security. If data is accessible to individuals (e.g. school or personal health accounts) would summary data be available to all? Should consider public testing – before money is spent releasing data, finding out if the public want / need it would be useful. However public testing itself could add to the burdens / costs, so a consistent approach is not sensible here.

Some of the arguments for making public data open (social growth, economic growth) do not seem to be backed up with any evidence? Is there any evidence that ‘Open Data’ will improve social or economic growth? If not, then what are the main reasons for doing it?

The quality of the data should be considered before it’s made accessible to all; data of poor quality is of little value to the public.

#### **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?**

Could use similar rules then those used for FoI costing.

Another approach could be to have subscribers who pay a fee – however the fee charged will need to be considered very carefully; it is important that costs are covered, but at the same time the Government should not be making money out of this.

A threshold could be used, and any costs incurred above this threshold should be charged to the requestor. However, this adds complexity because once it’s been paid for once, it could then be easily supplied to everybody. A uniform approach would be needed for how to calculate costs incurred.

The consultation suggests that data should be released as it is rather than spending time and resources to improve data quality. This is concerning; data of poor quality is of little value to the public. Where is the line being drawn on data quality?

#### **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 policy proposals should apply to any organisation paid for by the taxpayer – this should include data collected by a non-Government body but which the Government has paid for that data to be collected (so collected on behalf of the public). It could not, however, cover data collected by a non-Government body which is then sold to the Government and other organisations, as otherwise they would not be able to sell it to anyone other than the Government (because as soon as it is sold to the Government they make it public and free to all).

Could this also cover charities? These are organisations that the public give money to without really knowing “what they are getting” for their money.

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

It could be made mandatory, like the *Statistics and Registration Service Act 2007*. Where there are breaches, an external body (the United Kingdom Statistics Authority) has the power to “name and shame”.

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

The easier and cheaper it is to make data open, the more likely organisations are to comply. Making all data available through one website (data.gov.uk) makes sense for the public, but it adds extra burdens and bureaucracy to the organisations releasing the data. Also, being allowed to publish the data in any format would reduce burdens and possibly costs.

**7. 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, as there seems to be little alternative. But the additional costs of this should be considered (and weighed against the benefits). There needs to be central guidance available to all those organisations providing open data so that a consistent approach is used. This independent body should, of course, be complying with 'Open Data' themselves. Also, if it is decided that all data should be made available through one portal (data.gov.uk), then this independent body should have responsibility for that portal as well.

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

The existing safeguards / rules are only appropriate if they are followed and upheld (and previous data security breaches have shown the public that perhaps they are not). There must be consequences for any breaches resulting in personal data not being protected.

**9. 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?**

The financial and resource burdens on these bodies should be carefully monitored and assessed. This should include considering where any existing ICT projects will have to be scrapped or amended in order to comply with 'Open Data', and the financial implications of this. Will this all be covered by the independent review of privacy and transparency? Any additional costs to the website(s) intending to house all of the data, and any independent body set up / given additional powers will need to be considered.

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

Could there be a 'standard' clause which would need to be included in all Government ICT contracts? There will need to be central guidance made available.

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

The consultation suggests that data should be released as it is rather than spending time and resources to improve data quality. This is concerning; data of poor quality is of little value to the public. Where is the line being drawn on data quality? A similar issue concerns the publishing of new data rather than old data.

In some cases, publishing data can improve the quality – i.e. where data is supplied by another body, if the data is being published it is in their best interests to ensure that the data is as accurate as it can be. Also, if an organisation is using data to support policy then if the data is having to be published, then they will (hopefully) only base policy on sound data.

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

The suggested star rating within the consultation only looks at how the data is published, not what the data is and how useful it is. Not enough consideration has been paid to the quality of the data and how it could be misunderstood or misused. A star rating on quality could be more useful to the public.

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

Would it not make sense to have all data (or links to it) available through one portal (data.gov.uk)? training should then be provided to those responsible within departments / organisations / bodies for releasing the data.

**14. 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?**

Ultimately it should be cultural, but this involves behaviour change which takes time to implement.

Could each department have an Open Data Liaison Officer? This would be a person / team in each department / organisation who has ultimate responsibility for ensuring their organisations compliance with 'Open Data'. This would be a good way of "monitoring" data which is not currently published as they should be aware of everything produced in their department. They should be closely connected to the Survey Control Liaison Officer.

**15. What could personal responsibility at Board-level do to ensure that 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?**

It seems like the two roles are very closely linked, so would make sense to be done by the same person.

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

There should be clear rules / guidance around what should be made available to the public, and these should be complied with. Where they are not complied with, then enforcement would be needed.

It should be borne in mind that the public / taxpayers ultimately pay for the data, so a good reason is needed not to let them see it.

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

Any sector which takes the public's money, without clear evidence of what that money is spent on / where that money goes, e.g. the charity sector.

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

Some of the suggestions in the consultation sound quite bureaucratic e.g. data inventories. This adds to the argument that an independent body is needed to co-ordinate Open Data. Having data inventories, which should be supplied centrally to the independent body (annually?), means that there would be a central list of all data produced by the Government. This should help prevent duplication and unnecessary data being collected.

If we are making clear what data exists, data inventories should also include details on data quality and data limitations. This should be made public as it is key for enabling the data users to understand where the data has come from and how it can and should be used.

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

All data should be included in the inventory (not prioritised), and enough time should be allowed in order for organisations to be ready to release all data publically. Data should be published in a timely manner, i.e. if data is collected as evidence for a new policy then as soon as that policy is consulted on or implemented, the data should be made available.

Value could be established by perceived importance to the public, although this is not measurable without collecting new data!

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

Wherever data is used for policy making.

National / Official Statistics.

**21. What is collected 'unnecessarily'? How should these datasets be identified? Should collection be stopped?**

Where there is duplication across Government, or where sound data collected by an independent body exists, or where data can be purchased for less money than it would cost to collect the data ourselves.

Duplication could be identified through centrally collated data inventories.

Collection should not be automatically stopped. However, as soon as an issue is identified, those who have been collecting the data should get together to work out the best approach (deciding which collection to stop; collating resources; look into alternative suppliers; etc.). Something to consider with duplication is what happens if two or more organisations are publishing contradictory data? This will involve work from both / all parties to establish the reasons behind this. Releasing contradictory data will reduce public trust in official data.

**22. 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?**

Data quality should be a very important part of ‘Open Data’. The government should only be collecting data of adequate quality, otherwise it is a waste of money, should not be made publically available, and should certainly not be used as evidence for policy decisions. However publishing data of poor quality (which is labelled as such) would encourage organisations to improve the quality of their data.

Quality should be defined using a number of factors, including how the data is collected, the sample size, the amount of missing data, etc.

The National Statistician should be consulted on to compile a definitive way of measuring data quality.

Data should be published before use by the organisation, not just before publication.

**23. 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?**

It doesn’t matter whether data held centrally or by departments, as long as there is a central portal (data.gov.uk?) which contains (at least) links to all data.

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

All datasets should be published, in a timely manner. So datasets should be published as and when they are used for policy decision making or used in reports or other publications.

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

Data should only be published at a more detailed level if the data exists at a more detailed level.

Importance should be given to ensuring that the data that is published is of adequate quality, and clearly set out and labelled in a way that is most accessible and understandable to the public.

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

No, especially if it costs anything. However there is no point putting data in the public domain if it is not being used. Can “hits” be assessed for all datasets, certainly in the early stages, to see if data is actually being used?

Consultation invites views on:

**How we might enhance a “right to data”, establishing stronger rights for individuals, businesses and other actors to obtain data from public bodies and about public services:**

By publishing anything which has been paid for by the taxpayer.

There is an argument that data should be made easily accessible for comparative purposes, for instance rather than having to look at individual local authority websites / datasets for a piece of information, they are all held centrally so it is easy to make comparisons across all local authorities.

**How we might ensure collection and publication of the most useful data:**

Data should be comparable for it to be meaningful.

Have a data quality scale.

**How far there is a role for government to stimulate enterprise and market making in the use of Open Data:**

The consultation talks about “emerging evidence which suggests innovative use of Open Data will, in the long term, create the demand needed to embed the broader culture change that we have described in the public sector, in addition to stimulating the economy, offering new services and products that will transform the public and private sector, and creating jobs and opportunities for individuals.”

There is no reference about the source / detail of this evidence. This seems very ambitious and difficult to imagine how ‘Open Data’ alone can achieve this.

Who will pay for it?

How can it create jobs at a time when people are losing their jobs and budgets are being cut?

Has consideration been made about any companies who may base their business around collecting data and disseminating data? If the government is doing it instead, their business could suffer leading to job losses, not new jobs.