

CHRIS RUSBRIDGE RESPONSE (VIA E-MAIL)

I have some comments in response to this consultation.

Section 1 (Glossary)

1Q2: What tests should be applied to decide on openness? Response: it seems appropriate to use an initial presumption of openness, then to use the FoI (and/or EIR as appropriate) exemptions and exceptions as guides. However, these are confusing and sometimes apparently contradictory, so if the whole lot could be simplified, this would be good. But they MUST be consistent with one another.

1Q3 is very poorly written as a question: "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?" Response: first, "are not judged" begs the question "by whom?". Presumably this means "If the authority judges that the costs... are not vfm, to what extent should the requester be required to pay...?". There need in the first place to be firm guidelines on the type of data that should be published, costly or not (cost in this case probably relates either to the IT implementation or the cost of cleaning up the information for release). Secondly, these are not FoI requests, for which an existing regime applies; this is about proactive publishing. The "requester" is the people as a whole, and the Government.

1Q5: "What would be appropriate mechanisms to encourage or ensure publication of data by public service providers?" A staged process starting from encouragement, through defining best practice, through to statutory requirement, would seem appropriate.

Section 8.7

8.7Q1 see response to 1Q5.

8.7Q2 The Information Commissioner should have enhanced powers to act in the area of data publication; this is especially preferable to creating a new body who would inevitably clash with ICO rulings. But regulation etc should not be part of the ICO's remit; the current structure works reasonably well in the UK, and appears to work surprisingly well in Scotland as well (with its own SIC but using the ICO for Data Protection issues).

8.7Q3 There are of course increased risks to privacy when opening up data. Data Protection principles and legislation should continue to apply. However, the ability to combine otherwise non-disclosive information and gain a high degree of re-identification would increase as more datasets are released and can be cross-referenced. I don't know the extent to which this applies now, but (with careful wording) it should be an offence to identify individuals from anonymised datasets.

8.7Q4 There will of course be resource requirements, especially at first. However, once systems are in place costs may reduce to below current costs. This is because publishing Open Data requires quality data, and will (or should) have additional quality control mechanisms through public scrutiny (the "many eyes" principle). Poor quality data is a major source of costs in organisations, so better quality data will reduce costs. Cost savings will increase as other external data sources can increasingly be relied on, innovation opportunities will occur and the benefits of Open Data can be felt within the government as well as outside.

8.7Q5 is a daft question. How do you ensure any appropriate standards are followed in ICT contracts? This becomes part of public procurement policy.

Section 8.11 on standards

8.11Q2 Yes, there is clearly a role for government in establishing consistent standards, although as has been demonstrated with current OGD initiatives, working with the wider community makes sense. But only government can require government (on our behalf) to adhere to standards. However, it would be wrong to set up a separate UK Gov standards operation. It might also not make sense to bring BSI in too much; their approach of closed standards is wrong and incompatible with the required approach (and should be fixed!).

8.11Q3 Accreditation of information intermediaries seems a crazy idea just now. Who could tell which are good ones? It would place a filter between government data and re-use, which is distinctly NOT the point. Maybe there should be an offence of data mis-representation or corruption???

Section 8.12 No comment.

Section 8.15 Data Inventories

8.15Q1 There needs to be a way that data inventories are constructed automatically as part of the publication process, rather than manually created/updated. Every dataset requires a description and some kind of schema, both of which should automatically be made available.

8.15Q2 Data should not be prioritised for inclusion in the inventory; inclusion should be automatic on publication (and automatically removed should publication be revoked). The question about value relates to the dataset, not the Inventory entry.

8.15Q5 Data should be made public. Public data should be of high quality. However, quality problems should not themselves be sufficient to prevent release, unless they are of such magnitude as to have the potential to cause real harm. In the latter case, they must either be fixed as a matter of high priority, or the dataset should cease use even internally. Quality data is really important, but we all know that data problems occur in datasets. Exposing these issues (and providing meaningful feedback mechanisms) is an excellent (if sometimes embarrassing) route to improving quality.

Section 8.17 Government example

8.17Q1 It is of less importance whether datasets are held in departmental or central portals than that they are held in stable web addresses. It should be a breach of discipline to cause datasets to change web addresses without good cause (which does NOT include departmental or other reorganisations). Don't break data URIs. If this is stated as a central requirement, it is not too hard to follow; if un-stated, the consequences are extremely expensive and dysfunctional.

8.17Q2 and Q3 Priorities for dataset publication should be led by the Public Interest Test.

Section 8.22 Innovation

8.22Q1 There are many good and inexpensive ways to encourage innovation, but as recent history has shown, having good stocks of reusable quality data is the start.

Having said all that, it is not easy to define the kinds of datasets which should be Open. It is not everything; it is not even every dataset that doesn't fall into existing FoI exemptions. I don't know too much about how civil servants work, but if I think about datasets in the University sector, they will pop up all over the place. Millions of spreadsheets exist, for instance, which qualify as datasets by most definitions. Most of these are

UNCLASSIFIED

temporary, person-oriented datasets, often exploratory, sometimes making up for deficiencies in other datasets (eg accounting data more oriented to annual reports than to managing a department or project). It would be a great mistake to publish most of these by default. So there has to be some significance, some public interest, some strong relation to mission and/or to public accountability.

--

Chris Rusbridge

UNCLASSIFIED