

Comments for “Making Open Data Real” consultation

First, I would like to applaud the UK government for seeking to increase the availability of open data. I think its analysis of the benefits that can flow from doing so are correct. The consultation document is also extremely thorough, if a little overwhelming; in particular, I have found it difficult to respond directly to the questions, even though – or maybe because – they pinpoint the key issues well.

So I will try to answer some of them by first stepping back and looking at the bigger picture for open government data. One useful way of doing that is by re-visiting the 8 Open Government Data Principles that were drawn up by a group of open government advocates a few years ago (available from https://public.resource.org/8_principles.html). They are:

“Government data shall be considered open if it is made public in a way that complies with the principles below:

1. Complete

All public data is made available. Public data is data that is not subject to valid privacy, security or privilege limitations.

2. Primary

Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.

3. Timely

Data is made available as quickly as necessary to preserve the value of the data.

4. Accessible

Data is available to the widest range of users for the widest range of purposes.

5. Machine processable

Data is reasonably structured to allow automated processing.

6. Non-discriminatory

Data is available to anyone, with no requirement of registration.

7. Non-proprietary

Data is available in a format over which no entity has exclusive control.

8. Licence-free

Data is not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.”

I think some important actions flow from these principles that relate to the key questions raised by the consultation.

Point 1 addresses issues of privacy: open data only refers to data that is not subject to valid privacy concerns. Note, though, that there can be issues that arise from aggregating such apparently impersonal data. This is a very complex area, with no easy solutions, but is worth bearing in mind, not least in terms of dealing with possible future problems, when it turns out that some anonymous data isn't.

Point 2 is in tension with that privacy requirement, since it requires information to be given in its most detailed form. In many cases, it will be possible to give raw data; in others, data will need personal elements removed; finally, there will be a class of government data that simply cannot be divorced from the people it refers to, and which therefore is not released as open data.

Points 3 to 8, taken together, mean that data needs to be available to everyone, on-demand, in open formats and with minimal licensing. Open formats means those based on open standards - those that are Royalty-Free (RF), not Fair, Reasonable and Non-Discriminatory (FRAND.)

Moving on to specific questions raised in the consultation:

An enhanced right to data: how do we establish stronger rights for individuals, businesses and other actors to obtain, use and re-use data from public service providers?

Some kind of Open Data Commissioner is probably necessary to enforce transparency – and one with real powers to sanction. The simple existence of such a person with such powers would focus the minds of people working with public data on the issues involved, and help to create a presumption of publication. In terms of ensuring open data standards are embedded in IT contracts, the key is open standards, which means RF licensing, as explained above. If that is done, it is far easier to build other open data systems of top, possibly at a later date.

Setting transparency standards: what would standards that enforce this right to data among public service providers look like?

The best way to achieve compliance on standards is to ensure that open standards are deployed across all government IT. This will make subsequent extraction of data far easier, and far more uniform, and reduce the difficulty of making data available. That, in its turn, will reduce resistance by those who are tasked with providing the data, since it will not be so onerous. Establishing consistent standards across government would help here: it would mean that expertise in opening up data could be shared, and would avoid the need to re-invent the wheel every time.

Corporate and personal responsibility: how would public service providers be held to account for delivering open data through a clear governance and leadership framework at political, organisational and individual level?

I don't see this as any different from other areas where targets are set and monitored. Standard management practices can be applied once suitable metrics for opening data are in place, including the use of sanctions where necessary. The issue then becomes how to draw up those metrics, which is certainly a non-trivial task. In terms of oversight, it would be better to have a separate structure for monitoring privacy issues, since there is clearly a tension between opening up data and protecting privacy: it would be neither workable nor fair to expect one individual to be capable of getting the balance right all the time.

Meaningful Open Data: how should we ensure collection and publication of the most useful data, through an approach that enables public service providers to understand the value of the data they hold and helps the public at large know what data is collected?

The default should be that data is available unless there are good reasons not to publish. The best way for all concerned to appreciate what can be achieved is to make data available and let people loose on it, as is already beginning to happen. A key point is that the best uses of open data are often the ones nobody has ever thought of.

As government IT systems are gradually adapted to routine data publication, many of these issues will disappear. Data quality will remain something that will need examining, but it's hard to make general statements here. Better to get data out there and find out what adjustments need to be made afterwards: “release early and release often” is the key to success.

Government sets the example: in what ways could we make the internal workings of government and the public sector as open as possible?

I think that a central portal allowing access is absolutely critical, otherwise people will always find it hard to locate material they are looking for – and impossible to come across things they weren't. However, that does not mean that data should be stored centrally, just a set of pointers to the data held throughout different government departments. This would also allow the latter flexibility in how they hold and present that data on their own sites, with consistency being applied on the central site.

As for prioritisation, I think it is important to get some datasets out there as quickly as possible so that people can start using them. Then use feedback about what else they need, or how the formats/quality can be improved to drive the next wave of data release.

Innovation with Open Data: to what extent is there a role for government to stimulate enterprise and market making in the use of open data?

The government should be doing this as a matter of course through its use of open data as part of its day-to-day operations. If it has to do this as a standalone project there is something wrong. In other words, the culture of open data needs to permeate the way government works, rather than for there to be a conscious and possibly short-term stimulation of business activity based on open data. If that doesn't happen, then much of the laudable efforts to release this data will be in vain.

Glyn Moody