

Making Open Data Real: A Public Consultation

A response by the Open Source Consortium (OSC)
26 October 2011

Introduction

We encourage Cabinet Office to link this consultation to other areas of government ICT policy to ensure the initiative achieves the desired outcomes.

We were particularly impressed by the intent expressed by the Minister for the Cabinet Office both in the forward to “Making Open Data Real” (MODR) and in a radio programme discussing the initiative¹ in which he acknowledged that MODR was about “letting go” and that the outcome was unpredictable but would lead to greater transparency and accountability.

As with previous work in this area “letting go” enables the interested and the innovative to do things that might not otherwise happen.

This approach was taken up by the Commercial Secretary to the Treasury² in a recent visit to the Tech City start-up hub in London, for a discussion on how the government’s Open Data strategy can help technology entrepreneurs.

“Open Data is a key pillar to our technology state and this discussion will play a critical role in understanding the needs of the Tech City community on how we can best establish a culture of openness and transparency in public services.”

Having considered the underpinning process, apparently relevant related government policies and other potentially relevant but seemingly unaddressed material along with confusing official replies to requests for clarification

OSC is concerned that the consultation is fundamentally flawed from the outset.

1 Radio 4, 4 September 2011, “Beyond Westminster: The Data Tsunami”

2 <http://www.v3.co.uk/v3-uk/news/2108547/lord-sassoon-visits-tech-city-talk>

The consultation process

It is not clear that the consultation is well thought through, as indicated by the following, in no particular order:

- despite being issued in the middle of the summer holiday period, the deadline for response remained at the standard twelve weeks, despite guideline advice to the contrary
- there is no indication that there are any plans to publish the responses to this consultation
- It suggests that the outcome of this consultation will be a White Paper. Prior to any White Paper it is more usual to publish a Green Paper. Moreover a White Paper is a precursor to legislation and it is difficult to envisage what legislation is envisaged because none is discussed in the consultation.
- There is no Impact Assessment of the consequences of the consultation

We sought clarification with a Cabinet Office official with responsibility for consultations, concerning difficulty understanding exactly how many questions are being asked, specifically whether the questions identified in Section 8 "Policy Challenge Questions" are rhetorical or included as part of the consultation.

The reply stated:

"I am sorry that you have not found the consultation particularly easy to follow or clear. The Consultation questions to be answered are those set out in Section 8 on page 22. I agree that it is usual practice to summarise the questions at either the beginning or the end of a Consultation document but essentially it is the decision of the author how they choose to present the information"

Which highlighted the very difficulty identified for which clarification were sought. There are "questions for consultation" on pages 6, 25, 28, 30, 31-2, 33-34 and 36. A further request for clarification of the answer to the original enquiry remains unaddressed.

Open Standards

Annex 2 of MODR provides that:

“Public data will be published using open standards, and following relevant recommendations of the World Wide Web Consortium.”

However, there are two problems with this statement:

- is “public data” the same as “open data” as defined in the beginning of the consultation document?
- government does not have a definition of open standards³. It de-scoped the definition published in January 2011 to a draft and invited comment using a survey.

This survey, conducted in March 2011, was:

- erroneously described as a consultation in response to an FOI request⁴
- unhelpfully described both as a “informal consultation” and a “crowd-sourcing consultation” in the government ICT strategic implementation plan (ICT-SIP)⁵

The value of this survey might best be determined by reference to an analysis of the ICT-SIP by a respected IT commentator:

“Too much of the Cabinet Office’s Implementation Plan is given over to what has been achieved, such as the boast that “an informal consultation to crowd source feedback on Open Standards has taken place...” [who cares?]⁶

There are no obvious links from MODR to the Government ICT strategy, which unhelpfully uses the undefined term “common standards”⁷

3 Evidence to PASC

<http://www.publications.parliament.uk/pa/cm201011/cmselect/cmpubadm/writev/goodgovit/it63.htm>

4 http://www.whatdotheyknow.com/request/process_leading_to_publication_o

5 <http://www.cabinetoffice.gov.uk/resource-library/uk-government-ict-strategy-resources>

6 <http://ukcampaign4change.com/2011/10/24/governments-new-ict-plan-the-good-bad-and-whats-needed/>

7 Evidence to PASC supra

Further confusion is created by statements in September 2011 from the deputy government CIO in which he stated that

“[the government] is set to publish 10-12 key open technology standards including web and document formats that it will require all public sector bodies to use”⁸

“The initial set will not be set in concrete, but form part of an evolutionary process”

These statements are further confused by the ICT-SIP as it splits the work into two work streams in its timetable for delivery with its milestones:

- First release of a draft suite of mandatory Open Technical Standards published (December 2011)
- A set of open standards for data adoption established and progressed by government departments, driven by the Open Standards Board (June 2012)

Further rather than all applying to public sector bodies, the ICT-SIP indicates that it:

“provides a reference for central government” and “is not mandatory outside central government”

The ICT-SIP does not address the absence of a government definition of an open standard. Further the term “mandatory” is only applied to “technical open standards” and not to documents. The key milestones only provide:

Publish the findings from the crowd-sourcing consultation and the approach to identifying and mandating a set of technical standards (October 2011)

which to date has not been met and in any event is not helpful for MODR which has a closing date of 27 October 2011 for responses.

⁸ <http://www.ukauthority.com/NewsArticle/tabid/64/Default.aspx?id=3303>

It is unclear what how any of the foregoing is addressed by the ICT-SIP risk mitigation strategy:

The risk of protracted debates on architectures and standards result in long lead times or “academic argument”.

which will be mitigated by

Strong leadership and governance processes with efficient engagement channels.

It does not seem that the phrase “academic argument” is meant positively, why it was included in the risk mitigation strategy nor what is meant by “efficient engagement channels” as part of the mitigation.

The earliest reference to a government identified need for a definition for open standards that OSC has been able to identify is contained in the first action plan for open source software published in 2002^{9 10}

Although the 2002 document is not easy to find, version 2 was published in 2004¹¹ updated in 2009¹² and replaced by the one published in 2010¹³

It took nearly nine years until January 2011 for government to issue a definition¹⁴ for an open standard; one which OSC and others would have found acceptable.

This definition lasted until March 2011 when the definition was reduced to the status of a draft by the Cabinet Office standards survey¹⁵ described internally in a briefing note as a “consultation” despite it not remotely meeting the requirements for a consultation as provided in the government code of practice¹⁶

9 <http://www.ukauthority.com/NewsArticle/tabid/64/Default.aspx?id=588>

10 <http://www.internationallawoffice.com/newsletters/detail.aspx?g=99327ff6-e367-49bc-8e6d-6a658b246e87>

11 <http://www.epractice.eu/en/library/281032>

12 http://webarchive.nationalarchives.gov.uk/+http://www.cabinetoffice.gov.uk/government_it/open_source.aspx

13 <http://www.cabinetoffice.gov.uk/resource-library/open-source-open-standards-and-re-use-government-action-plan>

14 http://www.cabinetoffice.gov.uk/sites/default/files/resources/PPN_3_11_Open_Standards.pdf

15 <http://www.publications.parliament.uk/pa/cm201011/cmselect/cmpubadm/writew/goodgovit/it63.htm>

16 <http://webarchive.nationalarchives.gov.uk/+http://www.bis.gov.uk/policies/better-regulation/consultation-guidance>

An FOI request suggested that there was little underpinning thinking or analysis to support the publication to any definition¹⁷. This conclusion was supported by answers to four written Parliamentary Questions (PQs) published in September 2011:

Q. "To ask the Minister for the Cabinet Office whether his Department has considered the merits of fair, reasonable and non-discriminatory (FRAND) licensing obligations in respect of procurement policy Action Note 3/11, issued on 31 January 2011."¹⁸

A. The Government require that their ICT should be built on open standards, wherever possible, to improve competition and avoid lock-in to a particular technology or supplier.

Fair, reasonable and non-discriminatory FRAND specifications may present some difficulties for the open source software development model in terms of patents and royalties. To deliver a level playing field for both open source and proprietary software, open standards are needed.

Q. To ask the Minister for the Cabinet Office what assessment he has made of the effect of procurement policy Action Note 3/11 on industries that depend on royalties¹⁹.

A. Through specifying open standards in government ICT procurement, the Government are looking to improve interoperability, maximise efficiency and reduce cost to the taxpayer.

No formal assessment has been carried out on the effects of policy Action Note 3/11 on industries that depend on royalties or on industry that requires royalty free standards to operate. We are about to commence a review of this note and Government are actively engaged with industry on the impact of Action Note 3/11 which is being considered alongside feedback from the UK Government Open Standards Survey and other consultations.

Q. To ask the Minister for the Cabinet Office whether his Department carried out a cost-benefit analysis of the procurement policy Action Note 3/11 on the use of open standards when specifying ICT requirements issued on 31 January 2011; and if he will make a statement²⁰.

A. Cabinet Office did not carry out a generic cost-benefit analysis on the policy procurement note. Cost-benefit analysis should be carried out by Departments as

17 <http://www.opensourceconsortium.org/content/view/144/90/>

18 <http://www.theyworkforyou.com/wrans/?id=2011-09-08a.70111.h&s=speaker%3A24889#g70111.q0>

19 <http://www.theyworkforyou.com/wrans/?id=2011-09-08a.70112.h&s=speaker%3A24889#g70112.q0>

20 <http://www.theyworkforyou.com/wrans/?id=2011-09-08a.70115.h&s=speaker%3A24889#g70115.q0>

part of the procurement process for each specific implementation.

Q. To ask the Minister for the Cabinet Office what estimate he has made of the cost to industry of offering software and technology products to government on a royalty free basis²¹.

A. Government have no intention of demanding the intellectual property for all ICT solutions it specifies. The proper place for discussion of these rights is during contract negotiation.

The current policy states that intellectual property relating to solutions provided by the private sector for public sector contracts should remain with the party best placed to exploit them. This ensures that, wherever sensible, business can retain their IP to use with other clients and internationally.

As outlined in the Government ICT Strategy, published in March 2011, the Government is committed to creating a level playing field for open source software for Government ICT procurement. We recognise that open source solutions present significant opportunities for improved value for money and the stimulation of a more competitive ICT environment. We are therefore taking positive action to encourage the use of open source in departments, where cost is equal to, or less than, the lifetime costs of proprietary software.

The Government will publish guidance on intellectual property related to public procurement later this year, to raise awareness in the public sector and industry.

Interoperability

MODR Annex 2 also contains the statement that:

“Public data will be published in reusable, machine-readable form [...] At the moment a lot of government information is locked into PDFs or other unprocessable formats”

PDF is a published standard²², capable of being implemented without payment of royalties and so implementable in Free and Open Source Software (F/OSS). While PDFs are far from being unprocessable (e.g., if they are outputs from office productivity suites), as MODR acknowledges it is a far from ideal format for enabling data re-use.

This highlights another unaddressed issue with open data. Standards are only one part of the requirement to successfully provide open data, the other being interoperability.

²¹ <http://www.theyworkforyou.com/wrans/?id=2011-09-09a.70113.h&s=%22intellectual+property%22#g70113.r0>

²² Currently ISO 32000:2008

In addition to using suitable standards for publishing data, attention must be given to avoiding interoperability difficulties through excessive formatting of the data presentation or worse still, by embedding code in the data file, e.g., VBA or macros in a spreadsheet or text document.

We drew this problem to the attention of the deputy government CIO in February 2011²³ who indicated that he would arrange for his team to look into the matter²⁴. However an FOI request in May 2011, finally answered in August 2011²⁵ did not provide any evidence of progress.

Economic Impact Assessment

In the absence of an economic impact assessment the decision contained in MODR as to what data is in scope and not in scope seems arbitrary.

- standards

The government has published various studies on the importance of standards to the economy at large²⁶, of which government ICT and online public services play an increasing part directly and coercively (e.g., increasing compulsion to interact online with government departments).

- interoperability

Government departments have conducted studies on interoperability which could have formed part of an initial desk study. For example the Department for Education (Dfe) “interoperability review” was published in 2010²⁷

23 <http://www.opensourceconsortium.org/content/view/131/89/>

24 in a “tweet” dated 28 February 2011

25 http://www.whatdotheyknow.com/request/improving_interoperability

26 <http://www.bis.gov.uk/policies/innovation/standardisation/economic-benefits>

27 <https://www.education.gov.uk/publications/eOrderingDownload/Interoperability%20Review%20Report%20-%20Summary.pdf>

From an available summary:²⁸

“A three tier model for interoperability was considered. The review concluded that doing nothing is not a tenable option. The current lack of interoperability is estimated to be costing in excess of £300 million per annum across education, skills and children’s services.”

An analysis of the report raise familiar issues:²⁹

I believe the approach identified in the report is the right one, but doesn't stand a chance of being supported. Why?, because it requires agreement on standards [...], underestimates interference of public sector IT suppliers who prefer proprietary rather than open solutions, and wholly underestimates the scale of the task. Success would also require a hard line from a big central government

The study was undertaken by Cap Gemini and dismisses one available interoperability standard “SIF”. This decision not taken lightly³⁰ by the UK arm of the international parent body SIFA (UK). While the licence for the standard isn't immediately obvious from its website, the specification is publicly available without any copyright notices and we may read³¹

The SIF Implementation Specification is based on the World Wide Web Consortium (W3C) endorsed Extensible Markup Language (XML) which is not linked to a specific operating system or platform.

DfE has set up “rigorous processes”³² to take forward interoperability though including a new special interest group for interoperability but which seemingly has yet to meet.

- open data

There is a plethora of national and international data that could have been used to inform a central government desk study. It could have begun with the still available study undertaken for Ordnance Survey in 1996.³³

28 <http://www.scie-socialcareonline.org.uk/profile.asp?guid=e7298caa-86eb-4fb4-985f-f0d378518fa8>

29 http://kentrustweb.org.uk/CS/community/kcc_digital_curriculum/archive/2010/10/08/education-skills-and-children-s-services-interoperability-ouch.aspx

30 <http://sif.schoolsict.net/sifa-uk/sifa-uk-responds-to-the-dfe-summary-review-about-interoperability/>

31 <http://www.sifinfo.org/uk/sif-specification.asp>

32 <http://www.education.gov.uk/escs-isb/standardsadoption/a0076257/standards-adoption-lifecycle>

33 <http://www.ordnancesurvey.co.uk/oswebsite/aboutus/reports/coopers/index.html>

The study, entitled “Economic aspects of the collection, dissemination and integration of government's geospatial information” contained sections on

- key characteristics of information and the underlying rationale for government involvement with this area of the economy;
- then current UK government policy towards the provision of and charging for information;
- a review of the then UK market for information collected by government;
- comparisons with the arrangements for provision of information in certain other countries; and
- a brief review of the key economic issues associated both with pricing by government-owned organisations and with the regulation of privatised monopolies.

as well as a helpful bibliography for further reading.

Cabinet Office could have turned to its own study conducted in 2007 containing a literature review “The Power of Information”³⁴ which clearly indicated the additionality arising from allowing others to re-use public data.

There are freely available collections of national and international studies on the economic value of open data³⁵ as well as other related studies on open access³⁶

Finally there is the absence of any discussion relating to government as significant economic actor both directly and indirectly using a competition law perspective.

34 An independent study by Ed Mayo and Tom Steinberg in conjunction with Cabinet Office.

35 http://wiki.linkedgov.org/index.php/The_economic_impact_of_open_data

36 <http://svpow.wordpress.com/2011/10/22/economics-of-open-source-publishing/>

Responses to consultation questions

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

The terms “Dataset”, and “Information” are adequately defined

The term “Open Data” contains potential ambiguities. One ambiguity could be addressed by adding a phrase such as “without a requirement to pay royalties or imposition of other obligations” from

“Data which can be freely used, re-used and redistributed by anyone”

to

“Data which can be freely used, re-used and redistributed by anyone without a requirement to pay royalties or imposition of other obligations”

However the definition of “Open Data” (unamended) is immediately superseded in this context by defining open data produced by public services as being data released using the Open Government Licence (OGL) so the purpose of its inclusion here is unclear.

The OGL emerged suddenly a thread on “uk-government-data-developers” Google group³⁷ initiated by the Directgov innovate team relating to an original posting on their blog³⁸

In September 2010 the development of the OGL was announced (without consultation). and concern were expressed on this thread relating to the compatibility of this licence with other open licences. It is not clear that the issues raised on that thread were addressed.

One of the specific but unexplained issues that the OGL apparently addresses concerns database rights, however:

- The value or purpose of this licence does not appear to have been publicly established, as the discussion in the thread on making it compatible with other licences demonstrates

³⁷ http://groups.google.com/group/uk-government-data-developers/browse_thread/thread/62d625aa4e26e27b

³⁸ <http://innovate.direct.gov.uk/blog/open-source-software-licensing>

- Does the the OGL permit the re-use envisaged in the first part of this definition (including the proposed amendment) in which case what extra functions does the OGL serve?
- Will the government clarify the role of the OGL?
- If the OGL does not permit the re-use envisaged in the first part of the definition (including the proposed amendment) in what ways does it differ?
- Regarding software (data in its own way) despite considerable objection the developers of the OGL refused to universally enable software developed by or on behalf the public sector to be freely re-usable (both under an OSI compatible licence or proprietary). Will the consultation respond specifically to this issue?

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

The presumption should be to publish. Any tests applied should reflect that presumption. Existing safeguards and their oversight should suffice.

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?

Without an economic impact assessment this question becomes impossible to answer as there is no baseline for value for money. The general principle should be to release data free of charge as the marginal cost of such release is likely to be low-to-trivial

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?

Without an economic assessment of the potential value of open data any assessment of where to set the margin is in danger of being arbitrary.

Reversing the question, why not invite providers of public services to justify not releasing data?

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

What gets measured gets gone. A reporting requirement with performance penalties or rewards will assist.

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

An economic assessment of the value of open data would baseline this presumption. As has been expressed elsewhere on other topics what is wrong with mandation from the centre if the topic

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?

One implication of this question is that public bodies might be reluctant to release data.

The solution to this problem in an environment where the presumption is to publish would appear to be one for internal governance and accountability.

If in such circumstances it becomes necessary to use FOI legislation to obtain data (not in a reserved category e.g., personal data) the issue would not be one related to FOI but failure to comply with wider government policy and a failure of internal governance and accountability

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

If existing safeguards are inadequate then these would seem require attention regardless of this consultation.

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?

An economic assessment of the value of open data would enable the implications to be properly assessed. Until such an assessment is made, individual organisations will be able to conduct a tactical or operational assessment based on insular considerations.

Such an assessment would enable an informed discussion of burden. As with programme management rather than project management, in which some projects that do not have a robust business case individually are justified by their contribution to the programme.

How will we ensure that Open Data standards are embedded in new ICT contracts? What is the best way to achieve compliance on high and common standards to allow usability and interoperability?

In the absence of a definition of open data standards this question is unanswerable. If such a definition existed it would be a matter of imposing conditions during negotiations.

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

Yes, the government should be assessing how well its services are performing.

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

If the data is set free, we are unclear of the value offered by such intermediaries

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.

Existing legislation, governance and compliance policies should already cover this situation. There are no circumstances we can envisage that are not covered by existing legal safeguards.

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?

We consider this to be a question best addressed by officials mindful of their responsibilities under existing legislation

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

This would seem to be a question regarding performance and management of organisations and officials and perhaps best addressed within the public sector rather than have external parties attempt to suggest organisational development or change proposals.

Set in a context of a presumption to publish, the nature of any sanctions framework would appear to be different than if there is merely a recommendation that data should be published

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

We have basis for providing an answer to this question

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

Focusing on releasing datasets and indexing such sets is the priority. This will create a richer inventory over time by due process.

Focusing on the requirements of an inventory at the expense of publication of datasets could give rise to diversionary work and delay.

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

The need for prioritisation requires context. The absence of an economic impact assessment renders the establishment of value difficult. In circumstances where there is a presumption to publish, providing the publication is timely then over time this question becomes moot.

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

Collection of data would not appear to be within the scope of this consultation.

Any data collected by government that can be released should be released. The most important requirement is that data should be released quickly to enable maximum value to be both extracted and added

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

We have no specific response to this question as it would appear to require an understanding of the purpose and value of particular datasets in the absence of context.

Anecdotal evidence relation to the disposal of datasets concerning hearing tests of BBC employees over a long period of time suggests that the value of data changes in either direction depending on external unpredictable circumstances.

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?

A central point of access would encourage consistent interfaces and data formats across departments.

However this seems to be a question of internal organisation best addressed by officials but not at the expense of timely release of the data.

An emphasis on releasing data and the publication of URIs would seem to be the real issue.

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

The factors identified in a suitable economic impact assessment

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

This question is a false dichotomy. It is not either/or. Over time, in the light of a suitable economic impact assessment, the question will self answer.

The easier the data is to access the easier it is to publish. Focusing on ensuring interoperability based on a suitable definition of open standards.