

Making Open Data Real: A Public Consultation

Public Consultation Response

October 2011

Notice

This document and its contents have been prepared and are intended solely for HM Government's information and use in relation to the public consultation on "Making Open Data Real" as published in August 2011.

This document has 11 pages including the cover.

Document history

Job number:			Document ref: Public Consultation Response			
Revision	Purpose description	Originated	Checked	Reviewed	Authorised	Date
Rev 1.0	Initial Response	JM/VS	<Initials>	<Initials>	<Initials>	24/10/11

Table of contents

Chapter	Pages
Questions for Consultation	6
1. Do the definitions of the key terms go far enough or too far?	6
2. Where a decision is being taken about whether to make a dataset open, what tests should be applied?	6
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?	6
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?	6
5. What would be appropriate mechanisms to encourage or ensure publication of data by public service providers?	6
6. How would we establish a stronger presumption in favour of publication than that which currently exists?	7
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?	7
8. Are existing safeguards to protect personal data and privacy measures adequate to regulate the Open Data agenda?	7
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?	7
10. How will we ensure that Open Data standards are embedded in new ICT contracts?	7
11. What is the best way to achieve compliance on high and common standards to allow usability and interoperability?	8
12. Is there a role for government to establish consistent standards for collecting user experience across public services?	8
13. Should we consider a scheme for accreditation of information intermediaries, and if so how might that best work?	8
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?	8
15. 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?	9
16. Would we need to have a sanctions framework to enforce a right to data?	9
17. What sectors would benefit from having a dedicated Sector Transparency Board?	9
18. How should public services make use of data inventories? What is the optimal way to develop and operate this?	9
19. How should data be prioritised for inclusion in an inventory? How is value to be established?	10
20. In what areas would you expect government to collect and publish data routinely?	10
21. What data is collected 'unnecessarily'? How should these datasets be identified? Should collection be stopped?	10
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?	10
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?	10
24. What factors should inform prioritisation of datasets for publication, at national, local or sector level?	11
25. What is more important: for government to prioritise publishing a broader set of data, or existing data at a more detailed level?	11
26. Is there a role for government to stimulate innovation in the use of Open Data? If so, what is the best way to achieve this?	11

Introduction

The authors of this document, whilst permanent employees of a private sector supplier to government, express their own personal views on the public consultation document on “Open Data” as published by HM Government in August 2011.

The views expressed in this document reflect the private opinions of individuals.

Executive Summary

The principles outlined as a part of the Open Data consultation could provide benefit to the Government and the bodies seeking to use Government information. However, the generation of a clear, coherent and consistently applied scope for the data sets that are intended to be published present a considerable challenge.

Whilst common and consistent data will support removal of barriers to entry for Small and Medium Enterprises (SMEs), publication of all data for the sake of it, in whatever format it is collated will increase the risks of unreliable analysis and will create a burden for Government with regard to cost and administration.

The proposed 'Public Data Corporation' should provide an opportunity to provide an agreed structure within which Open Data can be collated using consistent measurements and definitions and an over-arching licensing model.

To deliver this, it is essential that sufficient consultation is included with information providers, industry and other user bodies.

Having a coherent regulatory framework will assist Government and users.

Questions for Consultation

This section contains details of the questions posed in the Public consultation document entitled “Making Open Data Real: A Public Consultation” as published by HMG in August 2011.

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

The definition of the key terms, whilst providing a wide scope for interpretation, should also include definitions of that which is not subject to the Open Data approach so as to make it clear from the on-set what datasets (if any) are considered out of scope for public sector organisations. In relation to non-government bodies providing public services, information about aspects unrelated to the delivery of their public service function are not in scope, does that imply that ALL public sector data is in scope? We assume certain public sector data will be considered out of scope and a clear definition of what can be expected should be provided.

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

In our opinion, existing legislation such as the FOI or the EIR would still have a role to play with regard to the publication of data. Additionally, HMG document markings should be reviewed so as not to provide a default ‘Restricted’ or ‘Confidential’ classification. Tests with regard to the publication of salary and personal information should also continue to be applicable in certain situations where the publication of data may compromise personal or business relationships.

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?

Currently the maximum ceiling for the costs to public authorities responding to FOI requests is £450 or £600 (for central government departments) represents a considerable amount of man-time with regard to salary and business costs. Additionally, the costs incurred by private sector organisations in responding to management information (MI) requests from public authorities, or FOI requests directed at public sector organisations is not considered by the individuals making these requests.

If it is Government’s aim to establish a common data set against which all public bodies would have to provide a base data set, as outlined in the Managing Public Money guidance (MPM) it should be acknowledged that that in itself will create a burden on those bodies providing the data. Consideration should be given to establishment of a baseline data set against which data is provided for free. Any information requests outside of the defined datasets could be chargeable with a proportion of the fees payable supporting the potential Public Data Corporation (PDC) in support of their activities and costs.

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?

Whilst Government should ensure that the requirement to provide data does not create unnecessary burdens on public bodies, particularly those organisations working with small budgets. There is a view that any organisation that is in receipt of public funds should have a responsibility to account for how those funds are spent. This would include all organisations currently subject to FOIA and any other bodies in receipt of public funds. The issue would be where to draw the line, i.e. in the event of private suppliers to public bodies who engage sub-contractor organisations to deliver work, should the subcontractor also be subject to such measures.

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

Whilst it is acknowledged that the establishment of a Public Data Corporation would provide a vehicle to manage and regulate the behaviour of those organisations required to make open data available, the ICO

would seem to be the natural 'owner' of the governance and regulatory scheme that would be required to support the Open Data and Transparency proposals in the absence of a PDC.

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

The opportunities for public bodies to hide behind claims of excessive time to produce must be minimised and support through a regulatory framework that compels publication in all but the most exceptional circumstances (i.e. to protect national security, personnel information etc). If appropriate legislative and regulative measures could be established by the responsible body (be it the ICO or a new PDC vehicle) to establish industry wide reporting criteria (i.e. for education, health, defence etc). Government should also consider a 'published by default' scheme that could be written into service provider contracts to those organisations supporting public bodies.

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, so long as the organisation remains independent and is not seen to be at the sole service of government. Any other option would require the establishment of a new body and the associated costs of doing so should be taken in to consideration in the current economic climate.

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

This point need to be considered carefully. Data should only be relevant to the service being provided, anything that does not directly influence an entitlement to a service or funding should not be collated within Public bodies. Open data should not be cross-referencable between data sets from different organisations. Any information that would support an ability to cross-reference information against another data set provided by another public body must be carefully considered. Whilst the ability to cross-reference data is vital between departments (i.e. the use of NI numbers to combat benefit fraud), the ability to make such comparisons using Open Data must be removed to ensure privacy.

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 potential to create additional burdens on public bodies as a result of implementing Open Data needs to be considered carefully. A thorough assessment of how the FOIA is delivered within existing public bodies should be undertaken. The exercise should encompass those bodies where the potential for creating unnecessary burdens is greatest, and consideration should be given to this both informally and formally. I.e. the current commitment to dealing with FOIA in public bodies should be benchmarked. Whilst the need for transparency and visibility of how public funds are spent is important, it should not be delivered so as to create a substantial increase within public bodies or add to the costs they incur to ensure compliance. The only way to ensure it does not is to benchmark the existing burden on these organisations prior to the introduction of any supporting requirements under the Open Data initiative. This could be achieved through taking a phased view to establishing reasonable boundaries and limits in cost to produce.

In summary, it would not be possible to measure the impact of introducing Open Data unless there is a clear understanding of what investment is being made in complying with existing FOIA requests.

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

Government needs to consider the approach to this carefully. There are a large number of ICT contracts already in existence with considerable time left on them. Whilst introducing new Open Data standards with regard to new ICT contracts would be relatively straightforward (once the legal and regulatory hurdles have been cleared), the legacy contracts should be amended to reflect any new requirements in support of Open Data. Much like Transparency clauses have been written into existing supply contracts, the requirements could be transitioned into existing contracts through change control. Where a supplier refuses to adapt an existing contract, Government should seriously consider whether or not the response is acceptable, and where this is not the case, the potential to re-tender a particular contract in the public interest should be considered.

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

Rather than undergo a costly exercise in defining a new set of high and common data standards, government should consider identifying current areas of best practice through the benchmarking of public bodies and existing data structures. HMG should seek to agree definitions of data terminology so that all organisations subject to the requirements of Open Data have a consistent understanding of common definitions for data fields and the content within them. Misinterpretation of information will be reduced and the ability to compare and integrate cross departmental data analysis will be increased. Consideration would also need to be given to the adoption of common reference standards i.e. for GIS related information the use of common OS references (be they 6 digit or longer) would allow for accurate and comparable reporting and cross referencing of information from different organisations. Common interpretation of expenditure on commodities should be defined, i.e. inclusive of VAT and expenses, or exclusive or power consumption being reported in Mw hrs or Kw hrs etc.

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

Government provides many different public services, whilst some standards for data collation can be made consistent to allow for greater comparison of perceptions of individual services, consideration also needs to be given to specific departmentally related data in order to identify local issues and areas of weak performance across individual departments and other public bodies. There is a need for balance so as to include the benefits of service providers and public interest in the public services they receive. The public (and other interested parties) require the localised information to make decision on services and in order to support delivery of the big society. However, public service providers and government itself has separate needs for consistent data in order to make informed decisions and direct comparisons across public service providers, markets and costs. Many organisations develop manage and analyse their own user experience initiatives, there is significant cost in the localisation of these initiatives and government has an opportunity to centralise the definition of information around user experience whilst also reducing the local development costs associated with this.

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

The accreditation of information intermediaries will only work effectively if the organisation responsible for the delivery of that accreditation (1) has enforcement powers to deal with ineffective or poor performance from accredited suppliers, (2) has some influence over the maximum costs that accredited organisations charge for access to information and (3) Maintains an ability to adequately meet the needs of information providers (public sector and related service providers) and also the needs of the proposed information intermediaries. This role could potentially be fulfilled by the proposals relating to the establishment of a Public Data corporation, but consideration should also be given to using an existing organisation such as the ICO to oversee delivery.

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?

To support this objective would require an organisation with independence and impartiality so as to ensure that public bodies and public service providers are treated mutually and identically. To support this objective a set of detailed and coherent definitions and guidelines would need to be produced to clearly communicate to organisations what is considered to be private information and what would be considered as confidential with regard to protecting national security. However, in order to ensure consistency in application, it is essential that an independent body or reviewer be appointed to oversee the application of privacy and confidential data so as to ensure that public sector organisations are not using the ability to withhold data under those categorisations unduly. This role could potentially be fulfilled by the proposals relating to the establishment of a Public Data corporation, but consideration should also be given to using an existing organisation such as the ICO to oversee delivery. Departments and public service suppliers should be aligned to a common set of objectives and requirements related to the provision of Open Data, and these should be applied and monitored independently and with a consistent application of requirements, and where required penalties to support compliance.

15. 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?

Whilst Board-level accountability already exists in support of a number of initiative and requirements such as health and safety, it must be recognised that it is not necessarily these individuals dealing with the day to day requirements that supports the development and application of supporting policies within organisations. The nature and scope of information that will be required to be covered with regard to Open Data is likely to require considerable support within public bodies and other organisations service providers. Unless HMG provides considerable detail in relation to specific requirements, consistency of data sets, consistency of interpretation, file formats etc, organisations will interpret requirements differently, delegate responsibility and create multiple layers of input and ownership. Board level responsibility would make sense, but it must also be recognised that the need to comply with and provide Open Data will in itself lead to increased costs in order to provide the information. Government needs to carefully consider the impact with regard to man-days, overhead costs and additional burden on the public purse that will be created as a result.

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

Without a sanctions framework it is hard to see how the Open Data agenda would operate. There are many initiatives that have been tried and failed as a result of ineffective or limited enforcement. Public sector consumption is huge, the organisations that will be required to comply with the Open Data requirements a large, complex and varied in the consumption of products and services. The scope of Open Data is equally large, would require information analyst support and given pressures to deliver business as usual, it is hard to see how public bodies would maintain a focus on the delivery of Open Data if there was not some form of sanctions framework in place. However, any such framework would need to be consistently applied and enforced. This role could potentially be fulfilled by the proposals relating to the establishment of a Public Data corporation, but consideration should also be given to using an existing organisation such as the ICO to oversee delivery.

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

None. If, as suggested in the consultation of a Public Data Corporation, a single organisation would be responsible for overseeing data definition, collation, publication and licensing, why would there be any need for dedicated sector transparency boards? Surely, this is only creating another layer to deliberate and interpret any public data definitions. Government should not provide the opportunity for individual sectors to deliberate the confidentiality or provision of data, this should be part of the role of a PDC or the ICO as long as they have the correct legislative, sanctions and frameworks in place support to ensure public bodies comply with and are bound by the requirements of Open Data why would you need another layer to deliberate sector transparency. The PDC should be the body to hear cases for withholding data from public bodies and public sector suppliers, and should do so consistently. The only exceptions should be those stated that relate to private personal information and that withheld in the interests of National Security, but the PDC should be the ultimate policy and decision making body and should have the authority to ask for evidence in support of such requirements and as such could even contain individuals from the different sectors in support of ensuring familiarity with the data within those sectors and providing a sounding board for instances where data is withheld.

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

There is a need to establish a clear and well defined framework of data sets and date inventories that are applied consistently across public bodies and public service providers. Comparisons and meaningful analysis of data can only be achieved through the application of common definitions and measurements. The objectives behind the consultation on Open Data will only succeed if there is (1) an overarching body or organisation with responsibility for delivery (2) A clear and well defined framework of data sets (3) a body responsible for monitoring and amending data requirements going forward and (4) establishment of a Open Data 'data warehouse' to support the collation, storage and access of information in support of making it readily accessible. Failure to support the initiative appropriately will result in data discrepancies, non compliance, inappropriate understanding of definitions and deterioration of confidence in the data from users and consumers.

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

The main issue with this is establishing the baseline. What data should be made “Open” across different Public bodies and departments? For all organisations in the public sector there is a vast amount of data that is gathered and required to support the effective delivery of those organisations. There are also a number of data items that are required to support existing voluntary or statutory reporting requirements. Herein lies the problem, not all of these data sets are in the public interest, some are personal and confidential, and some are internal data sets used locally. Under the auspices of Open Data that data which relates to the efficient and effective performance of public services should be given priority as should that which serves the public interest. The main issue here the body or organisation responsible for the setting of priorities for data sets for inclusion in a data inventory. The task is large and complex, a thorough understanding of how public sector organisations and suppliers are structured, their ability to store access and deliver data sets across the public services being delivered and the ability to compel public bodies and service providers to supply data in an appropriate and timely fashion all need addressing. Only an overarching organisation such as the proposed Public Data Corporation would be able to do so effectively and even that is assuming it has the correct expertise, understanding and operational remit to do so.

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

Any organisation operated through the use of public funds should fall under the requirements to capture, store and publish open data except where this involves the publication of personal data or data that is sensitive with regard to national security. As soon as this definition starts being diluted, the impetus and effectiveness of the Open Data requirements will be called into question. The definition of personal data and particularly that withheld under the auspices of national security must be clearly defined from the on-set.

21. What data is collected ‘unnecessarily’? How should these datasets be identified? Should collection be stopped?

The collation of data for the sake of it should be managed carefully. Individual organisations should be encouraged to collect the data that is essential to the delivery of their specific service. This will ensure that only pertinent data is harvested. The ability to cross reference data based on a common identifier must be minimised except for in relation to those services where such data is meaningful i.e National Insurance numbers being cross-checked across agencies to combat benefit fraud. The purpose and use of individual data sets within public bodies needs to be questioned. An overarching body such as a “Public Data Corporation” could be the delivery agent with a remit for considering that which is required and that which is unnecessary and provide appropriate guidance and if necessary, sanctions to ensure compliance.

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?

Providers should not be allowed to ‘polish’ data. Where appropriate, commenting on data or listing assumptions etc, should be encouraged so that the reliability and accuracy of data sets can be treated subjectively by the end user. With holding data due to concerns over accuracy or quality should only be permitted in cases where the data is so unreliable it would inappropriately affect comparison with other data sets from within the sector to which it relates. In such cases a body such as the proposed Public Data Corporation should retain the authority to work with that particular public body or supplier to raise data standards so that data can be published. It would be important though to ensure that such a body were given the appropriate authority to intervene and compel the organisation to change.

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?

Centrally. Part of the existing problem is the plethora of public sector organisations, their respective web sites and the inability to find data on them easily. Once defined, the only way to store the data and make it accessible easily will be through a centralised portal or repository. Any other solution would simply increase costs and cause confusion to data consumers.

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

Essentially that which is in the public interest. Accepting the vision behind the exercises relating to Transparency, Localism, Open Data and the potential creation of a Public Data Corporation, government must ensure that the combined impact is visibility, accountability and engendering of trust from the public that government, public bodies and public service providers are spending public fund effectively. Therefore data sets should be published at national, local and sector levels where public interest and the desire for data is highest.

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

Government should not publish data for the sake of it. Broadening the net of what is captured and reported will only create additional burdens on already tight budgets. The effort required to support a number of existing data sets in the public domain is considerable and the quality and inability to effectively cross reference or collate different groupings of such data already presents a considerable challenge to all concerned.

Publishing relevant data sets effectively will encourage confidence and improve information access and usefulness. Get the existing stuff correct and more refined first, then consider the gaps that still exist and how best information can be delivered that enhances both the public's desire for information and also provides a useful measure of an organisation's effectiveness when compared to others within the sector, region or nationally.

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

Absolutely, and that role is for government itself to become a savvy user of the data being collated and generated. What data helps drive government policy on a public body's effectiveness? How does government measure and rate its own performance? What data and in what format would provide the public and public bodies with effective information on which to compare services or performance? If government itself does not make use of the information that will be generated in order to deliver service improvements and delivery, what is the point of having the data? This should not just be an exercise of 'publish everything and let others interpret it' government itself has a responsibility to monitor, act on and improve data collation and development in order to provide the public with confidence on its performance, otherwise the exercise will be fruitless and simply another example of costly bureaucracy without delivering improvements and benefits.

Jake Master

Veronica Smith