

# Making Open Data Real: A Public Consultation

Response from:

Tim Manning, Freelance Business Architect. Former Design Lead and Business Architect on the UK Location Programme (January 2009 to November 2010)

Submission Date: 26 October 2011

Email: tim.manning@design4services.co.uk

## Introduction

This response covers each of the “questions for consultation”, with a further section covering additional comments.

## QUESTIONS FOR CONSULTATION

### As provided under “Glossary of Key Terms”

I am happy with the definitions presented here, although the term “Transparency” should also have been included, given its current usage within Government.

On (2), the presumption should be that all data is open, against which a number of ‘exception conditions’ apply. This is a significant area that needs some detailed proposals brought forward. The devil is in the detail here.

On (3), first we need to be clear by what we mean by ‘value for money’ and who is doing the judging. The ‘value for money’ question is most likely to come up when viewed from the ‘present state’ use of data and its associated cost base. But if one is seeking to create a new ‘future state’, where data is created in a form that is business interoperable, both for internal and external use, then this narrow view is inappropriate. In this future state, the cost of publication is likely to be very low.

In broad terms, I would say that a requester should only be required to pay if the data is not a by-product, but a core part of their Business Model (particularly in terms of their Revenue Model). Even then, it may be appropriate for the Business Model to be changed, e.g. the revenue stream being provided out of national funds, if it provides significant payback in terms of overall economic gain to UK plc. Location Identifiers are an example of this. Without an open, unified scheme for location identifiers, most of the benefits of Open Data are unlikely to be realised.

On (4), I’m not sure what you are trying to balance here. With an ever increasing range of public services being delivered by the private sector, it is important that the definition ensures that the presumption to publish data (open or otherwise) extends to all organisations engaged in the delivery of public services. I think your definition here achieves this, by it needs thorough testing.

On (5), the need is for legislation that establishes the presumption to publish, applies a universal definition of scope and an overarching framework across all existing related

legislation. The objective should be more Open Data and less costly red tape surrounding access, publication and re-use.

## **Policy Challenge Questions**

On (1) I would agree with the introduction of separate legislation to establish the presumption to publish. This should act as an overarching and compatible framework for all existing related legislation, cutting through and simplifying overall implementation in this area – simplified decision making (are we in or are we out, publish or not to publish), fewer FOIA requests, etc.

On (2), I would extend the scope of the Information Commissioner to encompass all aspects of Open Data, consolidating responsibilities in this area.

On (3), Open Data only relates to non-personal data. The issue that requires more investigation and the introduction of any necessary safeguards is where personal information is derived by combining and filtering non-personal data.

On (4), in the most part “Open Data” goes hand-in-hand with modern data and information architectures within an organisation. These allow the free movement of data within the organisation and its strategic partners. Redirecting this data to the outside world is a relatively simple step. And indeed, in many modern public sector service delivery models you need to do this anyway, as many of your partners are out there, e.g. social enterprises and commercial private sector partners.

On (5), make it a requirement. This needs to happen in relation to all government Business Interoperability standards. Open Data standards should be defined and made part of a refreshed and properly applied eGIF. Where possible, these should be based on International Standards.

## **Setting Open Data Standards**

On (1), the first step to achieving compliance is to establish a set of UK Standards and related technical guidance for the publishing of open data. This is already well advanced for geospatial information, through INSPIRE, and for a number of other information domains, but not all. Unfortunately data.gov.uk has not performed well when it comes to the adoption of internationally recognised Standards, devising its own unique approach to discovery metadata. This needs to be improved upon going forward.

With a Standards framework in place, compliance needs to be through a process of self regulation, central quality assurance monitoring and feedback. It has to be recognised that this is a journey, where the standards (e.g. for Data Quality and Linked Data) and the organisations need to be allowed to mature over time.

On (2), yes. These need to be scientifically based.

On (3), no. With an open market for creating information and other added value services from Open Data, the cream will float to the top. I see no need for accreditation and the consultation paper presents no argument for such a scheme.

## Corporate and Personal Responsibility

On (1), Public Sector organisations need to build the commitment to Open Data into their existing Operating Models. This needs to be an extension of their existing Data/Information Management support processes, which in turn will almost certainly need changing as a result of the need to publish.

The initial commitment should be covered in (published) departmental Business Plans. This should be followed up by a 'community of interest' based approach, similar to the one adopted by the UK Location Programme for the publishing of data against the INSPIRE Themes. These should be organised around the different information domains, as this will have a strong bearing on the data publishing operating model adopted.

On (2), it should definitely be a corporate responsibility and thus represented at board level. And it should definitely be part of existing responsibilities concerning personal data and privacy.

On (3), it should be a legal requirement and current responsibilities should be rationalised around the Information Commissioner.

On (4), all sectors would benefit from a Transparency Board, to differing degrees. The obvious common area is the social and environmental impact of the sector. But there will be specifics in each sector as well, e.g. in the case of the Financial Sector, greater transparency of international/national exposure to risk; and in the case of sectors with limited competition, e.g. the Energy sector, greater transparency of the operation of the market.

## Meaningful Open Data

On (1), ideally there should be only one central inventory catalogue ([data.gov.uk](http://data.gov.uk)), which holds discovery metadata records for all data, both open and 'closed'. A single view of what data exists.

In terms of actual usage, this catalogue may appear in many different forms and settings (e.g. within a thematic digital channel), but still derived from the common inventory.

On (2), value is best determined by the level of demand. Publish and let demand determine the value. If there is no external demand, it is down to the individual data requester to determine if the base internal demand justifies the cost of producing the data, or whether an alternative approach might be more cost effective.

On (3), where there is a demand, both already recognised and emergent from the initial publication of the data.

On (4), see (2). Within an Open Data 'system', data providers/publishers should through the feedback mechanisms established, review the value of the data they produce and alternative sources for this data. This feedback mechanism should over time remove any "unnecessary" datasets. This process needs to be established as part of their operating model.

On (5), data quality is a very difficult area. A dataset may be fit for the purpose it was originally collected for, but not fit for purpose for potential areas of re-use, e.g. in terms of business interoperability. If the data is to be improved for re-use, the question is then who

pays for the improvements. I think that this is a subject for later on in the 'maturity cycle'. We need to learn to walk first. Data should be published as is, with any 'fit for purpose' constraints being clearly stated in the discovery metadata.

### **Government Sets the Example**

The questions for consultation do not really reflect the preceding discussion. Regarding this, yes, government should publish the evidence base behind policy statements and the data underlying surveys. It is often this underlying data that is the more useful.

On (1), data should be 'held' as close to its point of creation as possible, i.e. within the originating organisation. This helps to ensure that it is effectively managed by those who know the data.

In the age of the Internet and Web Services, there is no argument for the creation of large centralised data repositories. It is the standards and their adoption that provide the 'centralisation', by creating a common business interoperability framework, which everyone works within.

No data should be held on a "portal" as such - a portal is just a point of access.

On (2), recognised demand, e.g. through FIOA requests. On likely emergent demand, leading data re-users are probably your best guide.

On (3), the priority should be to broaden the set of data published. The publishing of more detailed data should be driven by the demand for the higher level data.

### **Innovation with Open Data**

There is definitely a role for government to stimulate innovation in the use of Open Data. It requires significant social and economic change against the old models of data collection and use. This change will take significantly longer without central stimulation.

This stimulation needs to be multi-stranded, but there are almost certainly some key leverage points. Greater collaboration may well be one, but perhaps greater is how people within organisations think about data – where it comes from, how it is used and how it can be combined - data/information literacy.

## ADDITIONAL COMMENTS

### Terminology, Meaning and Scope

There is some confusion within the consultation paper regarding terminology, meaning, objectives and scope.

By “Transparency” (which strangely is not included in the glossary), my understanding is that this is referring to ‘transparency of Government’. But here it seems to mean “Transparency of data” (as in para. 4.1). This also appears to be its usage in sec. 7, under “Accountability”, where it seems to be deemed equivalent to “Open Data”.

Open Data is not the same as “access to data” (para. 4.1). The examples you give regarding access are where, as part of the service delivery model, the customer drives the process themselves, i.e. they have access to their ‘own data’ and/or the information they require and self serve. “Open Data” is where everyone has access to the data.

### Six Opportunities of Open Data

I would encourage you to do some more formal Benefits Modelling in this area, to clearly work through the Objectives (those listed here and any others); and the related Enablers and Benefits that will lead to their achievement. Open Data is one, but there are others, including Business Interoperability and Data Literacy. I’m aware that the UK Location Programme has done some work in this area fairly recently.

### Access to Own Personal Data

In para. 5.1 “Right of individuals to access and control their own service user records, e.g. a school or personal health record” is a very valid government objective, but really lies outside the definition of “Open Data” (as reflected in your own glossary of terms).

It is true that, rather confusingly, providing access to one’s own personal records is also referred to as ‘Open Data’ in some cycles, but as I have indicated above, this is more to do with service design and adopting modern approaches to service delivery within the public sector. I would strongly recommend that you de-couple these two initiatives in your policy document, ensuring that this separate objective is picked up elsewhere. It has its own set of unique challenges.

### Sustainability

In sec.4.7 you make the all important point of ensuring continued access to data once released. This is critical and needs to be a key feature within the policy document. It’s very easy to focus on getting things out (by what ever means), but if not subsequently maintained, this is no good for the creation of sustainable business applications.