**Open Data and Open Government in the UK: how Closely are They Related?**

Martin De Saulles

School of Computing, Engineering and Mathematics, Faculty of Science and Engineering, University of Brighton, Brighton, UK

mrd@brighton.ac.uk

**Abstract:** Over the previous decade a number of initiatives within European Union (EU) Member States, in particular the UK, have attempted to open up access to public sector data for broader economic and social uses. Driving some of these initiatives has been the ambition to increase the transparency of public bodies and, as a result, improve the democratic process. Other initiatives have had an economic agenda that see the opening up of public sector data for commercial exploitation as a way to stimulate economic activity and growth. In the UK, the launch of the data.gov.uk website in 2010 combines both these ambitions by making over 8,000 public data sets available for third parties to download and build information services on top of. This paper examines a sample of these data sets as well as some of the applications that have been developed from them and uses a conceptual model developed by Yu and Robinson (2012) in the US. The model provides a basis for determining the technical characteristics of the data (is it adaptable or inert) as well as the primary purpose of the data (is it to improve service delivery or public accountability). Based on the analysis of the sample of data sets from the data.gov.uk website, it is concluded that Yu and Robinson’s framework provides a useful basis for separating the technical characteristics of public data from the purposes to which they can be put. Further refinements of the model are suggested that would allow governments to benchmark their public data initiatives against programmes in other countries.

**Keywords:** egovernment, open data, open government

**1. Introduction**

Notions of open government can be traced back to the 18th century with Sweden’s constitutional legislation enabling a free press and revolutions in the United States (US) and France which had the ambition of curbing executive and legislative powers at their core. More recently, the United States Freedom of Information Act in 1966 and similar legislation across Europe since the 1970s have attempted to uphold the principle that providing greater access to information about the workings of the state is good for democracy. The United Kingdom (UK) can be seen as something of a laggard in this area, only passing its own Freedom of Information Act in 2000. While these pieces of legislation have made it easier for citizens and organisations to access government information resources at both local and national levels, it could be argued that they have not gone far enough in opening up the public sector to greater scrutiny. Most legislation forces public bodies to be responsive in the ways they provide access in that they deliver the information in response to requests from individuals. There have also been restrictions, particularly in Europe as to what those receiving the information could then do with it. While in the US there has been a tradition of allowing the private sector to freely commercialise public information assets, in Europe the commercial exploitation of public information has, until recently, been restricted (De Saulles 2007). A 2003 European Directive (2003/98/EC) has attempted to open up the European public information market but the response across the EU has been mixed (Janssen 2011). The UK has been one of the more enthusiastic implementers of this directive and enacted many aspects of it with the passing of national legislation in 2005 and the creation of an organisational infrastructure to manage its implementation.

More recently, there has been a growing realisation that legislation on its own is not sufficient to encourage a thriving market for the commercial and social exploitation of public data. The ways that the data can be accessed and the formats it is presented in will play a large part in what third parties can do with it (Berners-Lee 2009). This paper explores some of the issues and debates surrounding the relationship between open data and open government and, in particular, the data.gov.uk initiative in the UK. The results of an analysis of a subset of the data contained within the data.gov.uk service are presented and discussed using the framework devised by Yu and Robinson (2012) which provides a basis for describing both the technical nature of government data as well as the purposes to which it can be put.

2. Open Government and Open Data

The term, open government, is open to a number of interpretations relating to the democratic processes of the state, the accountability of public bodies and officials as well as access to social and economic information by individuals. More recently, it has been co-opted by sections of the technology industry where analogies between open government and open computing systems have been drawn and a movement calling for governments to be providers of open data has emerged. Janssen (2012) differentiates between those advocating the setting up of open data initiatives and more traditional campaigners for open government that argue for a right to information whilst acknowledging there is an overlap between them.

O’Reilly (2010) talks of ‘Government as a Platform’ and ‘Government 2.0’ by which he means using the potential of collaborative technologies, particularly those based around the World Wide Web (WWW), to help solve some of the civic issues faced by local and national public institutions. He argues that where governments are successful in service delivery it is when they encourage the development of adaptable platforms such as the highway system in the US which was funded and regulated by the state but which allows a range of services to be delivered across it. US economic growth during the latter half of the twentieth century, O’Reilly claims, was partly driven by the opportunities for the transport of goods and people across the country made possible by a modern road network. Technology platforms in the computing world such as the personal computing (PC) architecture set by IBM over 30 years ago and the application stores of Apple and Android, while not completely open, have stimulated massive innovation amongst hardware manufacturers and software vendors of all sizes. Zittrain (2008) calls this the ‘generative’ nature of some technology systems, in particular the Internet, by which third parties are encouraged to develop products and services to run on top of the enabling platform.

While it might be argued that there is a technologically deterministic aspect to O’Reilly’s linking of the characteristics of computing systems to government services, perhaps where public data is concerned the analogies have some relevance. Some public services such as healthcare, policing, rubbish collection and defence do not obviously lend themselves to being seen as platforms on which other services can be developed by third parties outside of government. However, where government is a provider of data the notion of a technology platform may be more useful. One of the key factors driving the popularity of the Internet as a network for accessing and sharing information has been the ease with which hardware and software developers have been able to create products and services which can run across this network. The standards governing the interconnection of devices to the Internet are open and do not require the granting of permissions or paying of royalties to any third parties or governing bodies. Similarly, as long as the agreed standards are conformed to, software such as web browsers and applications such as email, Facebook and YouTube can be deployed to run on these devices. Prior to the mass deployment of the Internet, proprietary networks such as CompuServe, Prodigy and AOL offered email and information services but did not interconnect with each other restricting the extent to which users could interact with friends and family on other networks. By applying the lessons learned from the Internet and other open systems to the provision of public data perhaps some of the economic and social benefits of unlocking this data can be realised.

**3. The data.gov.uk Initiative**

Since 2005 and the passing of the Re-use of Public Sector Information Regulations, the UK has been one of the most active EU Member States in attempts to stimulate a thriving data market for public sector information. One of the driving forces underpinning is a belief that economic value can be unlocked by allowing commercial entities to use public data to create profitable information services. Research carried out by PIRA (2000) for the European Commission compared the economic benefits of the relatively open US public sector information market with the closed one operating in the EU at the time. The authors of the study concluded that the economic benefits of a more open system which allowed the free and unrestricted commercial exploitation of public sector information far outweighed any short-term financial gains public bodies may obtain by selling data. A restrictive market, they argued, discouraged innovators from using this data to create commercially viable information services while an open market was a stimulus to innovation that ultimately created new companies, greater employment and, therefore, more tax receipts for the government.

While a desire to stimulate economic activity was a driving force of the European re-use of public sector information directive and the associated national legislation across many member states, more recently there have been moves to promote the social benefits of a more free data market. In the academic sphere this has resulted in a growth of open access journals and a belief amongst many in the scientific community that opening up access to the results of leading-edge research can only benefit society (Arzberger et al. 2004). However, this ethos has also permeated to public sector information where, arguably, the data may be less interesting than ground-breaking scientific research but can be put to socially beneficial uses. In the US, this has been applied to tracking and publishing details of where politicians and parties obtain their funding (Bender 2010, Newman 2010), to the freeing of corporate financial data from expensive databases that had been the preserve on investment bankers (Brito 2010). In the UK, one of the more high-profile initiatives has been the creation of crime mapping services that combine crime statistics with mapping services such as Google Maps to produce very localised maps of criminal activity around the country (Chainey and Tompson 2012).

Many of these not-for-profit services rely on the combining of different data sets, what is often referred to as mashups. Crucial to the success of these mashups is for their creators to have access to data which is in a format that can be manipulated via automated routines. This requires the originating data to be presented in a structured format that can be understood by computers i.e. machine-readable. One of the leading proponents of the need for public bodies to offer their data in machine-readable formats has been the British creator of the WWW, Tim Berners-Lee, who has argued the need for ‘linked data’. This builds on the principles behind the success of the WWW which allows for links to easily be made between web pages. This idea has been taken up by the UK Government which in January 2010 launched a web portal (www.data.gov.uk) providing access to many thousands of public sector data sets. As of 28 December 2012 it contained 8,981 datasets from 786 publishers comprising central government departments, local councils, National Health Trusts and an assortment of other public bodies. There is an understandable concentration of data sets from the larger bodies with the ten largest publishers accounting for 45% of all data sets. Much of this data is published under the terms of the Open Government Licence which was released by the UK National Archive in September 2010. This allows for the re-use and re-publishing of the information under the following terms:

“The Licensor grants you a worldwide, royalty-free, perpetual, non-exclusive licence to use the information subject to the conditions below:

You are free to:

* Copy, publish, and distribute the Information;
* Adapt the Information;
* Exploit the Information commercially for example, by combining it with other Information, or by including it in your own product or application.”

(The National Archive 2010)

The restrictions imposed by this licence are relatively minor with the main condition of re-use being an acknowledgement of the original source.

While the data.gov.uk initiative has been welcomed by many in the open data community, there are concerns about how innovative it really is. Janssen (2011) points to the growth in the data sets it contains from 2,500 at launch to over 6,900 18 months later but questions whether it is really presenting any new data that was not already available via other sources,

“Since data.gov.uk does not really seem to be demand-driven, but rather based on what datasets are laying in the public bodies’ cupboards and that may be considered interesting for the citizens by the government, one could wonder whether data.gov.uk will indeed be a considerable contribution to innovation.” (Janssen 2011, p451)

The UK coalition government which came into power in May 2010 has stated it aims to develop and extend the scope of the data.gov.uk initiative and has incorporated it into its ‘transparency agenda’ which aims to make the workings of the state more easily accessible to the electorate. However, a National Audit Office report in April 2012 was critical of how far these intentions were actually being realised (NAO 2012). While the report’s authors acknowledge that progress had been made with the Open Government Licence they were concerned that the disclosure of information by public bodies was not systematic and lacked consistency while it was difficult, if not impossible, for data users to be confident of the quality of many data sets. A UK Cabinet Office White Paper published in June 2012 reinforced the government’s commitment to open data and the data.gov.uk initiative but the proposals it outlined will take several years to be realised.

**4. Evaluating Open Government Data initiatives**

On the surface, it is relatively easy to measure and describe the structure and contents of the data.gov.uk portal. The site itself provides statistics on the number of datasets it contains, who the publishers are and how many visits the site has had. Some of these numbers have been presented above. However, evaluating the data sets on the basis of their function and format is less straightforward. The importance of machine-readable formats has already been described so whether a dataset is presented as a PDF document, HTML page or XML file will make a difference as to what third parties can realistically do with it. The nature of the content of the dataset is also important; a database of crime statistics which includes times, dates and map coordinates of the incidents may be put to very different uses than a document outlining the strategic objectives of a regional health authority. This is not to argue than one data set has a higher inherent social value than another but that its potential uses may be different. Yu and Robinson (2012) point out that the language used to describe initiatives such as data.gov.uk and other similar programmes around the world can mean different things to the technicians that build the systems and the policy makers that and politicians that promote them,

“The popular term “open government data” is, therefore, deeply ambiguous – it might mean either of two very different things. If “open government” is a phrase that modifies the noun “data”, we are talking about politically important disclosures, whether or not they are delivered by computer. On the other hand, if the words “open” and “government” are separate adjectives modifying “data”, we are talking about data that is both easily accessed and government related, but that might not be politically important.” (Yu and Robinson, 2012 p 181-182)

If this ambiguity is to be removed and a more concrete understanding of what “open government” actually means is to be gained then the authors believe it is important to separate the characteristics of the data from the reasons for which it is being disclosed and ultimately used. To achieve this they propose a stylised framework which describes the government data across two dimensions. The first dimension describes the structure of the data and how it is published and runs from ‘adaptable’ to ‘inert’. Adaptable data being that which can be easily manipulated and repurposed while inert data is presented in a format which makes further changes by third parties difficult or even impossible. The second dimension describes the data on a spectrum running between ‘service delivery’ and ‘accountability’. Yu and Robinson give the example of machine-readable bus timetable data that may provide convenience to individuals, aid commerce and generally help provide a higher quality of life as being on the service delivery end of the spectrum. Data that discloses details of political funding or expenditure by public bodies would be on the accountability end of the spectrum as it could be seen to be increasing transparency. They acknowledge that their definitions may be rather binary and that some data will not neatly fall at one end or the other of these dimensions but as a framework within which to consider initiatives such as data.gov.uk it provides a useful starting point for evaluating public sector information.

The following section describes an analysis of a subset of data.gov.uk which uses Yu and Robinson’s framework in an attempt to determine how useful it is as an evaluative tool.

**5. Research Methodology**

The primary objective of this research was to consider the value of Yu and Robinson’s framework as a way to measure the characteristics of data sets in initiatives such as data.gov.uk. When the research was carried out there were more than 8,000 data sets contained within data.gov.uk and an analysis of all was not feasible for reasons of time. Therefore, a sample of 100 was analysed with the selection based on random sampling using a random number generator. While 100 data sets is relatively small compared to the total, it was considered sufficient to apply Yu and Robinson’s framework and provide the basis for a discussion of its utility.

Once the 100 data sets had been identified they were evaluated according to the extent that they were adaptable or inert and whether their primary purpose appeared to be to improve service delivery or public accountability. A data set was considered inert if it was presented in a static format such as PDF, HTML or Microsoft Word or adaptable if it was presented in a more malleable format such as Microsoft Excel, CSV or XML. It was also considered inert if restrictions beyond those contained in the Open Government Licence governed its re-use. Whether it was designed for service delivery or public accountability was a little more complex and required a more value-based judgement. A relatively large number of datasets detailed the expenditure above a specific threshold of public bodies and others contained salary levels of public employees and similar data. These were considered to be aimed at increasing public accountability. Data sets which included mapping information, transport statistics and other data that could not obviously be used to hold public bodies to account were classed under service delivery. A score of 1 was given against each of the two dimensions being measured giving a maximum score of 2 for each data set. For example, if a data set was presented as an Excel file, was released under the Open Government Licence and contained details of public expenditures it scored 1 as adaptable and 1 for public accountability and zero for the other variables. On the other hand, if it was a PDF document which outlined milk prices in Wales then it would receive 1 on the inert scale, 1 for service delivery and zero for the other 2 variables. 5 of the data sets were excluded from the final result as it was not obvious what the primary purpose of the data was intended to be. Once the data had been collated it was presented as a 4 spoke radar chart, shown in Figure 1 which is based on Yu and Robinson’s own representation of their framework.



**Figure 1:** Summary Results

**6. The Results**

It is important to remember that the sample being analysed comprised 100 data sets meaning that any conclusions drawn should be tentative. However, the results show a definite bias towards the data having a service delivery bias at the expense of public accountability. The extent to which the data is adaptable for re-use is less distinct with only a slight bias towards adaptability.

Based on this analysis it might be argued that the data.gov.uk initiative has so far produced data which is less aimed at making the workings of government and public bodies more transparent and more focused on improving the delivery of public services. There also appears to be a lot of work required if the data it is making available is fit for the purpose of re-use by third parties wishing to manipulate it as almost half of the data analysed was inert. It should also be noted that most of the data in the sample considered to be adaptable were presented as Excel sheets and CSV files which, although manipulable, do not conform to the same machine-readable formats which can be found in XML files or well-documented application programming interfaces (APIs). If this more rigid definition of adaptable had been applied to the analysis then the vast majority of the data sets would have been considered inert. Dead links were also an issue for 5 of the data sets where a publisher had uploaded a record to the data.gov.uk service that either had an invalid URL linking to the data or had changed the URL at a later date and not updated the record. For anyone attempting to build a data service on the basis of such data this could be a major problem.

**7. Conclusions**

As already mentioned, this research presents a first step in testing the value of Yu and Robinson’s framework for evaluating government data. A number of judgements were made during the research process to decide where individual data sets from the data.gov.uk initiative sat on the dimensions of the framework laid out by Yu and Robinson. For 95 of the 100 data sets analysed this was a relatively straightforward process and produced results that provided a useful overview of the nature of the data contained within data.gov.uk. These judgements could be calibrated depending on the extent to which data characteristics are considered adaptable or not. On the basis of the findings presented here, it could be concluded that this framework offers the basis of a useful method for evaluating government data programmes and providing a more nuanced explanation of both their technical characteristics and potential social value.

Further research may wish to consider analysing initiatives from other territories so that national comparisons can be made. It would also be useful for more work to be carried out on creating formal definitions as to what constitutes adaptable data allowing governments to benchmark their efforts at opening up public data.

**References**

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P., and Wouters, P. (2004) “Promoting Access to Public Research Data for Scientific, Economic and Social Development”, *Data Science Journal*, Vol 3, November, pp 135-152.

Bender, E. (2010) “Case Study: followthemoney.org” in *Open Government* (eds. Lathrop, D. and Ruma, L.), O’Reilly, Sebastopol.

Berners-Lee, T. (2009) “Putting Government Data Online”, *Design Issues*, www.w3.ord/DesignIssues/GovData.html.

Brito, J. (2010) “All Your Data Belongs To Us: Liberating Government Data”, in *Open Government* (eds. Lathrop, D. and Ruma, L.), O’Reilly, Sebastopol.

Chainey, S. and Tompson, L. (2012) “Engagement, Empowerment and Transparency: Publishing Crime Statistics Using Online Crime Mapping”, *Policing: A Journal of Policy and Practice*, Vol 6, No. 3, pp 228-239.

De Saulles, M. (2007) “When Public Meets Private: Conflicts in Information Policy”, *Info*, Vol 9, No. 6, pp 10-16.

Janssen, K. (2011) “The Influence of the PSI Directive on Open Government Data: an Overview of Recent Developments”, *Government Information Quarterly*, Vol 28, pp 446-456.

Janssen, K. (2012) “Open Government Data: Right to Information 2.0 or its Rollback Version?”, *ICRI Working Paper Series*, http://ssrn.com/abstract=2152566

Newman, D. (2010) “Cast Study: maplight.org”, in *Open Government* (eds. Lathrop, D. and Ruma, L.), O’Reilly, Sebastopol.

O’Reilly, T. (2010) “Government as a Platform”, in *Open Government* (eds. Lathrop, D. and Ruma, L.), O’Reilly, Sebastopol.

PIRA (2000) *Commercial Exploitation of Europe’s Public Sector Information*, PIRA International,

Leatherhead.

Yu, H. and Robinson, D. (2012) “The New Ambiguity of Open Government”, *UCLA Law Review Discourse*, Vol 59, pp178-208.

Zittrain, J. (2008) *The Future of the Internet*, Yale University Press, New Haven