



Submission to the Cabinet Office consultation *Data Policy for a Public Data Corporation*

27th October 2010

Dear Sir/Madam,

Please find below a submission from the TaxPayers' Alliance to the Cabinet Office consultation *Data Policy for a Public Data Corporation*.

If you have any questions about our submission, please feel free to get in touch with myself (Dominique Lazanski) or our Research Director John O'Connell.

Yours Faithfully,

Dominique Lazanski
Head of Digital Policy, TaxPayers' Alliance
dominique.lasanski@taxpayersalliance.com
0845 330 9554

John O'Connell
Research Director, TaxPayers' Alliance
john.oconnell@taxpayersalliance.com
0845 330 9554

Summary

- Government data that is freely available should continue to be freely available. Additionally, data that does not compromise the privacy of individuals should be released forthwith at no cost to the individual user or prospective innovator of that data.
- Government data is already being created, maintained, and stored at a cost to taxpayers. They therefore should be able to access data without licensing or other fees that would constitute a double taxation on the data already created.
- The Government should not create a new institution to manage data. It is absolutely inappropriate and unnecessary to spend money to set up a new institution even if the intention is for it to be self-funded in the future. This is especially true at a time when necessary spending reductions must be made.
- By extension, the Government should not set up a new regulatory body to oversee a new Public Data Corporation or a new entity which may become privately held and funded in the future for public data.
- Licensing and data governance needs to be approached on a case by case basis when it comes to requests for innovative use of data or large data sets. There is no one size fits all scheme available for data because new ways to use data have yet to be invented.

Introduction

The purpose of this consultation is to collect ideas about the organisation, management, structure, and set up of what a Public Data Corporation (PDC) would look like and function as. The consultation highlights the current legal framework and importance of open data while taking into account new possibilities of the reuse of open data.

Two of the important aspects of this consultation are the charging for government data and possible licensing models for that data. In terms of charging, three models are proposed. First, the status quo model lets open data remain open and free of cost. Second, the harmonisation and simplification model would let some data be open for free and other data would be charged for. Third, the freemium model is proposed as a parallel model to what is happening in the private sector. That is, a basic set of data is allowed for free while the complete set is charged for.

The other aspect of this consultation involves licensing. Suggested models are the used-based model, a PDC wide single agreement with supplemental licenses, or a PDC wide model with a single license. There are positives and negatives involved in each of the licensing models; however the choice of which model or hybrid model to go with depends on what the final outcome of the Public Data Corporation structure will be and how the governance structure is set up.

Overall, the consultation focuses on ideas within Government that are already going forward. This consultation does not just come out of the need to streamline open and public data, but comes out of an income issue among all government departments. Free data generally allows for only one type of income: revenues from taxing new applications and services built on the back of open data. The Treasury will be the main recipient of this indirect type of revenue. While this would benefit the UK's economy and the Government's revenue as a whole, it would not increase the bottom line revenue for each individual government department.

The idea that individual government departments seek to benefit directly from open data is one of the premises of this consultation response. The Public Data Corporation will enable individual government departments to receive direct revenue through specific licensing agreements. With that in mind, the purpose of open data changes to one that does not look to benefit the public with better services or the creation of new jobs through innovative reuse, but one that seeks to maintain territorial rights of individual government departments and increase their income.

This response is laid out as follows. First, a brief overview of the academic setting is discussed. Then types of data and a Public Data Corporation itself is analysed. Finally, case studies are introduced to show how free and open data benefits not only the individual innovator, but the country as a whole.

Background and Setting

The idea of public data or open public data is nothing new. Countries, organisations, and businesses have been collecting, creating, and releasing data for more than thirty years or more. In the UK, the 1984 Data Protection Act first enshrined the idea of different types of data, as well as the principles of protection and access.¹ Since then both private and public entities have collected, used and released data.

In 2003, the European Commission adopted the Directive on the re-use of public sector information.² In 2006, the European Commission published the Commission decision on the re-use of public sector information³ which determines the condition of the reuse of documents held by the commission. In it, it says:

"Article 7 Principles governing charging:

1. The re-use of documents shall in principle be free of charge. 2. In specific cases, marginal costs incurred for the reproduction and dissemination of documents may be recovered."

Furthermore, on September 22 of this year the European Commission's Digital Agenda Commissioner Neelie Kroes announced the Commission's intention to publish a proposal revising the 2003 Directive. The proposal, due out in November, is likely to contain the following:

- Data formats, licensing, and best practice usage.
- Charging regimes including making data available for free as ensuring EC data is released as a part of doing business.
- Eliminating any loopholes or historical exceptions.⁴

This announcement indicates that the European Commission is thinking about how it deals with open public data and continues to think of most data as free, and freely available.

In 2008 the OECD produced recommendations for public data based on a number of working groups and workshop meetings. The result, entitled *The OECD Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information*,⁵ makes the case clearly for best practices and enhanced use of Public Sector Information. Many of the issues that arise in this consultation are clearly outlined in that document with recommendations including privacy, security, quality, integrity and copyright best practices and principles. It is worth noting that the 'reinvention of the wheel' thinking about the PDC is not needed because this document is comprehensive and inclusive, and outlines many issues the PDC consultation is dealing with.

¹ <http://www.legislation.gov.uk/ukpga/1984/35/contents> accessed October 7, 2011

² http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive/psi_directive_en.pdf accessed October 19, 2011.

³ http://ec.europa.eu/information_society/policy/psi/docs/pdfs/comm_decision_on_comm_reuse_psi/en.pdf accessed October 19, 2011

⁴ <http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/11/596> accessed October 26, 2011

⁵ <http://www.oecd.org/dataoecd/0/27/40826024.pdf> accessed October 7, 2011.

In particular, the OECD Recommendation makes note of the idea of pricing and says:

"Pricing. When public sector information is not provided free of charge, pricing public sector information transparently and consistently within and, as far as possible, across different public sector organisations so that it facilitates access and re-use and ensures competition. Where possible, costs charged to any user should not exceed marginal costs of maintenance and distribution, and in special cases extra costs for example of digitisation. Basing any higher pricing on clearly expressed policy grounds."

Furthermore, the OECD recently released a pre-paper on local content creation called *The Relationship between Local Content, Internet Development and Access Prices Main Findings and Conclusions*.⁶ (The final paper is due out at the time of this consultation deadline.) It says:

"Governments should embrace the idea of openness where public sector data is deemed to be available for use free of charge unless specifically exempted for protection of national security interests, personal privacy, the preservation of private interests or where protected by copyright, or the application of national access legislation and rules. When public sector information is not provided free of charge, it should be priced in a way that is fair, that facilitates access and re-use, and ensures competition. Where possible, costs charged to any user should not exceed marginal costs of maintenance and distribution, and in special cases extra costs for example of digitisation. Governments should make public-sector information available to as much of the population as possible. This includes provisions for those with special needs, including the elderly, persons with disabilities, the vulnerable, or with gender or cultural differences. Steps to introduce more inclusiveness will help promote the take-up of services and the potential for content creation and distribution."

So it is clear that from the 2008 working group – and this most recent paper coming out shortly – that the OECD recommends the release of public data for free or marginal costs so that it can be used, reused, and accessed in a free and fair way. Charging for open data, under the definition of the OECD, would limit access and exclude public members who have already paid for the creation of the public data through their taxes.

Furthermore, there is academic literature that suggests open and free data is the preferable option when releasing public data. In his 2008 paper entitled *The Economics of Public Sector Information*,⁷ Dr. Rufus Pollock comes to the conclusion that:

⁶ <http://www.oecd.org/dataoecd/4/41/48761013.pdf> accessed October 7, 2011

⁷ http://rufuspollock.org/economics/papers/economics_of_psi.pdf accessed October 10, 2011

"When it comes to charging 'users' of public sector information the case for pricing at marginal cost or below is very strong for a number of complementary reasons ... a) the mark-up to cover fixed costs is high, as marginal costs are such a low fraction of average costs b) the demand for digital data as with other information services is likely to be high and growing c) there are likely to be large beneficial spill-overs in inducing users to create new products and services based on the information... To conclude: most upstream, digital public sector information is best funded out of a combination of 'updater' fees and direct Government contributions with users permitted free and open access. Appropriately managed and regulated this model offers major societal benefits from in-creased provision and access to information-based services while imposing a very limited funding burden upon Government."

Dr. Pollock's full econometric analysis is available in his paper and well worth looking at to see how he arrived at his conclusions.

A recent McKinsey report from May 2011 looks at the use of what they call 'big data' or massive data sets. The report's findings include the fact that in Europe using big data sets could save more than €100 billion in increased efficiency and reduce overall fraud, while improving the collecting of tax revenues.

The report goes on to state:

"Government employees and their respective agencies can also benefit from making data available across agencies and organisational silos. This can reduce search times between and within agencies, which can make up a significant portion of Government employees' time. One tax agency was able to redeploy 20 percent of its full-time employees by gaining online access to a housing transfer deed databases rather than collecting the data from another Government agency on physical media (CDs) and manually searching through the data. Increasingly more Governments (at all levels) are beginning to adopt "open data" principles in which they make more raw Government databases available to the public. Data.gov.uk in the United Kingdom and the Aporta Web portal in Spain (www.proyectoaporta.es) are central Web sites that are examples of this trend, in which external stakeholders can access more and more large databases. Such efforts have unlocked a tremendous amount of innovation that combines data from multiple sources (e.g., "official" Government information from law enforcement and public works with "unofficial" citizen-reported information from social media) to create services such as hyperlocal news that describes events specific to a city block. Other examples, including expectmore.gov and Dr. Foster Intelligence, which provides health care information to UK citizens, are designed to measure program performance directly."



The European Commission, the OECD, McKinsey and academics have noted – as early as 2008 – that charging for open data is unsustainable and incentive hindering.

Types of data and cost of data

To understand the costs of data we must first look at the types of data involved. There are two main types of data available. These types of data occur regardless of the format that they are released in. They are:

- Static data – non-updating or time sensitive data that is created once. Population figures, annual GDP, tax payments or health treatments all fall into this category.
- Dynamic data – constantly changing data often relating to geo-temporal location. This includes bus travel times, travel directions, stock market changes, and timely expenditures.

The marginal costs associated with both of these types of data are different. Static data is created once, though long-term storage is needed. Formatting static data can take some time as different forms, like XML or JASUN take some quantifiable effort. However, once data is formatted and released the data is not updated. It is a snap shot, so to speak, of data available at that time and place.

Making this type of data available online costs nothing except the human time it takes to upload the data to the web site. Data needs to be stored either way – either freely available online or within secure storage. Both require time and money so the default position should be to aim for publication, as data is created and stored regardless.

Marginal costs are often associated with dynamic data. In the case of these data sets developers need to access data by 'pinging' the server on which it sits. This is due to the dynamic and changing nature of the data. Data like live train times is constantly changing due to delays, accidents, and the like. There are marginal costs associated with the server traffic and server load.

The question is then, how does the UK Government account for their costs in terms of serving this data and time spent managing the data? There are a number of ways to save money and reduce the cost burden on the Government, and therefore taxpayers. First, though it is ideal to have data in usable formats, it is more important to have the data open and free to use. Consumers of the data at this point really just want to access it and are willing to take the time to re-format it if necessary. Releasing data now as it is will immediately alleviate the time and cost it takes for formatting. Although it is important to stress that data should be in a usable format anyway, and the optimal result is the publication of data that can be manipulated.

Second, releasing all data that is eligible to be requested under the Freedom of Information Act before it is asked for would, in time, greatly reduce the costs burden on FOI teams. At first, there may be a slight increase in FOI requests as people engage with the data. Storage and release would become more responsive in time though. So in time, the need for FOI teams would wane, thereby cutting the costs to Government. Departmental expenses, budgets, and compensation are examples of FOI-able data that isn't always available online.



Third, access to dynamic data does incur a cost so how does the Government recoup that? For large scale, enterprise access to data from large corporations a license should be agreed to on a case by case basis. One approach is to recoup costs by charging for marginal costs only, or for two trade enterprise access for server deployment. In other words, companies who would want large scale access to data could provide access in return for hardware or hosting, if the costs were equal or the same to marginal costs.

Finally, one way to address issues of access to dynamic data is to use scalable, cloud computing services. This way the cost of accessing the data would be based on use of data. Cloud computing allows for scalability, but also allows for the ability to track and pay for usage of data through open APIs. In using cloud computing, the UK Government could use one of the small to medium startups in the UK. This way the Government could support UK based businesses while storing data to for all of the UK to use.

When it comes to releasing data and making that data free to access and free to use, it comes down to good governance and a bit of creativity to find economical solutions. The next section will deal with the PDC directly and look at alternative solutions.

Public Data Corporation

We don't support the creation of a new organisation to manage and maintain the opening up and licensing of Government data. Creating a new institution without a strong business case which would spend taxpayers' money is absolutely inappropriate and unnecessary. It is difficult to make a case for a new institution when spending reductions have to be made across the board.

However, if the Government brings together existing organisations like the Ordnance Survey and the Met Office, that would streamline and reduce the cost of their operating expenses. In time, it would be worth revisiting the idea of creating a PDC, provided it had a robust business case.

We don't see the new PDC as an entity that will reduce the operating cost of any of these individual organisations. In fact, by creating a centralised corporation for dealing with data, government departments and organisations will have to add an additional process to releasing the information. Instead of releasing the data on existing departmental and organisations' websites, data will have to be collated and maintained by the PDC. This process could increase cost of opening data by adding an extra layer of administration and may delay the opening up of data by requiring an extra step in overall release process.

An issue in creating the PDC is the lack of direct revenue received from opening up data by government institutions and departments other than the Treasury. This means that there may be a reluctance to engage properly with an open data agenda or policy. For example, if the Foreign Office opens data up for free, they don't see a direct benefit from their open data, by way of additional tax revenue from an entrepreneur who uses it. The Treasury sees the benefit of the tax revenue raised instead. It is near-impossible to identify a direct link between the tax revenue and which open data set that revenue came from. Departments can't be 'paid back' based on tax revenue.

It is obvious that with a PDC, this won't be a problem. Charging for data will allow departments and organisations to directly account for their data based income. Charging for data makes it easy to identify a 1 to 1 relationship between data and income. For example, the Home Office can release data through the PDC knowing that if an individual or business buys it, the revenue can be accounted for and directed back to the Home Office.

But accounting for the revenue based on opening data supports the idea of a double tax. Taxpayers have already paid for the creation of that data. That data, whether or not it is opened up, is already collected, analysed and stored at taxpayers' expense. The choice, however, is in whether or not we use the data that we already paid for. Some individuals and businesses will choose to innovate on this data and some will not.

As consumers we pay a third time for an app or a website in which this data is presented after it is innovated upon. However, consumers, if they are charged, are paying not for the data again, but for the service and curatorial aspects of the data that has been created by the use of open data. This is not an additional tax, but a market and consumer choice.

There is no doubt that there is a cost to government departments and organisations by opening up data. We have made suggestions to overcome that issue in the previous section. Another is the licensing back of an application or data at a minimal or no cost to the government department from a third party. This is one possible solution to creating value for both the users of the data and the department from which it came. Crucially, it is also one that makes the case against of charging for data.

So what are the next steps then if the Government doesn't create a Public Data Corporation? Clearly there should be some best practices around opening up data, but what should those be?

In talking to companies like Pearson⁸ and the Guardian Data Store⁹, they have had challenges in opening up their data for commercial use and some of their experiences can be used and applied to government data. Both have offered their commercial data in the form of tiered costing, but both have faced challenges in that this kind of one-stop-shop doesn't suit all kinds of developers, even with tiered offerings.

The experiences both have regarding licensing is to be agile and creative. Developers, from the person in the bedroom to Microsoft, are going to have different needs when asking for data, different formats, and APIs to access that data. The concern with a PDC is that though centralisation of that data might appear to make this process more efficient, in reality different government departments – and organisations – will have different expert knowledge and experience in dealing with their own data. One cannot assume that developers who want to use Ordnance Survey data are going to be looking for the same kinds of things that developers who want to access tax data are hoping to find and use. And public organisations dealing with specialised health data will surely have more expertise in preparing and disseminating this information than a centralised body. In other words, the ability to respond to such varying needs is extremely important. And the ability to negotiate and manage bespoke licensing is also key.

Also, this means that the Government simply must engage with the developer community and allow for that community to give feedback and other comments. Developers know what they want, or can at least experiment with data. And they also know what doesn't work through that process. It is absolutely imperative to have some type of community engagement. And this just doesn't mean simply a website or Twitter, but it means really engaging – meeting with developers, holding development days, and ensuring voices who want to be heard are actually heard.

Overall, the way that the PDC is proposed to look right now means that it should not be created. We encourage the idea that Government departments and other public organisations open up as much data as they can, while being agile in their management of licenses and outreach to developers.

⁸ <http://developer.pearson.com/> accessed October 26, 2011

⁹ <http://www.guardian.co.uk/data> accessed October 26, 2011



Data as stimulus

Releasing data for free could stimulate the economy too. There is no doubt that in opening data up the use of it creates better services and more useful information. But opening up data also creates the opportunity for innovation. Through that innovation, new jobs and businesses are created providing growth in the economy. Open data should be part of the Government's growth agenda. Open data acting as a catalyst for economic growth is one of the main reasons not to charge for it.

The subsequent section discusses a number of case studies. They show that releasing open data has boosted economic growth and the quality of services.

Case Studies

The US Government's site Data.gov estimates that there are 21 different international websites which host released open data.¹⁰ And over 172 agencies and sub-agencies of the United States alone have released open data sets. This is an extraordinary number compared to just a few years ago.

There are a number of case studies available from around the world including the success of data.gov.uk and here we examine four.

Open Data exposes \$3.2 Billion Canada tax fraud

David Eaves of Vancouver, Canada, reported that he had a colleague who analysed the charitable sector in Canada. After much wrangling to get the tax returns of Canadian charities and after additional work to put the data in a readable and process ready format, Mr. Eaves' colleague discovered fake charities and fraudulent tax returns.¹¹ If the tax returns were open and available as a matter of course for all public data, this tax fraud might have been revealed earlier saving the Canadian Government over \$3.2 billion. At the time of this case study, the Canadian Government was actively pursuing the fake charities.

Kenya Open Data initiative contributes to GDP

On July 8, 2011, Kenya launched its open data platform called Kenya Open Data.¹² At a press conference for the launch, Bitange Ndemo, secretary of the Kenyan Ministry of Information and Communications, said that by releasing data he planned on seeing the ICT sector increase its contribution to 15 per cent of GDP.¹³ Among the applications already developed are the Huduma,¹⁴ a web and text based local constituency concerns and grievance system; and Msema Kweli,¹⁵ which tracks development funds spending. Both are community applications that have already saved many thousands of shillings in their up to date nature by providing immediate communication tools.

New York City: Real Time Bus locations at a fraction of the tender cost

OpenPlans, a not for profit IT development charity, launched a pilot project to put the B63 bus route online in February 2011.¹⁶ The project in itself isn't a big one, but the idea came from work that OpenPlans was undertaking with New York City's Metropolitan Transit Authority (MTA). The MTA was looking to expand their real time bus service, but it was estimated that \$140 million was needed to install devices needed for the service. OpenPlans spent just \$265,000 developing an open platform for using publically available data that

¹⁰ <http://www.data.gov/opendatasites#mapanchor> accessed October 7, 2011

¹¹ <http://eaves.ca/2010/04/14/case-study-open-data-and-the-public-purse/> accessed October 10, 2011

¹² <http://opendata.go.ke/> accessed October 10, 2011

¹³ <http://radar.oreilly.com/2011/07/open-kenya-Government-data.html> for press conference quotes. Accessed October 10, 2011.

¹⁴ <http://huduma.info/>

¹⁵ <http://msemakweli.ihub.co.ke/pages/home.php>

¹⁶ <http://openplans.org/2011/02/01/openplans-puts-mta-bus-locations-online/> Accessed October 10, 2011

MTA bus drivers already input at the start of their bus routes.¹⁷ The savings were substantial for both the MTA and taxpayers.

Todd Park and US Healthcare data

In June 2, 2010, Todd Park, the Chief Technology Officer at the US Department of Health and Human Services and a successful entrepreneur, showcased work of various startups based upon the release of US health data. Tools such as iTriage¹⁸ helps users look for local medical treatment for specific issues. Mr. Park suggests that this is only the beginning and millions if not billions of dollars can be generated from opening up such data, while new solutions and better healthcare also results from such information.¹⁹

There are many more case studies available and happening right now. The Federal Communications Commission, in conjunction with the Knight Foundation, just closed a round of competitions for the development of new apps in the US based on public open data. The prizes amount to around \$100,000 and were submitted on the Apps for Communities portal.²⁰ Who knows what new and inventive ways information will be used, repackaged and delivered to users? The only way we can find out is to release the data.

¹⁷ <http://secondavenuesagas.com/2011/02/01/along-the-b63-an-in-house-real-time-tracking-solution/> for open software development and project funding. Accessed October 10, 2011.

¹⁸ <http://www.itriagehealth.com/>

¹⁹ <http://www.theatlantic.com/technology/archive/2011/06/can-todd-park-revolutionise-the-health-care-industry/239708/> accessed October 10, 2011

²⁰ <http://appsforcommunities.challenge.gov/>

Conclusion

The conclusion to this consultation response is clear: do not create a Public Data Corporation. It may seem tempting for the Government to charge and receive direct financial benefits from opening up data, but public data should remain free. Taxpayers have paid for it. Also, genuinely open data will encourage innovation and economic growth. Streamlining licensing, output formats and release dates are, of course, necessary and continue to be so across Government. But continuing with the data.gov.uk platform is suitable at this time.

The Public Data Corporation will cost additional time, money, and expertise all of which does not exist in a coherent form at this time. Spending even more money on the creation of a new entity is not necessary. And charging for data that has already been created and stored thanks to taxpayers' money is not acceptable.

Open data brings benefits such as increased tax revenue, new job creation, economic growth and better and efficient services. Ensuring that these goals are maintained can happen without the creation of a new and costly Government entity. In fact, they'll happen more effectively if departments and public organisations simply released data in a useable format.