

Evaluating the Autumn Statement Open Data Measures

Submitted on 06 Dec 2011 by

(this guest blog posting is [cross posted from Tim Davies' personal weblog](#))

For open data advocates, the recently published Chancellor's Autumn Statement, underlined how far open data has moved from a small geeks issue, to an increasingly common element in Government policy. The statement itself included a section announcing new data, and renewing the argument that Public Sector Information (PSI) can play a role in both economic growth, and public service standards.

1.125 Making more public sector information available will help catalyse new markets and innovative products and services as well as improving standards and transparency in public services. The Government will open up access to core public datasets on transport, weather and health, including giving individuals access to their online GP records by the end of this Parliament. The Government will provide up to £10 million over five years to establish an Open Data Institute to help industry exploit the opportunities created through release of this data.

And accompanying this the Cabinet Office [published a paper of Further Detail on Open Data Measures in the Autumn Statement](#), including an updated on the fate of the proposed Public Data Corporation consulted on earlier in the year. Although this paper includes a number of positive announcements when it comes to the release of new datasets such as detailed transport and train timetable data, the overall document shows that government continues to fudge key reforms to bring the UK's open data infrastructure into the 21st Century, and displays some worrying (though perhaps unsurprising) signs of open data rhetoric being hijacked to advance non-open personal data sharing projects, and highly political uses of selective open data release.

In order to put forward a constructive critique, let us take the governments intent at face value (the intent to use PSI and open data to promote economic growth, and to improve standards in public services), and then suggest where the Open Data Measures either fall short of this, or where they should otherwise give cause for concern.

A strategic approach to data?

Firstly, let's consider the particular datasets being made available: there are commitments to provide train and bus timetable information, highways and traffic data, land registry 'price paid' data, Met Office weather data and companies house datasets all under some form of open license. However, the commitments to other datasets, such as key ordnance survey mapping data, train ticket price data, and the national address gazetteer are much more limited, with only a limited 'developers preview' of the gazetteer being suggested. There appears to be little coherence to what is being made available as open data, nor a clear assessment of how the particular datasets in question will support economic development and public accountability. If we take seriously the idea that open government data provides key elements of infrastructure for both enterprise and civic engagement in a digital economy, then we need a clear strategic approach to build and invest in that infrastructure: focussing attention on the datasets that matter most rather than seeing piecemeal release of data [1].

Clear institutional arrangements and governance?

Secondly, although the much disliked 'Public Data Corporation' proposal to integrate the main trading funds and establish a common (and non-open) regime for their data, has disappeared from the Measures, the alternative institutional arrangements right now appear inadequate to meet key goals of releasing infrastructure data to support economic development, and removing the

inefficiencies in the current system which has government buying data off itself, reducing usage and limiting innovation.

The Open Data Measures propose the creation of a ‘Public Data Group (PDG)’ to include the trading funds who retain their trading role, selling core data and value-added services, although with a new responsibility to better collaborate and drive efficiency. The responsibility to promote availability of open data is split off to a ‘Data Strategy Board (DSB)’, which, in the current proposal, will receive a subsidy in it’s first year to ‘buy’ data from the PSG for the public, will in future years rely for it’s funding on a proportion of the dividends paid from the PDG. It is notable that the DSB is only responsible for ‘commissioning and purchasing of data for free release’ and not for ‘open’ release (the difference is in the terms of re-use of the data), which may mean in effect the DSB is only able to ‘rent’ data from the PDG, or that any data it is able to release will be a snapshot in time extract of core reference data, not a sustainable move of core reference data into the public domain.

So – in effect whilst the PDC has disappeared, and there is a split between the bodies with an interest in maximising return on data (PDG), and a body increasing supply of public data (DSB) – the body seeking public data will be reliant upon the profitability of the PDG in order to have the funding it needs to secure the release of data that, if properly released in free forms, would likely undermine the current trading revenue model of the PDG. That doesn’t look like the foundation for very independent and effective governance or regulation to open up core reference data!

Furthermore, whilst the proposed terms for the DSB terms state that “Data users from outside the public sector, including representatives of commercial re-users and the Open Data community, will represent at least 30% of the members of DSB”, there are also challenges ahead to ensure data users from civil society interests are represented on the board, including established civil society organisations from beyond the technology-centric element of the open data community (the local authority or government members of the board will not be ‘open data’ people, but simply data people – who want better access to the resources they may already be using; we should be identifying similar actors from civil society to participate – understanding the role of the DSB as one of data governance through the framework of an open data strategy).

Open data as a cloak for personal data projects and political agendas?

Thirdly, and turning to some of the other alarm bells that ring in the Open Data Measures, the first measures in the Cabinet Office’s paper are explicitly not about open data as public data, but are about the restricted sharing of personal medical records with life-science research firms – with the intent of developing this sector of the economy. With a small nod to “identifying specified datasets for open publication and linkage”, the proposals are more centrally concerned with supporting the development of a Clinical Practice Research Datalink (CPRD) which will contain interlinked ‘unidentifiable, individual level’ health records, by which I interpret the ability to identify a particular individual with some set of data points recorded on them in primary and secondary care data, without the identity of the person being revealed.

The place of this in open data measures raises a number of questions, such as whether the right constituencies have been consulted on these measures and why such a significant shift in how the NHS may be handing citizens personal data is included in proposals unlikely to be heavily scrutinised by patient groups? In the past, open data policies have been very clear that ‘personal data’ is out of scope – and the confusion here raises risks to public confidence in the open data agenda. Leaving this issue aside for the moment, we also need to critically explore the evidence that the release of detailed health data will “reinforce the UK’s position as a global centre for research and analytics and boost UK life sciences”. In theory, if life science data is released digitally and online, then the firms that can exploit it are not only UK firms – but the return on the release of UK citizens personal data could be gained anywhere in the world where the research skills to work with it exist.

When we look at the other administrative datasets proposed for release in the Measures the politicisation of open data release is evident: Fit Note Data; Universal Credit Data; and Welfare Data (again discussed for ‘linking’ implying we’re not just talking about aggregate statistics) are all proposed for increased release, with specific proposals to “increase their value to industry”. By contrast, no mention of releasing more details on the tax share paid by corporations, where the UK issues arms export licenses, or which organisations are responsible for the most employment law violations. Although the stated aims of the Measures include increasing “transparency and accountability” it would not be unreasonable to read the detail of the measures as very one-sided on this point: and emphasising industry exploitation of data far more than good governance and citizen rights with respect to data.

The blurring of the line between ‘personal data’ and ‘open data’, and the state’s assumption of the right to share personal data for industrial gain should give cause for concern, and highlights the need for build a stronger constituency scrutinising government open data action.

Building capacity to use data?

Fourthly, and perhaps most significantly if we are taking seriously the goal of seeing open data not only lead to economic development, but also to better public services, the measures contain a dearth of funding or support to truly support the sorts of skills development and organisational change that will be needed to have effective use of open data in the UK.

The Measures announce the creation of an Open Data Institute, with the possibility of £10m match funding over 5 years, to “help business exploit the opportunities created by release of public data” which does have the potential to address much needed research to the gap in understanding and practice on how to build sustainable enterprise with open data. However, beyond this, there is little in the measures to foster the development of data skills more widely in government, in the economy and in civil society.

We know that open data alone is not enough to drive innovation: it’s a raw material to be combined with others in an information economy and information society. There are significant skills development needs to equip the UK to make the most of open data – and the Measures fall short on meeting that challenge.

A constructive critique?

Many of the detailed measures from the Autumn Statement are still draft – subject to further consultation. As a package, it’s not one to be accepted or rejected out of hand. Rather – there is a need for continued engagement by a broad constituency, including members of the broad based ‘open data community’ to address the measures one-by-one as government works to fill in the details over coming months.

Footnotes

[1] An open data infrastructure: The idea of open data as digital infrastructure for the nation has a number of useful consequences. It can help us to develop our thinking about the state’s responsibility with respect to datasets. Just as in the development of our physical infrastructure the state both invested directly in provision of roads and railways, has adopted previously privately created infrastructure (the turnpikes for examples), and encouraged private investment within frameworks of government regulation, a strategic approach to public data infrastructure would not just be about pre-existing datasets having an open license slapped on them – but would involve looking at a range of strategies to provide the open data foundations for economic and civic activity. Government may need to act as guarantor of specific datasets, if not core provider. When we think infrastructure projects, we can think critically about who benefits from particular projects: and can have an open debate about where limited state resources to support a sustainable open data infrastructure should go. The infrastructure metaphor also helps us start to distinguish different sorts of government data, recognising that performance data and personal data may need to be handled

within different arrangements and frameworks from core reference data like mapping and transport systems information. In the later case, there is a strong argument to secure a guarantee of the continued funding of these resources as public goods, free at the point of use, kept in public trust, and maintained to high standards of consistency. Other arrangements are likely to lead to over-charging and under-use of core reference datasets, with deadweight loss of benefit – and particularly excluding civic uses and benefits. In the case of other datasets generated by government in the day to day conduct of business (performance data; aggregate medical records, etc.), it may be more appropriate to recognise that while there is benefit to be gained from the open release of these (a) for civic use; and (b) for commercial use, this will vary significantly on a case-by-case basis, and the release of the data should not create an ongoing obligation on government to continue to collect and produce the data once it is no longer useful for government's primary purpose.)