

Big Data and the Implications for Government

Abstract: Big data is more than the sum of the parts; it is about scale but also how we inter-connect once disparate data sets, and mine and analyse them. It is something that has started to make a significant impact everywhere from financial markets to supermarkets. And, as Andy Williamson explains, it is starting to become an important factor in the development of public policy and for the delivery and analysis of public services. Big data offers us powerful new ways to see the world around us but this comes with the challenges of ownership, privacy and misuse: it can be used for good or to constrain people. We have to ensure that, as the uptake of big data increases, our legislation and practice keeps pace, ensuring that mistakes and misuse are prevented.

Keywords: big data; government policy; public sector

WHAT IS 'BIG DATA'?

'Big data' is one of those troublesome new terms that is used far more than it is properly understood. It is, like many new terms emerging from the digital realm, a constatable discourse: fluid, evolving and shifting with use, re-use and wider adoption.

Inherent in the name is some concept of scale, 'big' data is clearly more than your shopping list! But big data is not simply about the scale of the data but the scale of the inter-connectedness, the relationships that exist between large and sometimes disparate data sets. So, 'big data' is data linked together, to create a digital picture that is bigger than the sum of the parts. Data is passive, it exists and we can use it or not, and link it with other data or not. Big data recognises this and therefore a common understanding of the concept is that it encapsulates the subsequent action taken with the data set. Mayer-Schonberger and Cukier (2013) suggest that big data refers to the things one can do only on a large scale that cannot be done at a smaller one, including:

To extract new insights or create new forms of values, in ways that change markets, organizations, and the relationships between citizens and governments, and more.

In other words, it's not having 'big data' that makes the difference, it is what we do with it that matters.

THE ORIGINS AND GROWTH OF BIG DATA

Large data sets have been with us for a long time, banks and other large institutions have long held significant

volumes of data. What is different today is that cheap storage and computer processing power makes it more possible, smarter statistical and computational methods give it value and purpose and new ways of linking datasets (not least the internet and open connectivity standards) give them meaning and context.

So, big data are large volumes of inter-connected data that can be stored, processed and shared in new ways to give us a richer, deeper analysis and picture of what that data represents. This introduces us to the concepts of 'data mining' and 'data analytics', where powerful new computer processing techniques are used to discover, process and analyse vast, inter-linked data sets to understand such things as patterns, trends and sentiment. This has enormous commercial, as well as public sector potential and the organisations that become early adopters in this market describe significant competitive advantage (Wall, 2014). Think of this in terms of your Tesco Clubcard at one, perhaps relatively innocuous, level through to the detailed mining of private and potentially personal health data at another end of the spectrum. In between lie all sorts of things that today we take for granted but perhaps aren't as well informed about as we might think. Facebook is a good example of this, it is responsible for the creation of almost unimaginably vast amounts of data every hour of every day; and this data is harnessed to target advertising at us and to gain complex understanding of its users. It is a relatively obvious truism in the online world that, if the cost is zero, then you are the product. This is certainly true in the case of Facebook.

As you can see, the application of big data goes far beyond the technical. It very quickly escalates to issues of privacy, consent, the legal framework and the politics that

surrounds it, as well as a natural human fear of the unknown and the inherent distrust that this creates (Harris, 2014).

Thanks to the internet the amount of data that we create is growing at an unprecedented rate. In the past, data was primarily created by the corporate and public sector, held privately and used internally. The rise of internet-based networks has changed this so that data is now more available and the further evolution of the internet into the social web has led to a massive growth in 'data' production. And most of this is created by individuals on platforms such as Facebook, You Tube and Twitter.

It is estimated that there is now some six zettabytes¹ of digital data and that this is increasing by around 50% year on year (Meeker, 2014). According to IBM, during 2012 there was 2.5 billion gigabytes of data generated every day. They go on to note that approximately 75% of that data is unstructured, meaning that it comes from sources such as text, voice and video, rather than the more familiar kind of structured and often proprietary data that is held in traditional databases (Wall, 2014). Of all this data, it's been estimated that only 34% of it is 'useful' (there is a lot of machine-generated data that has no value beyond its original use, for example) but that only 7% of data has been tagged to give it any context and meaning and only 1% of all data that exists has actually been analysed (Meeker, 2014).

For data to be big, then, we have to satisfy three primary conditions (Livingstone, 2013):

- The data must have **volume**. We post, and Twitter distributes, about 600 million tweets every day. That's 42,000 tweets in the time it just took you to read the last two sentences.
- And if Twitter tells us one thing about what big data means, it's about **velocity**. This is data in real time, happening now and constantly changing and growing.
- The data can come in a **variety** of shapes, sizes and formats. Not just structured data but comments, blog posts, photographs, video and music.

And if this is what makes data 'big' then to qualify as useful, usable and tenable, big data also needs to exhibit a second set of characteristics:

- It must have **validity** because unless we can logically confirm that the data is correct we cannot rely on it.
- The data has to be testable for authenticity. Without **veracity** it is impossible for us to use the data to draw any reliable conclusions.
- Data of itself has no value but a key attribute of big data is that it is useful in ways beyond its own existence and therefore a **value** can be ascribed to it.
- Big data is about connecting and linking what might previously have been seen as disparate sources. It must have **visibility** so that it can be analysed.

However this does not mean that 'big data' is the same as 'open data', in other words it must be visible to the tools using it, which might be closed and proprietary, not necessarily to the public at large.

HOW BIG DATA IMPACTS THE PUBLIC SECTOR

If big data is a child of the internet revolution where does it fit into public service delivery? To understand this we need to approach the subject from two angles. First, technological adoption happens right across society so it is inevitable that, whilst governments might not be innovators or early adopters of new technology, they will eventually adopt some of these new tools. However, digital does not happen in isolation, our society is not technologically deterministic and it is important to recognise that the context for 'big data' in government is cultural. Over the last 30 years, many of the governments in the developed world have transformed into technocratic elites, driven by process improvement and service delivery – the language of managerialism not democracy. These systems of government have become more corporatised, often outsourced and very often – despite legislation to protect the public from this – opaque. More recently still the culture of austerity witnessed in the UK has led to a further retrenchment in government. There is a drive to efficiency and digital platforms are seen as critical drivers of that efficiency. Public sector big data in the UK exists too in the context of 'digital by default', whereby the delivery of public services are optimised for digital delivery and use (Cabinet Office, 2014).

Though I have already said that 'big data' is not necessarily 'open', in the government context much of it is likely to be. It is, therefore, worth briefly exploring the context for the increasing push towards transparency and open data in the public sector. This has taken on a prominent focus within the digital government agenda and rightly so. Though it's just one aspect of digitising our democracy, data are a vital engine to drive better decision making and to level the playing field for citizens. Open public data actively supports better, more active democracy.

There are many reasons why governments are moving towards transparency and adopting open government principles and many more reasons why citizens should demand that data be open by default. Whilst the G8 governments have already committed to exactly this (subject to legal restrictions, such as national security), implementation lags well behind intent (Chan, 2014). Perhaps a stronger driver of big data is austerity. Better use of government data sets through connection, analysis and application matters in terms of efficiency too, as Yiu (2012, p.7) noted:

there is scope to improve the overall efficiency of government operations, to accelerate efforts to reduce fraud and error, and to make further

inroads into the tax gap (the difference between actual tax collected and theoretical liabilities).

He estimated a potential cost-saving to the UK government alone of between £16 billion and £33 billion a year through the use of 'big data' and associated techniques. As Yiu notes, that is the "equivalent to £250 to £500 per head of the [UK] population". A large part of this will come through the improvements in operations and processes that can be made as a result of a better understanding what is happening. As Margetts (2012) observes, big data allows governments to draw not only on their own data and metrics but to mine a wide and rich landscape of other data sources, so they could, for example:

use data from social media for self-improvement, by understanding what people are saying about government, and which policies, services or providers are attracting negative opinions and complaints, enabling identification of a failing school, hospital or contractor, for example. They can solicit such data via their own sites, or those of social enterprises. And they can find out what people are concerned about or looking for, from the Google Search API or Google trends, which record the search patterns of a huge proportion of internet users.

This is no panacea, however, there are significant challenges in opening up our data and actively linking it to create 'big data' for public policy purposes. Not least, we cannot ignore the issues of privacy, how (or whether) the rights of individuals can be protected and what methods can be employed to guarantee an acceptable level of certainty in terms of anonymising data sets (Stough & McBride, 2014). Such issues come starkly into focus when governments suggest selling health or taxation data to corporates, as was recently mooted in the UK (Williamson, 2014).

Whilst this transformation will occur in part because governments get smarter at using the data in terms of mining and analysis at a service delivery and a transactional level, there are also benefits to be had at policy level. But for these to occur it will be critical that civil servants and the public become more effectively literate in their use and understanding of big data in the future:

The explosion of data and the power to manipulate it gives intimate insight into people's lives at a near population scale. This could fundamentally change social policy, just as mapping the human genome has affected medicine (Perrin, 2014).

To summarise, some of the benefits of big data for the public sector include:

- **Sharing** of data across currently disparate government and public sector agencies, though this will be subject to stringent controls and monitoring.
- **Learning** through the availability and interconnectedness of data and the use of new tools, recognising that this requires new skills, new kinds of information and literacy and that it can also increase complexity.
- **Personalising** data becomes increasingly possible through the real-time linking of data and massively increased granularity. The inverse of this is obviously concern around confidentiality and anonymity so data must be controlled carefully and, where appropriately, aggregated without any identifying features.
- Big data sets connected together help with **solving** complex problems and provide the opportunity for predictive analysis, examples of which include predicting patterns of neighbourhood crime or pharmacy stocks through analysis of prescriptions and health records. We must, however, ensure that this isn't done at the expense of personal stories and people's lived experiences.
- Finally, big data is a valuable tool for governments because it allows **innovation** as part of the wider digital and transformational economy, both inside the public sector and in the private sector.

WHAT THIS MEANS FOR CITIZENS

Big data presents the opportunity for governments to personalise public services making them more targeted, efficient and effective. Conversely, it creates an opportunity for misuse, mass surveillance, breach of privacy and an Orwellian state monitoring of citizens. So suggests The White House in their review of big data in government (Executive Office of the President, 2014, p.49):

Big data will enhance how the government administers public services and enable it to create whole new kinds of value. But big data tools also unquestionably increase the potential of government power to accrue unchecked.

Digital technologies can act as a disrupter or an enabler. As Singapore has discovered, big data can be used to foil terrorism or to mitigate the impact of flu, though this comes with a trade-off that the state must track and monitor people's activities (Harris, 2014).

Big data has clear economic benefits but there are also democratic benefits too and these are why public data must be opened up and made accessible. Open data (and by association, open sets of big data) provide accountability. As a subset of the wider philosophy of public transparency (and, conversely, the public's right to information) there is an increasing assumption, matched by the availability of technology that makes it easy, that

governments will share information and can now be held to account. One key issue here is trust: transparency and trust are “intrinsically linked” (Cabinet Office, 2013, p.31). Unfortunately, few of us trust our government: in the UK 90% of us believe that government is acting in the interests of small elites (Hansard Society, 2013) and Hager’s (2014) recent expose of ‘dirty politics’ in New Zealand suggests that this assumption is unlikely to be either ill-founded or restricted to the UK. Better access to better quality information can support the re-building of public trust in democratic processes (Williamson, 2011).

At the personal and social level, open access to big data increases public choice. The UK Government describes how citizens can take control of their lives through direct payments, personal budgets, entitlements or choice. Big data can be used to provide comparative information and enable meaningful choice in public services. Examples of this are the comparative analysis of service quality and supporting wider public debate and effective user engagement, leading to better input into policy and improved service design. As Williamson and Sande (2014) observe, we must consider the benefits of making big data available to citizens, so that they can be more informed and better resourced when they engage with government.

I’ve already mentioned that the potential benefits of big data are to some degree offset by the risks. Big data requires strong governance and a lack of process and standards can hamper its veracity (Martin et al, 2013). The strong risk emanating from big data is personal privacy and the related issue of public awareness. Sciencewise (2013) described public understanding as “generally low” and that open data was seen as “an abstract issue with unclear benefits to everyday life”, where there “may also be a lack of clarity for members of the public over exactly what is meant by ‘open data’”. In particular, the Sciencewise research highlights that the

public appears to struggle with the concepts of private and public data. And privacy issues (perceived or real) lie at the heart of concerns for many people over open and linked data. So much so that it is clearly articulated in government policies that all open data must be anonymised. O’Hara (2012) recommends that privacy needs must be addressed through the development of good practice, including pre-release evaluation of data sets and the creation of transparency panels, that the technical as well as legal aspects of privacy (which currently dominate the discourse) must be addressed and that, where data has privacy implications appropriate controls must be considered.

CONCLUSION

Big data has a big future in government. It will help us to make better decisions, to better understand the macro policy landscape as well as ourselves and our neighbourhoods. But it comes with a set of trade-offs and risks that are currently poorly understood by the public. There is a risk of inappropriate surveillance, of personal data leaking out and a strong need to educate the public to become ‘data savvy’.

Big data can and should be used for good, not to constrain people. It can be used to empower front line staff in the public sector to make effective and holistic decisions and to be fully informed. It can be used to make policy better informed and to model future services. But these are early days and current legislation struggles to keep pace with technological innovation. It is going to be vital that we develop clearer policies and frameworks about who owns, accesses and shares data. Big data is often framed in technocratic and economic terms. Whilst it has clear economic benefits in terms of both efficiency and innovation, it is the social and cultural impact that will determine its acceptance or otherwise.

Footnote

¹ Zettabyte = 10^{12} bytes, or 1 billion terabytes (where 1 terabyte = 1,000 gigabytes and 1 gigabyte is approximately a single episode of Coronation Street when streamed online).

References

- Cabinet Office (2010). Efficiency review by Sir Philip Green. London: HM Government.
- Cabinet Office (2013). Open Data White Paper: Unleashing the potential. London: HM Government.
- Cabinet Office (2014). Digital by default: service standard. Available online: <https://www.gov.uk/service-manual/digital-by-default>
- Chan, J. (2014). Policy Deep Dive: Local open data policy in Canada. Washington DC: Sunlight Foundation Available online: sunlightfoundation.com/blog/2014/07/11/policy-deep-dive-local-open-data-policy-in-canada
- Executive Office of the President (2014). Big data: seizing opportunities, preserving values. Washington, DC: The White House.
- Hager, N. (2014). Dirty politics: How attack politics is poisoning New Zealand’s political environment. Nelson: Craig Potton Publishing.
- Hansard Society (2013). Audit of political Engagement. London: Hansard Society.
- Harris, S. (2014). The Social laboratory. Foreign Policy. Available online: http://www.foreignpolicy.com/articles/2014/07/29/the_social_laboratory_singapore_surveillance_state

- Harris, D. (2014). Why big data has some big problems when it comes to public policy. Gigaom. Available online: <http://gigaom.com/2014/08/27/why-big-data-has-some-big-problems-when-it-comes-to-public-policy>
- Livingstone, R. (2013). The 7 Vs of Big Data. Available online: <http://rob-livingstone.com/2013/06/big-data-or-black-hole/>
- Margetts, H. (2012). The promises and threats of big data for public policy-making. OII Policy and Internet Blog. Available online: <http://blogs.oii.ox.ac.uk/policy/promises-threats-big-data-for-public-policy-making>
- Martin, S., Foulonneau, M., Turki, S., Ihadjadene, M. (2013). Open data: Barriers, risks and opportunities. ECEG 2013, Como, Italy.
- Mayer-Schonberger, V. and Cukier, K. (2013). *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt.
- Meeker, M. (2014). *Internet Trends 2014*. San Francisco: Kleiner Perkins Caulfield Byers.
- O'Hara, K. (2012). Transparent government, not transparent citizens: A report on privacy and transparency for the Cabinet. Available online: http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/61279/transparency-and-privacy-review-annex-a.pdf
- Perrin, W. (2014). We need a comprehensive review of the way Britain approaches the benefits and challenge of data. Labour Digital Government Review.
- Sciencewise (2013). Public views on open data. London: Sciencewise.
- Stough, R., & McBride, D. (2014). Big data and U.S. public policy. *Review of Policy Research*, 31(4), 339–342.
- Wall, M. (2014, Mar 4). Big Data: Are you ready for blast-off? BBC Business News.
- Williamson, A. (2011). Disruption and empowerment: Embedding citizens at the heart of democracy. *Journal of eDemocracy and Open Government*. Available online: <http://www.jedem.org/article/view/52>
- Williamson, A. & Sande, M. (2014). *From arrogance to intimacy: A handbook for active democracies*. London/Gothenburg: Democratise/Co-lab by Preera.
- Williamson, A. (2014). Open government risks becoming an oxymoron. Available online: <http://www.andywilliamson.com/index.php/oxymoron-open-government>
- Yiu, C. (2012). *The big data opportunity: Making government faster, smarter and more personal*. London: Policy Exchange.

Biography

Dr Andy Williamson, of Democratise, is an internationally recognised expert in digital democracy, democratic innovation and the transformation of citizen engagement. An experienced consultant, researcher and commentator, he works with parliaments, local and national governments and NGOs right around the world. Andy holds a PhD from monash University, Melbourne and is an RSA Fellow.