Policy Exchange

research note
March 2012

A Right to Data

Fulfilling the promise of open public data in the UK

Chris Yiu

With a foreword by Tim Kelsey

@PXDigitalGov #opendata

Executive Summary

A piece of data or content is open if anyone is free to access, use, reuse and redistribute it – subject only, at most, to the requirement to attribute and share-alike.

The business of government has always involved quantities of data. For centuries almost all of this public data has been closely guarded by the state. Governments provided access to data on a need-to-know basis and, for the most part, citizens didn't need to know.

The balance is, however, starting to shift. Advances in information and communication technologies mean that, for the first time in human history, it is technologically feasible for every citizen to have access to every piece of data or content generated by their government. And as this same technology drives fundamental changes in our economy and society, it is becoming clear that, with the right protections, opening up public data will deliver considerable benefits.

In this research note we review the state of open data policy in the UK. We find that the direct cost to the Exchequer of giving away key datasets like maps and postcode data may be far lower than sometimes thought – perhaps in the region of £50 million a year. The potential benefits are notoriously difficult to quantify, but are likely to be orders of magnitude greater.

We make three recommendations for government on open data policy:

1. **That the government enshrine a right to public data in legislation.** This should make all of the data collected or created to support the day-to-day business of government open: easy to access and free at the point of delivery, without restriction on use or reuse. Important protections for personal data, national security and Ministerial advice should be provided.

2. **That responsibility for open data rest at board level in every public sector body.** A statutory right to data will be most effective when mindsets and behaviours in the public sector are geared toward open ways of working. Leading and inspiring a culture of openness will be a greater challenge than simply implementing new technology.

3. **That every public sector body define its public task and associated data requirements.** The public sector's role is to deliver public services. Any activity based on leveraging public data to develop commercial products or services should, ultimately, be spun out. This will help prevent crowding out and foster innovation around the application of public data.

The impetus for these changes is threefold.

The first is economic. Data has become part of the fabric of the modern economy. It is an asset on a par with communications networks, transport infrastructure and payment systems. We identify a number of firms where the core of their business is working with open data. It is important to remember, however, that every business relies on critical datasets to locate customers, organise supply chains and analyse performance. By making it easier to organise economic activity efficiently, open public data will have wide benefits for businesses and consumers alike.

The second is social. Advances in processing power, storage and social networks mean that many of us are used to routinely dealing with large quantities of data. We know more about our world than ever before, and as citizens increasingly expect this to extend to the workings of government. In today's digital society the guiding maxim of "need to know" for public data is largely obsolete.

The third is political. For the public sector, open data is a necessary and important step on the road to open government. Transparency about public sector activities and decisions boosts accountability. And armed with public data, citizens can make informed choices about the services they consume and help drive improvements in public service delivery.

Against this backdrop there is no longer any reasonable justification for keeping publicly funded data under lock and key. Indeed, the potential for an open data revolution to permanently transform our economy, society and polity is within touching distance. Our recommendations provide the structure necessary to realise the potential of open public data in the UK.

We believe that a statutory right to data is required to guarantee openness as the new default, and to provide clear protections for personal data, national security and Ministerial advice. We believe that strong open data leadership at board level in every public sector body is needed to ensure senior officials take responsibility for ensuring that the spirit of this right to data is adhered to. And we believe that requiring every public sector body to define and focus on its public task is necessary to ensure that everyone is clear about what data the public sector is collecting and for what purpose.

Taken together, our recommendations would deliver on the government's manifesto and Coalition agreement commitments on open data, benefit our country significantly and set a leading example for the rest of the world. We urge the government to implement them.

Foreword

SEVEN years ago, British heart surgeons took the brave decision to publish data on their patient outcomes. Their transparency has saved lives; survival rates have measurably improved. In parts of India, where villagers do not have access to the internet, local officials are painting data in spreadsheets on the walls of communal building asking locals to identify fraudulent welfare claimants. In Brazil, officials publish details of expenses within 24 hours of their payment. Corruption has fallen and trust in government improved.

Sharing data – transparency – may turn out to be the most important public policy of our time. The British government has put it at the heart of its agenda: it is key to accountability and choice but also to transforming public services and promoting social innovation and economic growth. We now publish more datasets (currently around 13,000) than any other country in the world, in machine-readable formats and under the Open Government Licence.

Nine months ago, in July 2011, the Prime Minister announced release of open data in health, transport, criminal justice and education – from data on GPs to county court outcomes. Last November, the Chancellor confirmed the release of more raw data: the public weather service, for example, is now available for free re-use – meaning that real time feeds from the weather stations around our country are openly available for the first time.

As this report notes, the Cabinet Office recently ran a consultation on Open Data policy – and the overwhelming conclusion was for far more, not less. Later this year, we will be publishing proposals to strengthen the citizen's Right to Data, outline how public services will encourage citizen feedback and participation and describe new approaches to Open Data collaborations between government, business and the voluntary sector.

The underlying concept is simple: that government should be a source of information for citizens; and citizens should be a source of information for government. This is the engine of effective, equitable and efficient 21st century society. Open Government is just one part of the objective; an Open Society requires transparency from the whole community.

The Open Government Partnership symbolises that collaborative future: an international initiative which brings together more than 50 nations and many civil society organisations in a network to make transparency and participation real at home. The UK becomes co-chair with Brazil this Spring.

Transparency is not easy. At the beginning, it can prompt hard questions about how little government knows about itself, and it challenges many cultural norms in public service delivery and in society more widely. But the benefits and opportunity of Open Data outweigh any inaugural issues: the question for all of us – and the central focus of this paper, which we welcome – is how we make it real in our communities.

Tim Kelsey

Director of Transparency and Open Data

HM Government

1. Prologue

> *The Open Definition sets out principles to define 'openness' in relation to content and data – that's any kind of material or data 'from sonnets to statistics, genes to geodata'. The definition can be summed up in the statement that "a piece of content or data is open if anyone is free to use, reuse, and redistribute it – subject only, at most, to the requirement to attribute and share-alike."*
>
> *-- OpenDefinition.org*

Over the past decade the open data movement has gained momentum around the globe, from humble beginnings in forward-thinking public bodies to Tim Berners-Lee's memorable call for "raw data now" at TED 2009.[1]

This research note provides a framework for the next phase of the open data debate in the UK. We step back from a detailed discussion of individual data releases, and argue that to realise maximum economic and social benefits, policymakers need to embed open data principles deep into the mindsets and mechanics of the public sector.

2. Data is now part of our shared digital infrastructure

Before exploring the policy choices around public sector data, we need to identify the appropriate conceptual framework for thinking about data in the context of the modern economy. We contend that so-called "core reference data" – maps, postcodes, timetables and statistics – are now an integral part of the fabric of economic life. So we class this data as part of the nation's digital infrastructure – a category that contains the (ever growing) range of technologies and support factors that underpin the modern business world, from mobile telecommunications networks and broadband internet, through common standards like email and HTML, to electronic payment systems and services.

[1] Tim Berners-Lee on the next web, TED 2009

Virtually every firm in the country relies on this core reference data to run their operations and supply chain, identify and serve their customers, and plan the future development of their business. Every time a shipment is scheduled or an address is looked up from a postcode, for example, this core reference data comes into play. Moreover, a significant number of firms now devote at least some of their time and effort to leveraging this data, in combination with data fusion and analytics, to create value in the form of new information and insights of interest to their business customers and consumers.

Moreover, individuals are themselves accessing, using and consuming increasing quantities of data. Some of these channels for data consumption are well known: think fold-out Ordnance Survey Explorer and Landranger maps, the school and university league tables published in many national newspapers, abstracts of National Statistics. Others are more recent: in-car satellite navigation systems, and apps and websites for visualising almost any data one might imagine.

3. We are interested in non-personal data collected or created as part of the public task

In conjunction with the digital infrastructure concept introduced above, we need a way to identify which data is in scope for the policy recommendations in this note.

For the avoidance of doubt, we place *personal data* (regardless of whether it is held in the private or public sectors) out of scope for the main discussion in this paper. We also hold over a detailed review of how the public sector could make better use of its own *big data* assets, either internally or through properly controlled *data sharing* with the private sector, for another occasion.[2]

We define *public data* as any non-personal data collected or created as part of the public task. We propose only one exception to this definition: non-personal data relating to national security, which necessarily does not belong in the public domain.

This definition puts a range of data in scope for our discussion. The table below provides a (non-exhaustive) illustration of the sorts of data we have in mind.

[2] The term big data typically refers to very large, granular datasets. Organisations can often interrogate this data to generate insights about their business or customers. In the public sector context, big data assets might include large datasets held by e.g. HM Revenue & Customs or the NHS.

Table 1: The public data universe		
Types of data	**Public sector users**	**Examples**
Management information and other operational data collected or created as part of the day-to-day business of government	• Whitehall departments • Government agencies and non-departmental public bodies • Local government institutions	• Departmental accounts and organisational structures • Registers of public property • Public spending data • Non-personal data related to the health and education systems • Calendars of public holidays and other key administrative info
Geographic and other data describing the physical and environmental characteristics of the UK	• Emergency services • Military • Coastguard • Mountain rescue • Environment agency / departments	• Maps and charts • Meteorological data • Hydrographical data • Postcode and address data files
Administrative data describing the business environment and public realm	• Tax and welfare administration • Planning authorities • Courts and justice system	• Company registrations • Land and property registrations • Planning data • Vehicle and mileage data
Data related to public transport networks and transit systems	• Public transport operators • Public infrastructure bodies	• Timetables • Fare schedules • Planned and emergency works • Transit loading data
National statistics describing the UK economy, state and society	• Whitehall departments • Government agencies and non-departmental public bodies • Local government institutions	• National accounts • Surveys of attitudes and behaviours • Other statistics collected for analysis and / or the public record
Data created in the course of conducting publicly-funded research	• Universities and other educational establishments	• Various mass datasets • Published research

The common thread running through all of this data – be it reference or real-time, geographic or financial – is that it is integral to the efficient and effective discharge of the duties of public sector bodies. If this data did not already exist then we would expect the state to assemble it. Unsurprisingly then, the primary consumer of this wide range of public data is the public sector itself.

4. The best price for public data is free

Data has many interesting economic characteristics. Two factors are especially important for this discussion. First, public data typically involves high fixed costs – so high, in fact, that it would not normally make sense for multiple bodies to collect or create the same data. Second, digital data can typically be duplicated at very low cost – so low, in fact, that in many cases zero is a good approximation for the marginal cost of data provision.

Taken together, these two factors have significant implications for any government that is considering reselling public data.

The cost structure described above implies that public data provision is a natural monopoly.[3] So if the state seeks to optimise the direct financial return from data provision, it will set a price for users high enough to (a) recover all of its costs and (b) maximise supernormal profits. This may be very profitable indeed for the public bodies concerned. But it imposes a deadweight loss on the economy, as too many users are priced out of the market.

From a public policy perspective, a superior approach is to set price equal to marginal cost. This eliminates deadweight loss and ensures allocative efficiency. The problem, of course, is that in our case it would also leave the data provider running at a loss. In these circumstances there are two broad ways through. The first is for the state to cover the fixed costs of data production, funded out of general taxation and accepted, like other fixed costs in the public sector, as part of the general financing requirement for executing the public task. The second is to find a way to levy a minimally-distortive charge on data users (often done through average-cost pricing, but preferably by implementing some form of Ramsey pricing to focus charges on inelastic activities).[4]

We will return to the practical implications for governments looking to move from charging for public data to providing it for free later in this note.

[3] See Baumol (1977) for an exposition of the natural monopoly concept
[4] See Ramsey (1927) for the seminal discussion of maximising consumer surplus subject to monopoly cost recovery

5. Free public data supports competition and innovation

Allowing public data to be duplicated and used for free will help ensure that there are no missed opportunities for public data to add value to private sector activities. This approach is the best way for a country to sweat the data assets that its government has already collected on behalf of its citizens.

Precisely how free public data is used will vary from firm to firm, and is not something that the state can predict or should be overly concerned with. Typical applications might include firms increasing the efficiency of their operations – using maps and transport data to schedule more efficient delivery routes, or using macroeconomic projections and consumer data to forecast future demand. All of this is good for the economy. Millions of business owners, shareholders, customers and the environment all stand to benefit in one way or another.

Some firms will go further and orient their entire business model around building commercial value on top of public data sets (for some early examples see the table later in this chapter). Subject to the usual concerns about market power and illegal activities, this again is not something for the state to be overly concerned about. Where these businesses earn a sustainable rate of return they pay tax, provide jobs and make an important contribution to our overall prosperity.[5]

Many of the applications of public data will deliver incremental improvements to existing business processes, products and services. Some applications will be truly disruptive, where the firm in question brings an entirely new proposition to market. Successful, radical innovations may be one in 100 (or maybe even fewer). The best way to promote them is to roll the dice as many times as possible. In any economy where information is increasingly important, free public data is an important precondition for this to happen.

As the vast majority of the economic benefit from open public data will accrue in small gains spread across a very large number of firms, simply counting the number of businesses built around open data is not a sufficient metric to gauge impact. Nevertheless, to give a flavour of what is happening, the table below provides a handful of examples of the new breed of non-governmental businesses and organisations using data as an input to their activities.

---

[5] Of course, in the internet age there is no reliable way to restrict access to UK public data to UK citizens or businesses. We are relaxed about this: if the marginal cost of providing data is close to zero, then access by those based outside the UK does not impose an additional cost on the UK taxpayer. Indeed, all countries stand to benefit from each others' open data initiatives, as new insights flow from data shared, compared and analysed across countries.

Table 2: Data businesses and organisations		
Domain	**UK examples**	**Worldwide examples**
Business and consumer sphere	• BusIt: helps users in London plan how to make their journey by bus • Google Maps: helps users in London plan their journey by public transport • Chomaroma: takes your Oyster travel data and makes it into a game where every journey counts in a competition for the city • Placr: a UK public transport data aggregator, providing feeds to third-party apps and websites and bespoke travel services • ScraperWiki: an online tool for extracting useful data from the internet to be reused in apps or investigated by researchers and journalists • Spotlightonspend: a managed service that helps local authorities and other public bodies put their spending data online	• ITO World: provides web-based services for transport professionals and transport users • iSOCO: enriches company databases with additional data available online, especially from open repositories • Boreda: Provides real estate information using open data from the Swedish National Board of Housing, Building and Planning on building energy efficiency • AMEE: aggregates and automates access to the world's environmental and energy information to help businesses with environmental insights • OpenCorporates: provides free and open company data on over 30 million companies worldwide
Accountability, education and engagement	• OpenlyLocal: provides users with information on all UK councils and over £14 billion of spending data • Where does my money go?: uses public data to show users how and where their taxes are spent • Rewired State: organises and runs hack days, where data and software prototypes are developed in just one or two days	• Gapminder World: promotes the use and analysis of facts in education about global development • SeeClickFix: provides US citizens and local government with a portal to track local issues • OpenSpending: enables users to explore, track and analyse government spending data • KelQuartier: provides free access to factual information on 42,000 neighbourhoods and small towns in France

We have also heard (anecdotal) reports of projects that never made it into development, or took a different turn to the route originally envisaged, because the cost of access to the necessary public data inputs proved prohibitive. By definition it is impossible to count how many ideas or business propositions never made it off the drawing board due to difficulties accessing data.

6. Making this work in practice: open public data

For public data to power real-world economic activity we need a way to operationalise the economic ideal of free public data, which implicitly assumes no hidden or transactions costs. The Open Knowledge Foundation provides a benchmark in their Open Definition, which they summarise as follows:[6]

[6] Open Definition

- A piece of content or data is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and share-alike

The UK government has adopted the spirit of this definition in its own Open Government Licence, under which users are free to:[7]

- Copy, publish, distribute and transmit the information
- Adapt the information
- Exploit the information commercially

(Subject to requirements to acknowledge the source of the information, to not mislead users or misrepresent the data, and to remain within the confines of the law.)

Technology also has a part to play in delivering open public data. Data needs to be published on the web, be machine readable and be encoded using non-proprietary formats. All of these factors will help minimise the barriers to the use and reuse of data. The pragmatic use of application programming interfaces (APIs) in particular matters for data that is too expansive to copy in its entirety or changes on a very frequent basis (e.g. high resolution maps and live public transport data respectively). Looking forward, ensuring conformity with standards on linked data and the semantic web will be increasingly important for innovative and creative uses of data to flourish.[8]

7. Where we are today

From the outset, this government has promised to be the most open UK administration ever. The very first paragraph in the Conservative *Technology Manifesto* pledged that "we will legislate to enforce the freedom of government data... [and] create a powerful new right to government data, enabling the public to request – and receive – government datasets".[9] The Liberal Democrat policy paper *Preparing the Ground*, carried at the 2011 party conference, reaffirmed the position that "there should be an assumption that public non-personal data belongs to the nation, so should be freely available".[10] These sentiments are reflected in the Coalition's *Programme for Government*, which pledges to "set government data free" and "create a new right to data".

---

[7] Open Government Licence
[8] Semantic Web
[9] Conservative Technology Manifesto, Conservative Party, 2010
[10] Preparing the ground: stimulating growth in the digital economy, Liberal Democrats, 2011

The Cabinet Office has since set out what it refers to as the six opportunities of open data:[11]

- Accountability (the public scrutinising politicians and public bodies)
- Choice (helping people make informed decisions when they consume public services)
- Productivity (benchmarking costs and other performance metrics)
- Quality (boosting engagement and enabling outcomes-based commissioning)
- Social growth (supporting localism and a more informed public debate)
- Economic growth (enabling a new market based on public sector information)

Following consultations led by the Cabinet Office and the Department for Business, the Chancellor of the Exchequer used his 2011 autumn statement to announce the creation of a new Open Data Institute (directed by Tim Berners-Lee and Nigel Shadbolt), along with the release of previously unavailable datasets from the Met Office and public transport providers. The government is also establishing a new Data Strategy Board and Public Data Group to guide the development and release of public sector data.[12]

These announcements are the latest developments in an ambitious programme of openness and transparency initiated by the current administration (and building on initiatives put in place by the previous government, including the Freedom of Information regime and re-use of public sector information regulations). The most high profile examples of action so far have been related to government spending – in particular moves to publish detailed data on senior civil service salaries and the COINS database of public spending decisions. The journey has already thrown up important lessons: for COINS in particular although large volumes of raw data were released, there was significant demand for reports and summaries in user-friendly formats.

The Cabinet Office has marshalled much of the central government data released by Whitehall departments onto a single website at data.gov.uk, powered by the same technology underpinning PublicData.eu and the IATI registry.[13] At the time of writing the UK government's central data portal contains over 7,900 distinct datasets.[14]

---

[11] Making open data real: a public consultation, Cabinet Office, 2011
[12] Further detail on open data measures in the autumn statement, Cabinet Office, 2011
[13] CKAN
[14] Data.gov.uk, HM Government

Large volumes of data have also been released locally. In 2010 the Royal Borough of Windsor & Maidenhead was first to publish linked local spending data online, and has since been followed by over 30 other local government bodies on the SpotlightOnSpend.org.uk portal.[15] In London the Mayor's office has prioritised the development of the London Datastore, which hosts a wide range of datasets related to the capital.[16] In Manchester a range of public bodies have come together to collect and publish their data under the DataGM banner.[17]

The next step in the government's transparency programme looks likely to be the publication of an open government white paper in the course of 2012, building on the open public services white paper published in 2011.[18] This would seem like a natural place to consolidate the government's position in the wake of the Chancellor's autumn statement and the two consultations carried out during the second half of 2011: *Making open data real* (published by the Cabinet Office) and *A consultation on data policy for a public data corporation* (published by the Cabinet Office and BIS, in its role as the parent department for the Shareholder Executive).[19,20]

8. It is time for a statutory right to public data

In this note we have argued the economic case for open public data. Beyond the benefits for consumers, businesses and the economy, open public data can contribute to a strengthened society and enhance democratic accountability. For the UK to gain the most from open public data, the people using it need to be confident that all public data is and will remain open by default (and not only open when this is convenient for ministers or officials).

Access to open public data needs to be a right, not a privilege.

We therefore recommend that the government, in the lifetime of this Parliament, should enshrine a comprehensive right to public data in legislation.

[15] Spotlight on spend
[16] London Datastore
[17] DataGM
[18] Open public services white paper, Cabinet Office, 2011
[19] Making open data real: a public consultation, Cabinet Office, 2011
[20] A consultation on data policy for a public data corporation, Cabinet Office and Department for Business, 2011

This should require that all non-personal data collected or created to support the day-to-day business of government be made open: easy to access and free at the point of delivery, without restriction on use or reuse. Important protections for personal data, national security and Ministerial advice should be incorporated into this legislation, to provide clarity on where open data ends.

This would deliver on the spirit of the Coalition Agreement, which stated that "we will create a new right to data so that government-held datasets can be requested and used by the public, and then published on a regular basis".[21]

We recognise that moving to an open default for public data will take time. Importantly, although there are important technical aspects to resolve around government systems and software, the primary challenge is not technological but cultural. This takes us to the second of our three recommendations.

9. Strong leadership is required to manage culture change across the public sector

The culture change required to execute an open default for public data should not be underestimated. Experience with the Freedom of Information Act has shown how resistant some parts of the public sector can be to opening up to the public.[22] More generally the challenge is to shift mindsets and behaviours on data from need to know to need to restrict (where the latter category would be drawn as tightly as possible, to protect personal data and national security).

Leading successful change will take a combination of:

- Fostering understanding and conviction: helping public servants, at all levels of every organisation, to understand why the move to open public data matters for their organisation and business unit, and to want to make the transition
- Building skills and capabilities: in particular ensuring that public sector bodies have the talent and teaching required to deliver open public data and work effectively with the technology and processes involved

[21] The Coalition: Our Programme for Government, HM Government, 2010
[22] Freedom of information: the first five years, UK Parliament research briefing, 2010

- Identifying and making the most of positive role models: celebrating examples of organisations, teams and individuals doing the right thing on open public data, and ensuring that senior leaders walk the walk on open public data even when other pressures bear down on their time and attention
- Formal rewards and sanctions: recognising activities related to open public data when setting work objectives and carrying out performance appraisals, rewarding public servants when the job is done well and not shirking difficult conversations when there is room for improvement

Like many large organisations, Whitehall is traditionally – though of course not uniformly – bad at this sort of broad and fundamental change management. We do not see a role for further bureaucracy to attempt to impose top-down rules across a diverse range of government departments, agencies and other public sector bodies. Instead we believe that existing structures should be more clearly focused on enabling an open default for public data.

We therefore recommend that responsibility for open data policy should rest at board level for every public sector body, with a CIO or other board member who will champion open data.

A statutory right to data will be most effective when mindsets and behaviours in the public sector are geared toward open ways of working. Leading and inspiring a culture of openness will be a greater challenge than simply implementing new technology. Senior officials will need to ensure that that a right to data is adhered to in spirit as well as in the letter of the law, and will have a role to set and enforce minimum standards on accessibility and privacy protection.

An important part of this remit will be to upgrade the public sector's concepts of risk and opportunity cost around open public data. During the transition to open public data there will be many instances where specific datasets are not as correct, complete or visually appealing as their owners might like. Our conversations with developers confirm that many suspect that this is the case, and nevertheless would rather have the data published and iterated than held in a lengthy queue for review. Consequently, there is an important role to play in monitoring incentives – public sector managers should have more to fear from failing to open up public data than from opening up public data that is initially of substandard quality.

Beyond this, other priorities for public sector open data leaders might include working together (or perhaps through the new Data Strategy Board) to:

- Propose some common standards for how open public data is delivered, including on acceptable digital formats and API structures. Public sector bodies should of course be free to exceed these standards where doing so is cost effective.
- Consider how best to evolve data.gov.uk into a common repository or directory of open public data from across the UK public sector. This might include hosting or linking to derivative datasets created by users so that they are easy for others to find.
- Help data owners to avoid inadvertently revealing personal details when delivering open public data (we return to this and related issues in a later section of this note).

The final responsibility for this group should be to ensure that the public sector's data activities are clearly focused on collecting and creating only the data required to execute the public task. This takes us to the last of our three recommendations.

10. Good open public data will follow from a clear focus on the public task

The current public sector data landscape involves a number of bodies (including in the newly consolidated Public Data Group) that exist primarily to support the public sector but are also engaged in commercial activities. We do not dispute that many of the products and services these organisations develop and sell are both interesting and useful. But locating this activity in the public sector ultimately has two major drawbacks.

Firstly, conflating public and commercial activities risks distorting price signals within the public sector. In organisational setups where public servants can be deployed onto commercial activities, it can become hard to separate and understand the cost of the minimum data activity required to execute the public task. Put another way, there may be an incentive for public sector managers to gold-plate data products on the basis that part of the cost can be passed through to commercial customers. Whilst this may deliver *good* products, it is not clear that it delivers the *best* products for the taxpayer or for private sector data users.

Secondly, as in other aspects of economic life, public sector production happens at the cost of crowding out private sector activity and innovation. The incumbent public sector producers inevitably have a very strong advantage in the marketplace, which unavoidably deters competition. A lack of competition in turn translates into weak incentives to innovate – and even where innovation does happen it may not be focused on the activities that will deliver the greatest economic benefit.[23]

More fundamentally, engagement in commercial business ventures risks distracting public sector bodies from their primary duties as providers and enablers of public services.

We therefore recommend that every public sector body should clearly define its public task and the data requirements that this entails, and then focus ruthlessly on this remit.

Once this is done, any activity based on leveraging public data to develop commercial products or services should, ultimately, be spun out. This will ensure that the entities undertaking this activity are exposed to the full rigour of market competition. Where sound business models underpin strong commercial propositions, these new private entities will thrive and customers will benefit from businesses that have a strong incentive to meet their (changing) needs.

The taxpayer may even enjoy a windfall gain as they are privatised, and the remaining slimmed down public sector data owners will be able to run leaner operations with fewer sales, marketing and commercial staff (the Ordnance Survey accounts, for example, show around 150 people employed in sales and marketing).[24]

We recognise that pursuing this agenda will take time. The steps outlined above on open public data will be an important first step, as they will *inter alia* help level the playing field for firms looking to build their own commercial products around public data.

The end goal must, however, be to remove commercial activity from the public sector. The playing field will never be truly level while private firms have to compete with products developed in-house at the organisation that also collects or creates the raw public data.

---

[23] As evidenced, for example, by hundreds of forum pages discussing potential improvements to Ordnance Survey products and services – but which no entity other than the Ordnance Survey has the power to implement.

[24] Annual report and accounts , Ordnance Survey, 2010

To be clear: we are not suggesting that all public data activity should be privatised. Indeed, we argue in this note that there is a strong efficiency case for the public sector to collect or create the data it requires to execute the public task. But where products or services are built on top of this data, we know that once the early development phase is over, a competitive market is likely to do a much better job than a public sector monopoly at driving cost down and quality up.

11. How to measure and benchmark progress

As we move from the status quo to open public data, it will be important for leaders in the public sector to understand their organisation's performance and track progress.

Simply counting the number of datasets published, hits on a data website or downloads of different files will not be sufficient. Although these sorts of metrics are (relatively) easy to get hold of, when they become targets or success criteria the impact on incentives in an organisation can be perverse and counterproductive.

Instead, the leaders of public sector bodies, perhaps working with the Data Strategy Board, need to track a much deeper sense of the extent to which the principle of open public data is being adhered to in their organisations. To get the ball rolling we propose looking at performance and progress through three lenses:

- Leadership and culture: to what extent is open public data embedded in the mindsets and behaviours of the people in the organisation?
- Accessibility and technical: how far has the organisation leveraged proven technical tools to deliver on its open public data agenda?
- Community satisfaction: how well is the organisation's open public data activity responding to its users and meeting the needs of individuals and businesses?

We also propose tracking explicitly the gap remaining to be closed on (a) providing public data for free at the point of use, and (b) ceasing or spinning out commercial data activities.

Figure 1 illustrates how these criteria could be operationalised into a light-touch public data scorecard for assessing the progress of individual public sector bodies. Public sector bodies could use this or something like it to measure their own internal performance, and as a framework for an independent assessment of their open public data activities.

Figure 1
Open public data scorecard for <INSERT ORGANISATION NAME>

Policy Exchange

Key performance indicators

	LEVEL 1 SURVIVING	LEVEL 2 ENGAGED	LEVEL 3 ADVANCED	LEVEL 4 DISTINCTIVE	STATUS 1 to 4
Leadership and culture	☑ Responsibility for organisation's public data assets identified	☐ Senior staff are clear on the challenges ahead and their role in leading change ☐ Open data strategy clearly reflected in business plan and annual report	☐ Open public data strategy published and continuously refined ☐ Open public data principles built into all staff objectives and appraisals	☐ Staff dissent if others' actions or behaviours conflict with open data ☐ Culture fosters collaboration and innovation on data	#
Accessibility and technical	☑ Organisation complies with FoIA and data protection legislation	☑ Key datasets accessible online ☑ Process in place to deal with data requests efficiently	☐ Most datasets accessible online ☐ Open and machine-readable formats adopted as default ☐ Future technical milestones set and published	☐ All public data accessible in (close to) real time ☐ Linked data standards met ☐ Open dialogue with community on future technical improvements	#
Community satisfaction	☑ Public sector bodies respond to data requests	☐ Timely and helpful responses to public requests for data ☐ Online data easy to find and navigate ☐ Crystal clear what is available, what is not, and why	☐ Community consulted on priorities for change ☐ Policy stable enough for private firms to develop around open data	☐ Open dialogue with public sector on future of public data ☐ Open public data is the norm, returning to restrictions unthinkable	#

Data freedom

OBJECTIVES	STATUS & TARGET
• No charge for any data collected or created to support the public task	**Not met Target: 2012**
• All commercial data activities ceased or spun out into separate entities	**Not met Target: 2013**

Overall assessment

A
B
C ▶
D
E

Summarise the overall position for this organisation with respect to open public data, and identify the key priorities for (a) immediate improvements and (b) long-term success in embedding open public data

Sign off (independent): <XXX>

Updated: <INSERT DATE>

ILLUSTRATIVE

We recognise that the "distinctive" open public data performance aspiration on this scorecard is very high. We believe it is right, however, to set ambitious goals for progress from the outset so that everyone is clear about what public sector bodies should be aiming for.

The Cabinet Office seems well placed to take this forward in the first instance.

12. How to manage the Exchequer impact

The economic analysis in this paper makes the general case that the overall net economic benefit of moving to open public data will be overwhelmingly positive. Indeed, econometric attempts by others to quantify the general benefits of open public data – in terms of increased output, employment, efficiency and the like – have delivered estimates in the order of £ billions per annum.[25] Broadly speaking the calculations involved use estimates of potential revenues from reselling public data, the proportion of sales going to other parts of government, the price elasticity of demand for data and a multiplier effect when this data is made open. The welfare gains include new products and services built on open public data, an expansion of complementary products and services, reduced transaction costs and more efficient operations for data users, and gains from more timely and accurate information for the public sector.

It is important to remember that these estimates are top-down. They do not – and cannot – attempt to individually identify and add up all of the different ways in which businesses might benefit from open public data. This is as much a practical problem as anything else: it is difficult to count up the businesses that don't currently exist or focus on data, but which might start up or grow if public data were more open.

Regardless of the broad potential benefits of open public data, there is no escaping the fact that the status quo in the UK currently involves some public sector bodies reselling public data, and earning revenues from commercial activities as a result. In the current fiscal climate, questions about the affordability of moving to open public data are therefore unavoidable.

---

[25] Quantifying the range of diffuse benefits involved is difficult. The study quoted in the Conservative Technology Manifesto put an estimate of the net benefits from opening up public data in the UK at around £1.6 to £6 billion per annum – the figure at the top of this range being the one most frequently quoted, see Pollock (2009, 2010). Related estimates made by extrapolating from experience in other developed countries are not out of line, see Vickery (2011), and the overall benefits from the European Commission's new open data strategy are estimated at around €40 billion per annum, see Digital agenda: turning government data into gold, European Commission 2012.

Our best estimate, building on existing work done in this area, is that the recommendations in this paper can be implemented with minimal impact on the public finances. To understand why, it is helpful to sort the issues around the potential Exchequer impact into three categories:

- Moving from closed public data to open public data
- Moving from reselling public data to providing it for free
- Spinning out commercial activities

From closed to open

The vast majority of cases (by volume) will fall into the first category, i.e. where public data is currently collected or created to support delivery of the public task, but not made more generally available. In these cases a move to providing open public data does not jeopardise any existing revenue streams.

We recognise that there may be some transition costs if technical and systems changes are needed to deliver open data. But for publishing raw data the basic system requirements are usually low. If change is unavoidable, developments should use open standards and be implemented using agile techniques to keep costs as low as possible – with open data requirements built into the design specification for new public sector systems as standard.

From charging to free

Some important cases, however, fall into the second category. Here the transition to open public data will necessarily eliminate the revenue stream that the government currently earns from reselling public data. For the main public data bodies producing core reference data, we estimate the total potential loss of data revenue, if no mitigating action is taken, at perhaps £50 million per annum (and of course the people currently paying for this data would gain by the same amount, as they would no longer be charged).

In most cases the revenue at stake from the pure resale of basic public data is small, perhaps at or below the £1 million per annum mark for the Met Office, Land Registry, Companies House and DVLA.[26] These figures are just a fraction of one per cent of the total cost of operations, and should be manageably absorbed into general public funding for these bodies.[27]

The most frequently cited exception is the Ordnance Survey, where commercial and consumer sales in 2009-10 totalled around £39 million (and about £10 million was revenue from paper maps).[28,29] As outlined earlier in this note, there are two broad approaches to closing the gap whilst maintaining marginal cost pricing.

The first is to close the gap by extending the public funding already provided to the organisation. For the Ordnance Survey, trading revenues (from both public and private customers) are currently around £113 million, so with no other changes this would mean increasing the £74 million public funding component by around 50 per cent to keep the organisation's finances in balance.[30]

The second is to close the gap by switching charges onto activities where the price elasticity of demand is lowest. This would not entirely remove costs from the private sector, but would aim to shift them onto a less distortionary basis. As a result, fees and charges would be less of a barrier for those wanting to make use of the data.[31]

Spinning out commercial activities

The calculus of spinning out commercial activities related to public data is complex, and (ironically) the data that is currently available on public sector activities in this area is not sufficiently granular for us to attempt a quantification of the likely Exchequer impact. We can however make the following general observations:

---

[26] Models of public sector information provision via trading funds, Newbery et al, 2008.

[27] For completeness, much of the turnover for these organisations comes from other sources, e.g. charges for registrations and updates, or fees for other commercial services that the entity provides.

[28] Annual report and accounts, Ordnance Survey, 2010. The split of paper map sales between public and commercial users is not broken out.

[29] The Met Office also earns commercial revenues, perhaps in the region of £32 million a year, though it is unclear how much of this is attributable to raw data. Announcements on opening up Met Office data were made as part of the autumn statement package. See Annual report and accounts, Met Office 2011.

[30] This estimate may be on the high side – if some of the revenue-generating activities were spun out into the private sector, the increase in public funding required for the remaining entity could be considerably smaller.

[31] One potential candidate that merits attention is registration charges, i.e. applying a (statutory) charge every time the master data set needs to be updated due to building or land use changes. One estimate suggests this might level out at an average data charge of around £50 to £100 on each planning application made, see Funding options for trading funds and other PSI holders, Pollock

- The government may enjoy a windfall gain from privatising these activities. Provided they are commercially viable as private entities, they should be attractive to investors.[32] We recognise that the precise timing of divestment will need to be considered carefully, to ensure that broader market conditions do not artificially constrain the potential sale proceeds.
- The windfall gain from privatisation may, however, be below the net present value to the Exchequer of the state continuing to run commercial operations. This will be the case if opening these activities up to competition drives out any monopoly profits that the public sector is currently earning (of course the net benefit for the economy as a whole would still be positive, as competition eradicates deadweight loss).[33]

The combination of these two factors most probably implies a neutral or (small) negative impact for the Exchequer in present value terms, albeit with a one-off upfront gain. Offsetting this:

- Price signals in the public sector will be more transparent, as there will be greater clarity on the cost of collecting or creating data to support the public task. This should help public sector bodies to focus on collecting and creating data only when it is truly required, and to avoid the temptation to gold-plate data requirements. It should also encourage public sector managers to seek out efficiencies and synergies on data-driven execution of the public task (provided they are properly incentivised to be prudent with public funds).

This countervailing factor should deliver a positive benefit for the Exchequer. In the absence of more detailed information our best estimate is that the combined impact of these effects in present value terms will be broadly neutral.

This topic clearly merits further investigation, starting with an examination of the commercial activities of major organisations in the Public Data Group. HM Treasury and the Shareholder Executive / Department of Business seem well placed to take this forward.

[32] And if they are not commercially viable and not required to execute the public task, it is clear that the taxpayer should not be covering the cost of keeping them going.

[33] This assumes that the outcome is a competitive market rather than simply replacing a public sector monopoly with a private sector one. We would look to the general competition and markets regime, overseen by the Office of Fair Trading, to protect consumers against the accumulation and abuse of market power.

13. Important protections and the limits of open public data

We have argued that a move to open public data, including a statutory right to data, will deliver important economic and social benefits for the UK. Open public data is not the same, however, as an open-ended requirement for all of the information that the government creates or holds to be made available for public scrutiny.

Although the focus of this note has been on non-personal public data, there are three closely-related areas that merit additional consideration: national security, policy advice, and privacy.

National security

Clearly there will be some non-personal data collected or created by the state that does not belong in the public domain. Where making a dataset available would pose a material risk to national security, the data in question should be exempt from any open public data requirements. In some cases, acknowledgement that the data exists at all may give rise to security concerns – in which case it may be acceptable for the government not to comment.

We believe that legitimate national security concerns are the only substantive justification for withholding non-personal data collected or created to support the public task. At the boundary there are risks in both directions: for sensitive data to be released inadvertently, and for more data to be held back than is really justifiable. The existing machinery of government and Parliamentary mechanisms designed to provide oversight of the security and intelligence services should therefore have a strong role in policing data policy in this area.

Policy advice

Many aspects of public service, in particular in Ministerial departments, require public servants not just to execute public policy but to advise on potential changes to policy itself. It is essential that this advice is based on evidence (and this is recognised explicitly in the HM Treasury Green Book and other government guidance).[34,35] Good advice will usually also incorporate an element of expert judgment, drawing on the knowledge and experience of the officials involved. This makes it important to protect space for a free and frank discussion of the issues and options.

---

[34] The green book: appraisal and evaluation in central government, HM Treasury, 2003
[35] The magenta book: guidance for evaluation, HM Treasury, 2011

A protection of this sort is built into the Freedom of Information Act, where Section 35 provides a potential exemption for disclosures that relate to "*the formulation and development of government* policy" and Section 36 a potential exemption for information that would "*inhibit the free and frank provision of advice or exchange of views, or otherwise prejudice the effective conduct of public affairs*".[36]

In this note we defined public data as any non-personal data collected or created as part of the public task. This will often go beyond raw data to include information that has been cleaned up, labelled or mapped in order to aid comprehension and further analysis. Clearly data may be processed or combined to differing extents – but wherever the process is properly auditable and the data is essentially factual (or positive, i.e. statements about "what is" without approval or disapproval) then it should be in scope for open public data treatment.

At the other extreme, we suggest that information where the primary content is subjective (or normative, i.e. statements about "what ought to be" which cannot be proved or disproved) then it should fall outside the boundaries of public data.

In practice, the evidence supporting policy advice should be thought of as public data and opened up accordingly. The advice itself should not be automatically available under the open data banner – though as we move ever closer toward open government and a more participative approach to policy making, more of these discussions may take place in the public domain.

Personal data and privacy

The distinction between non-personal and personal data is complex and not always intuitive. This matters because in many instances the data collected or created to help execute the public task will, either explicitly or implicitly, identify individual citizens.

To comply with the spirit of a right to data whilst respecting the individual right to privacy, public servants dealing with personal data need to find a way to transform it into non-personal data. The easiest way to do this is to report only very broad-brush aggregates (the average GCSE pass rate at school X, say) or to withhold the data altogether.

[36] Freedom of information act, 2000

Of course, data that is provided only in a highly abstracted form is far less interesting and useful than a more granular description of the topic in question. This has led policymakers (and companies) to develop a range of techniques that aim to preserve interesting information whilst protecting privacy. These include (but are not limited to):

- Anonymisation or deidentification: stripping personal details (e.g. names) from records
- Pseudonymisation: replacing personal details with identifiers (e.g. names with numbers)
- Obfuscation: reducing precision in the data set, e.g. replacing values with ranges
- Perturbation: introducing random errors into individual records whilst preserving descriptive statistics e.g. the mean and standard deviation
- Access and query controls: restricting what information is disclosed at runtime e.g. by denying requests that would identify individuals, or limiting the minimum size for query sets

None of these are perfect, however, and there is material risk that data subjected to these or similar techniques before being made open can still be interrogated to identify individuals. In particular, the risk from so-called "jigsaw identification" where an adversary combines multiple datasets to uncover personally identifying information, can be significant.[37]

Recognising the challenges in this area, the government commissioned a detailed review of transparency and anonymisation.[38] This concluded that transparency and privacy can be compatible – but that getting it right will require a sophisticated and mature approach to handling sensitive data. We support the recommendations of this review. In particular, we believe that the government needs to be transparent about the challenges it faces. Subjecting deidentification techniques to peer review before they are used to generate open data, and having a transparent process in place to deal with vulnerabilities as they are identified, seems like the most scientific and effective response to a complex and rapidly-evolving environment.[39]

[37] To take a trivial example: if your name is stripped out of a database but your postcode and date of birth are left intact, then armed with only a small amount of contextual information – who lives in the handful of homes that comprise a postcode area, and how old they appear – I may easily be able to pinpoint references to specific people.
[38] Transparent government, not transparent citizens: a report for the Cabinet Office, O'Hara, 2011
[39] Many of these issues also apply to policy on controlled government data sharing with trusted third parties. One oft-cited example is the sharing of (deidentified) healthcare data with pharmaceutical companies. This sort of tightly drawn data sharing, with restrictions on access and reuse, is not open data in the sense we have been discussing.

14. Epilogue

This research note marks our contribution to the debate on open public data in the UK. We have argued that there are significant economic, social and political benefits to pressing forward on open public data. In our view, there is no longer any reasonable justification for keeping publicly funded data under lock and key.

Indeed, the potential for an open data revolution to permanently transform our economy, society and polity is within touching distance. We believe that a statutory right to data is required to guarantee openness as the new default, and to provide clear protections for personal data, national security and Ministerial advice. We believe that strong open data leadership at board level in every public sector body is needed to ensure senior officials take responsibility for ensuring that the spirit of this right to data is adhered to. And we believe that requiring every public sector body to define and focus on its public task is necessary to ensure that everyone is clear about what data the public sector is collecting and for what purpose.

In the coming year we look forward to news on the future of the Data Strategy Board and Public Data Group, the launch of the new Open Data Institute in London, and perhaps even a white paper marking out where the open government agenda is headed. We will maintain a close interest in this area, and urge the government to implement our recommendations.

Policy Exchange

research note
March 2012

To follow the hyperlinks in this document, please retrieve a digital copy from the Policy Exchange website at: http://www.policyexchange.org.uk

Acknowledgements

We are grateful to a wide range of individuals and organisations for their support, encouragement and critical challenge during the production of this research note.

About the Digital Government Unit

We are helping policymakers unlock the potential of technology: for an innovative digital economy, smarter public sector and stronger society. For more information on our work programme please contact Chris Yiu, Head of the Digital Government Unit.

Email: chris.yiu@policyexchange.org.uk or Twitter: @PXDigitalGov

About Policy Exchange

Policy Exchange is the UK's leading think tank. We are an educational charity whose mission is to develop and promote new policy ideas that will deliver better public services, a stronger society and a more dynamic economy. Charity Registration Number 1096300.

Trustees