

The Big Data Opportunity



Making government faster,
smarter and more personal

Chris Yiu

 @PXDigitalGov #bigdata



The Big Data Opportunity

Making government faster,
smarter and more personal

Chris Yiu



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. Registered charity no: 1096300.

Policy Exchange is committed to an evidence-based approach to policy development. We work in partnership with academics and other experts and commission major studies involving thorough empirical research of alternative policy outcomes. We believe that the policy experience of other countries offers important lessons for government in the UK. We also believe that government has much to learn from business and the voluntary sector.

Trustees

Daniel Finkelstein (Chairman of the Board), Richard Ehrman (Deputy Chair), Theodore Agnew, Richard Briance, Simon Brocklebank-Fowler, Robin Edwards, Virginia Fraser, Edward Heathcoat Amory, David Meller, George Robinson, Robert Rosenkranz, Andrew Sells, Tim Steel, Rachel Whetstone and Simon Wolfson.

About the Author

Chris Yiu is Head of the Digital Government Unit at Policy Exchange. He directs research on public policy in the era of digital communications, high technology and big data. Chris was born and brought up in London. He holds a first class degree in economics and a master's degree in economics and finance, both from the University of Cambridge.

© Policy Exchange 2012

Published by

Policy Exchange, Clutha House, 10 Storey's Gate, London SW1P 3AY

www.policyexchange.org.uk

ISBN: 978-1-907689-22-2

Printed by Heron, Dawson and Sawyer

Designed by Soapbox, www.soapbox.co.uk

Contents

	Acknowledgements	5
	Executive Summary	6
1	Data, Data, Everywhere	9
2	Big Data 101	10
3	Why this Matters for the Public Sector	13
4	Some Practical Ideas	17
5	The Size of the Prize	21
6	Big Data Also Needs Sound Judgment	23
7	Capability	25
8	Power and Responsibility	28
9	Recommendations	30
	Postscript: What Leaders Need to Ask	33

About the Digital Government Unit

We are helping policymakers and politicians unlock the potential of technology: for an innovative digital economy, smarter public sector and stronger society. For more information on our work programme please feel free to get in touch.

Email: chris.yiu@policyexchange.org.uk

Twitter: [@PXDigitalGov](https://twitter.com/PXDigitalGov)

Acknowledgments

We would like to thank the wide range of individuals, businesses and other organisations that shared their perspectives, insights and critical challenge with us during the course of our work. Any errors and omissions remain, of course, our own.

We are particularly grateful to EMC Corporation for the support, challenge and encouragement that they contributed throughout the project.

Executive Summary

The modern world generates a staggering quantity of data – and the business of government is no exception. Across the public sector, extraordinary quantities of data are amassed in the course of running public services – from managing welfare payments and the National Health Service, through to issuing passports and driving licences. In the arena of tax alone, HM Revenue & Customs reportedly holds over 80 times more data than the British Library.¹

The term *big data* has come to refer to these very large datasets, and *big data analytics* to refer to the process of seeking insights by combining and examining them.

Regardless of the stance a government chooses on openness – i.e. decisions on making public data free to use, reuse and redistribute – an abundance of data and computing power gives the public sector new ways to organise, learn and innovate.

The purpose of this short report is threefold: to inspire policymakers around the opportunity for data and analytics to transform public service delivery, to sound a note of caution about the challenges this agenda poses for the public sector, and to make recommendations for how government might begin to realise the former whilst addressing the latter.

The opportunity for public service transformation is real. For citizens, the application of data, technology and analytics can cut paperwork, get questions answered more quickly, help people find and claim the benefits they are entitled to, and tune front-line services more closely to individual needs and behaviours. Enhanced use of data and analytics in the health arena, for example, could help ensure patients in care homes receive the right medicines at the right times, or help hospitals further personalise patient care and advice to minimise readmissions after surgery. In the welfare arena, better segmentation and personalisation could help identify the support that unemployed people need and get them into long-term work.² Here and elsewhere, smarter, more personal public services would help bring interactions with government up to the quality of experience we are often used to as consumers of commercial products and services.

At a macro level, 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). We estimate that achieving cutting-edge performance could in time save the public sector up to £16 billion to £33 billion a year – equivalent to £250 to £500 per head of the population.

Big data technologies alone are not, however, a silver bullet for transforming the public sector. Underlying data issues like quality, standards and bias still need to be recognised and addressed. And governments must have the capability to conduct, interpret and consume the outputs of data and analytics work intelligently. This is only partly about cutting-edge data science skills. Just as important – if not

¹ On the trail of the offshore tax dodgers, BBC, December 2011

² Personalised Welfare: Rethinking employment support and jobcentres, Policy Exchange, 2011

more so – is ensuring that public sector leaders and policymakers are literate in the scientific method and confident combining big data with sound judgment. We have included some initial questions for public sector leaders to ask of their organisations at the end of this report.

Governments will also need the courage to pursue this agenda with strong ethics and integrity. The same technology that holds so much potential also makes it possible to put intense pressure on civil liberties. Both governments and businesses are exposed

to tensions when attempts to extract value from data assets collide with individuals' wishes not to be tracked, monitored or singled out. Of course, for governments the motivations for accumulating data and analytics capability are less about profit; the risk is that curiosity trumps a demonstrable and defensible public policy rationale. We can and must hold our leaders to the highest possible standards.

“We estimate that achieving cutting-edge performance could in time save the public sector up to £16 billion to £33 billion a year – equivalent to £250 to £500 per head of the population”

Recommendations

1. A new Advanced Analytics Team should be established in the Cabinet Office, with responsibility for identifying big data opportunities and helping departments to realise them.

This should be modelled in part on the approach taken with the Behavioural Insights Team, with primary objectives to:

- Work with departments to transform three major areas of public policy delivery by applying data and analytics in new, imaginative and/or more sophisticated ways.
- Spread awareness, understanding and demand for cutting-edge data and analytical tools and techniques amongst senior leaders in the public sector.
- Achieve savings and benefits for central government, over and above existing plans, worth at least £1 billion.

To foster a sense of urgency the team should publish a progress review within one year, with a clear plan for how savings will be captured, and after a further year be subject to sunset review.

2. Government should adopt a Code for Responsible Analytics, to help it adhere to the highest ethical standards in its use of data and analytics.

Important elements of the code might be to:

- Put outcomes before capabilities. Data and analytics capabilities should always be acquired on the basis of a clear and openly communicated public policy justification, and never for their own sake.
- Respect the spirit of the right to privacy. Auxiliary data and analytics should not be used to infer personal or intimate information about citizens (e.g. reproductive status or sexual orientation). Where this data is needed for public policy reasons, consent should be sought explicitly.

- Fail in the lab, not in real life. A sandbox environment and synthetic data should be used to test all major big data initiatives – after which they should be subject to intense scrutiny and peer review before a decision is made on implementation.

Just because government can do something with big data, that doesn't mean that it *should* do it. In the final analysis, if a Minister would not be comfortable putting themselves or their family under the sort of scrutiny required by a departmental big data initiative, then that initiative should not make it into government policy.

1

Data, Data, Everywhere

In the long run, the choices a government makes about capturing, exploiting and releasing data may be some of its most defining economic and social decisions. A government's attitude to data sets the tone for businesses and citizens, and the framework for public policy initiatives that may persist for decades.

Earlier in 2012 we published *A Right To Data*, which examined the economic case for *open data* in the public sector.³ We argued that all non-personal data collected or created to support the day-to-day business of government should be made open: easy to access and free at the point of delivery, without restriction on use or reuse. Opening up core public data assets for non-government users to put to new, as-yet unknown uses is an essential stage in building our nation's digital economic infrastructure.

This report is about a strategy for *big data* in the public sector, and how this could make a real difference for public services and citizens. Our discussion of the challenges and opportunities around big data is distinct from – albeit in many ways complementary to – the issues around open data. Regardless of the stance a government chooses on openness, an abundance of data and computing power gives it new ways to organise, learn and innovate. It also poses new challenges on civil liberties and privacy that must be dealt with, and raises the bar on the talent and capability required in the public sector. Embracing the big data opportunity will take leadership and ethical integrity of the highest order. For the governments that succeed, the benefits will be both significant and long lived.

3 *A Right to Data: Fulfilling the promise of open public data in the UK*, Policy Exchange, March 2011

2

Big Data 101

The world's capacity to store, broadcast and compute information is growing exponentially. The numbers involved have already passed well beyond the scales we are used to in our everyday lives.

Counting across all forms of storage, from mobile phone memory to DVD, Blu-Ray and hard disks, we estimate that the world's installed capacity to store information will reach around 2.5 zettabytes this year. One zettabyte is equal to one trillion gigabytes. If we stored all this data on DVDs and piled them up, the stack of discs would stretch one-and-a-half times the distance from the earth to the moon. What's more, this figure is growing by over 50% year-on-year. If the world's storage capacity continues to grow at this pace then it will reach nearly 100 zettabytes by 2020 (see Figure 1).⁴

Computing power is also continuing to advance. Moore's law famously predicts that the number of transistors that can be fitted on a microchip will double approximately every 18 months to two years. By 2014 the combined capacity to compute information on the world's mobile phones will exceed that of all the world's supercomputers by an order of magnitude. The combined computing capacity installed across the world's videogame consoles will be an order of magnitude greater still.⁵

Against this backdrop of staggering growth in data storage, computation and broadcast capacity, businesses and governments are learning to cope with the huge quantities of data that they now generate and store on a daily basis. This *big data* revolution is arguably one of the most important global trends for the coming decade.

The literature on big data can be somewhat elastic in its suggested definitions. This is not surprising: the extremely rapid pace of progress on data storage and computational capabilities renders an absolute definition of big data in terms of bytes or bandwidth impractical. A more pragmatic definition – and the one we will adopt for this paper – is that big data refers to datasets that are too awkward to work with using traditional, hands-on database management tools. This might be the case for a number of reasons – including (but not limited to) limitations in traditional methods' ability to capture, store, manage or analyse data on the scale in question.

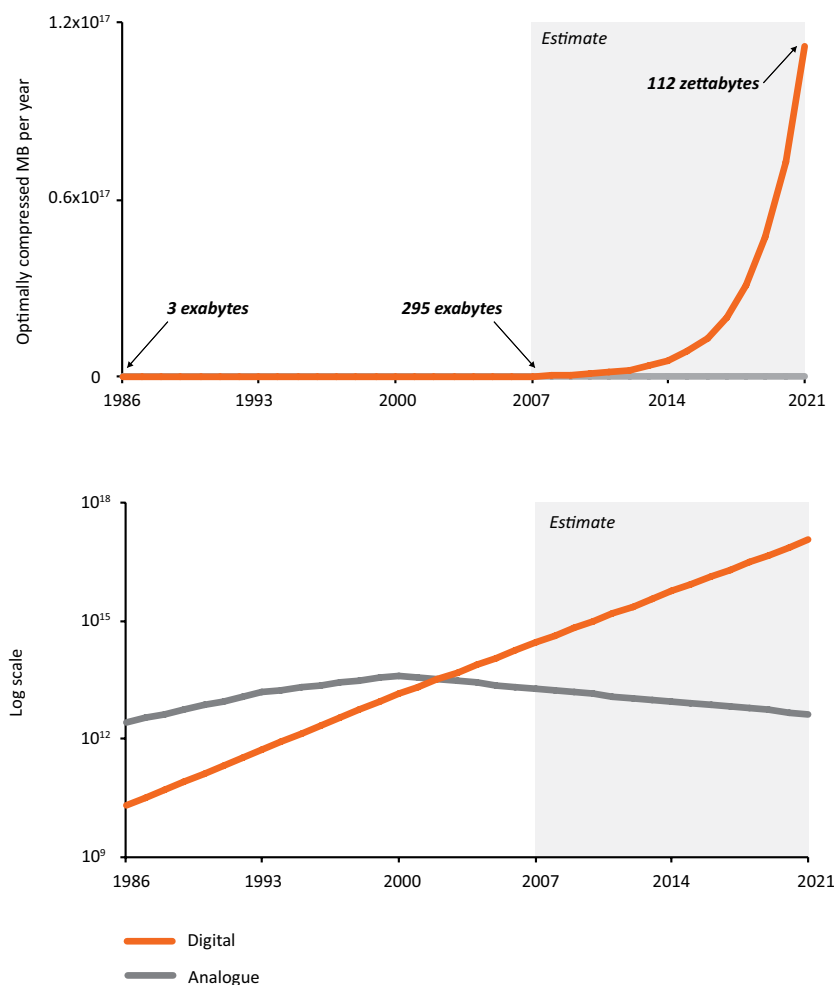
Box 1: Definitions

Big data: datasets that are too awkward to work with using traditional, hands-on database management tools.

Big data analytics: the process of examining and interrogating big data assets to derive insights of value for decision making.

⁴ Author's calculations based on Hilbert & López, 2011

⁵ Ibid

Figure 1: Growth in global storage capacity, 1986 to 2021

Source: Hilbert and López, Policy Exchange analysis

For some organisations, a requirement to deal with very large datasets has existed for some time. Companies running large retail operations, large logistics operations, or trading in the financial services sector, are typical examples of big data as part-and-parcel of running a modern business. The public sector has its own long-standing stores of big data. As a direct result of dealings with very large numbers of people and/or complex processes, organisations like HM Revenue & Customs, the Department for Work and Pensions, the National Health Service and the Met Office all have very large quantities of data distributed through their organisations.

The important shift in recent years has been leaders recognising big data as a significant source of value and competitive advantage, rather than just something that has to be coped with when operating at scale.

Examples of businesses taking advantage of big data and analytics are well documented. Many of these stories permeate our day-to-day lives – anyone who has used a Tesco Clubcard or run a Google search has dipped a toe into the world of big data (and been touched by the privacy implications thereof). The businesses that have mastered big data technologies, and the changes in mindsets

and culture that are needed alongside them, are using them to generate new sources of value for consumers and shareholders alike.

Many of the big data tools and techniques pioneered in the private sector will have a role to play in the public sector. In the United States, for example, the White House has a formal initiative up and running to use big data to accelerate the pace of discovery in science and engineering, strengthen national security, and transform teaching and learning (see Table 1).⁶

We must recognise, however, that unlocking the full potential of big data for governments is about more than investing in technology and replicating commercial innovations. Important distinctions in the objectives of public sector organisations and the unique nature of the state/citizen relationship raise the stakes significantly.

Table 1: Selected big data initiatives in the United States⁷

Area	Initiative	Description
Department of Health and Human Services	Center for Medicare & Medicaid Services	"Data visualization tools, platform technologies, user interface options and high performance computing technologies – aimed at using administrative claims data to create useful information products to guide and support improved decision-making."
National Institutes of Health	Patient Reported Outcomes Measurement Information System	"A system of highly reliable, valid, flexible, precise, and responsive assessment tools that measure patient-reported health status... provides tools and a database to help researchers collect, store, and analyse data related to patient health status."
National Institutes of Health	The Cancer Imaging Archive	"Aims to improve the use of imaging in today's cancer research and practice by increasing the efficiency and reproducibility of imaging cancer detection and diagnosis, leveraging imaging to provide an objective assessment of therapeutic response, and ultimately enabling the development of imaging resources that will lead to improved clinical decision support."
Department of Veterans Administration	Algorithms for Text Processing to Capture Health Events	"A research-based surveillance program that relies on newly developed informatics resources to detect, track, and measure health conditions associated with military deployment."
Department of Energy	Atmospheric Radiation Measurement Climate Research Facility	"A multi-platform scientific user facility that provides the international research community infrastructure for obtaining precise observations of key atmospheric phenomena needed for the advancement of atmospheric process understanding and climate models."
National Aeronautic and Space Administration	Advanced Information Systems Technology Awards	"Seeks to mature big data capabilities to reduce the risk, cost, size and development time of space-based and ground-based information systems and increase the accessibility and utility of science data."
National Science Foundation	Focused Research Group on stochastic network models	"[Seeks to] differentiate knowledge in a network from randomness. Collaborators in biology and mathematics will study relationships between words and phrases in a very large newspaper database in order to provide media analysts with automatic and scalable tools. A cross-disciplinary Ideas Lab [will] generate ideas for using large datasets to enhance the effectiveness of teaching and learning environments."

⁶ Big data is a big deal, Office of Science and Technology Policy, March 2012

⁷ Excerpts from Big data across the Federal Government, Office of Science and Technology Policy, March 2012

3

Why this Matters for the Public Sector

Properly executed, big data analytics can have a real and direct impact on the way policymakers work and citizens interact with governments.

In this chapter we provide a broad outline of how big data has the potential to improve public administration, services and the citizen experience. We have deliberately set high ambitions for how public sector leaders should be thinking about big data. Seeking revolution rather than evolution is a defining characteristic of big data visionaries. In that spirit, here are five classes of big data opportunity that are relevant to the public sector.

1. Sharing

The public sector is made up of many thousands of different organisations, ranging from large departments of state to individual schools, surgeries and libraries. Each of these organisations knows many things about its operations and the people that it deals with. Finding ways to share or link this data together has the potential to save time for citizens and money for taxpayers.

The UK does not have a single national identity database – and the Identity Cards Act 2006 was repealed by the coalition government in 2011 – so it is not possible to retrieve reference information relating to an individual citizen from a single authoritative source. With previous generations of technology, this often meant there was no alternative to storing multiple copies of the same data (or simply coping without some of the data altogether). Modern technology, however, can enable fragments of related information to be matched and linked together quickly and non-persistently. This can be used to streamline transactions – reducing the scope for errors and avoiding asking people to provide the same information multiple times.

The process for obtaining a UK driving licence is a good example of data sharing in action. When you apply for a driving licence online, the Driver and Vehicle Licensing Agency (DVLA) requires a photograph and signature for your new licence. If you have a UK passport then the DVLA will try to capture these electronically from the information already held by the Identity and Passport Service (IPS).⁸ This is often held up as an example of how simple changes can deliver practical improvements for end-users.

2. Learning

Big data analytics can be an immensely powerful tool for helping organisations to learn about how they work. Traditionally, managers and leaders have looked

⁸ Driver licensing online, Directgov

to a relatively small set of key performance indicators to assess the health and efficiency of their organisations. Digitisation has massively increased the quantity of management information available, the resolution and frequency at which it is captured, and the speed at which it can be processed. Data on inputs, outputs, productivity and processes can all be captured and recalled in more comprehensive detail than ever before.

Of course, all this information is useless unless it is used to generate insights that leaders can act on. Fortunately, advances in analysis and visualisation tools (interactive charts, infographics, deep zooming applications, etc.) mean it is now feasible to bring granular and up-to-date evidence to bear on leadership challenges. This applies across the board – from analysing and optimising the performance of different business units, through to gathering and acting on feedback from citizens on service delivery.

In many instances, important sources of big data for learning live outside traditional organisational boundaries. One increasingly important source is the information shared publicly via social media. In the business sphere, leading organisations like Procter & Gamble use cutting-edge tools to scan for relevant feedback and comment, which can then be sent straight to the screen of the individuals that need to see it.⁹ This is agile, real-time learning at its best – and a far cry from the traditional approach of reading and replying to pen-and-paper correspondence and management memos as the primary source of data on how an organisation is working.

3. Personalising

One of the defining features of big data is its granularity. We have moved from an era of knowing things at a very macro level to knowing things at a very personal level. Once, I might have known that the average adult in the UK burns around 1,800 calories a day. Now, armed with the latest personal tracking technology, I can also know that I have burned precisely X calories in the last day, putting me in the Yth percentile of the distribution for my gender and age bracket, and representing a change of Z% on my average for the previous week.

This granularity in big data opens up new possibilities for personalising services. When a service provider knows something specific about a user then there is an opportunity to tailor the service offered accordingly. This will be most useful when the data in question relates to the user's needs, and when the personalisation is done in a way that is salient for the transaction being undertaken or service being used. In healthcare, for example, data and analytics might be put to good use helping people to avoid readmission after surgery, based on an analysis of risk factors related to an individual's circumstances. This sort of personalisation to achieve an outcome is what matters; superficial personalisation or irrelevant cross-selling is at best pointless and at worst counterproductive.

Once again, in the commercial world the power of big data for personalisation has been used to great effect. Amazon is famous for pioneering collaborative filtering – the system that generates highly personalised recommendations based on its knowledge about purchase histories, product ratings and reviews. Similar principles are at work when Facebook and other social networks serve up personalised news feeds based in part on the user's position and relationships in the social graph. The result is content and services that feel more relevant to the

⁹ Inside P&G's digital revolution, McKinsey Quarterly, November 2011

user without them having to expend the significant effort that would be required to filter and prioritise a mass of unsorted material. There are lessons here for government about how it presents large quantities of information to citizens online.

4. Solving

One of the most powerful aspects of the big data revolution is the unification of large datasets with advanced analytics for problem solving. With advances in algorithms and machine learning, data-driven insights can be extracted from information sources that are too expansive or complex to tackle using traditional desktop analysis. This ability to spot patterns and solve problems beyond human mental capabilities has led to two main sources of insight derived from big data.

First, very large and/or multidimensional datasets can be examined to look for previously hidden patterns and correlations. Sometimes this can validate positions that were previously supported by common sense, practical experience or received wisdom. On other occasions, this sort of analysis can deliver entirely new insights into the underlying dynamics of a population, market or business. The Livehoods project, for example, looks for patterns in location-aware social network activity to identify neighbourhoods in urban areas based on overlapping social patterns rather than traditional geography.¹⁰

Second, big data opens up the realm of reliable predictive analytics. By examining the relationships embedded in large datasets it is possible to build a new generation of models describing how things are likely to evolve in future. This approach can be combined with scenario planning to develop a series of predictions for how a system will respond to different policy choices. The state of the art in predictive analytics can deliver forecasts for some domains with a very high degree of precision, providing an auditable, scientific basis for making decisions in complex systems.

Using big data analytics for problem solving has other advantages beyond seeing deep and far. Increased use of computational techniques can free up an organisation's staff to focus on tasks where human beings continue to outperform computers, increasing overall productivity. And working rigorously to quantitatively link insights to supporting evidence can provide a check against any bias inadvertently introduced by the individuals involved in the process. "Evidence-based policymaking" is an oft-repeated Whitehall mantra; big data analytics can help embed this in the culture where previous efforts have failed.

5. Innovating for growth

The market for big data analytics, tools and technologies is dynamic and evolving rapidly. As part of the broader digital and knowledge economies, the businesses and organisations that are leading innovation in the big data space, along with those deploying big data in their organisations, have an important role to play in supporting economic growth.

Where big data and analytics are used to identify cost savings and increase efficiency, they can contribute to a direct improvement in productivity. By identifying areas of underperformance and reallocating resources to their most productive uses, the overall performance of the organisation in question can be improved. Where the operation is commercial this can contribute to increased

10 The Livehoods project: utilizing social media to understand the dynamics of a city, Cranshaw et al, 2012

value for shareholders. Where the organisation is in the public sector, there can be benefits for citizens both as taxpayers and as end users.

Beyond direct benefits for organisations that deploy big data analytics, there is further scope for new markets in big data related tools and products to emerge. The businesses that develop big data technologies are at the cutting edge of the digital economy – ranging from large, established players to numerous smaller and start-up companies offering innovative and niche services. The government has set an ambition for the UK to become the technology hub for Europe. An important second-order benefit of striving to capture big data opportunities in the public sector will be the economic upside from the partnerships with big data businesses that this will entail.

Of course, support for the big data industry should never drive the decision-making process for government – the primary concern must always be taxpayers and citizens. But where commercial partnerships make sense, this will send an important signal of confidence for the big data and related sectors of the UK economy.

4

Some Practical Ideas

Big data technologies make it possible to set up smart, sophisticated organisations and deliver rich, personalised user experiences. As customers, many of us experience tailored services and optimised business processes backed by big data on a daily basis. In this chapter we set out five proposals for increased use of big data analytics in the public sector that we believe merit attention.

Table 2: Potential UK public sector applications

Opportunity	Summary	Precedent
Real-time management information	<p>Routinely capture data created in the day-to-day business of government, monitor it to detect how departments are performing, and analyse it to identify opportunities to reduce waste or increase efficiency.</p> <p>Use cutting-edge data visualisation tools both as an aid to analysis and to provide senior officials, Ministers and their advisers with real-time, interactive facts and figures on public sector performance.</p> <p>To give an idea of how this might look, we have mocked up a simple visualisation example (see Figure 2). Note the inclusion of real-time data, analysis and rich interactive elements including rewind/predict and zoom in/out.</p> <p>Easy-to-consume data on tap would reduce the time and effort spent disputing the facts and free up more time for working on policy solutions.</p>	<p>Businesses with strong competition and cost pressures – e.g. retail, logistics, energy, transport, social networks – all contain examples of businesses that track inventory, performance and financials both in detail and in close to real time.</p> <p>The DCLG Open Data Communities website provides a selection of statistics on local government performance and outcomes. An interactive dashboard displays data for individual local authorities, along with how they compare to others in England.¹¹</p> <p>Google Maps already includes live traffic data as an overlay for many regions.¹²</p> <p>The National Rail Enquiries website displays live departure and arrival boards for UK rail stations.¹³</p> <p>At Gatwick Airport’s new South Terminal security area, lanes are colour coded with screens displaying live queuing times, enabling passengers to choose where they want to queue.¹⁴</p> <p>Speedtest.net allows users to test the performance of their broadband connection in real time, to compare the data with past performance and against millions of other users.¹⁵</p>

¹¹ Open Data Communities, Department for Communities and Local Government

¹² Google Maps, Google

¹³ National Rail Enquiries, Association of Train Operating Companies

¹⁴ Gatwick Airport opens new £45 million South Terminal security area, Gatwick Airport Limited, 2011

¹⁵ Speedtest.net

Opportunity	Summary	Precedent
Countering non-compliance, fraud and error	<p>Accelerate the use of analytics and data fusion to make further inroads into tax compliance, welfare and benefits fraud and errors</p> <p>This could help:</p> <ul style="list-style-type: none"> • Focus caseworkers on individuals most likely to be in breach/where the most money is at stake • Identify and prioritise the channels and communications that are most effective at ensuring compliance • Eliminate errors by reducing avoidable data entry and flagging potential errors at the point they enter the system 	<p>Government departments that administer benefits, grants and other application-based processes to obtain public funds will soon be required to subject applications to automatic anti-fraud checks before payment is made. HMRC is already using this approach in the Tax Credits regime, and the DWP are seeking to integrate a similar approach into Universal Credit.</p> <p>HMRC and the DWP are using credit reference agency data to verify claimant circumstances on a payments-by-results basis, and are seeking opportunities for online transaction monitoring to identify criminal behaviour in real time.¹⁶</p> <p>In the financial services industry, automated real-time checks for fraudulent transactions have reduced the average losses for fraud for credit card firms to about 0.1% of transactions.¹⁷</p> <p>Businesses providing online loans track and combine a wide range of data – from publicly available records to real-time telematics – to build up a sophisticated judgment about the risk a lender presents and on what terms an offer should be made.¹⁸</p>
Transforming and personalising the citizen experience	<p>Accelerate the use of analytics and data fusion to make public services feel more responsive and tailored to individual citizens and households.</p> <p>This could help:</p> <ul style="list-style-type: none"> • Ensure households receive the benefits and other support they are entitled to, but fail to claim • Respond to individual public enquiries faster and with more useful information • Detect public concerns and priorities at an early stage, and adjust policy accordingly 	<p>The analytical techniques used to combat fraud and error could be turned around to help individuals – particularly in vulnerable situations – ensure they are receiving all of the support they should be.</p> <p>In the welfare arena, better segmentation and personalisation could help identify the support that unemployed people need and get them into long-term work.¹⁹</p> <p>Advances in machine learning have allowed many companies to offload routine enquiries to virtual agents. At the cutting edge, IBM's Watson supercomputer has demonstrated significant competence in its ability to parse natural language, idiom and context.²⁰</p> <p>Sentiment analysis based on social media and other sources is already used by many businesses as a complement to traditional focus groups and media analysis. This allows businesses to respond to changing events and react quickly if customers are not satisfied.</p>
Improving health, care and patient outcomes	<p>Further optimise the use of scarce health service assets by combining and analysing large datasets to advance risk models, personalise health advice and reduce waste e.g. reducing avoidable readmissions after surgery and reducing medicine waste.</p>	<p>The Department of Health recognises the potential for health and care professionals to use connected information and technology to improve services, inform decisions and deliver safer, more integrated care. This covers a wide range, from ensuring care workers give the right medicines to the right person in a care home, through to secure data linkage services to serve research and life sciences needs.²¹</p> <p>A number of big data initiatives in the United States are focused on capturing, combining and analysing large clinical datasets as the basis for scientific discovery in the realms of cancer research, heart and lung research, neuroscience and infectious diseases.²²</p> <p>Google Flu Trends has demonstrated the ability to identify influenza trends based on changes in search activity. Its forecasts lead the Centers for Disease Control by a fortnight.²³</p>

16 Tackling fraud and error in government, Cabinet Office, 2012

17 Big data: crunching the numbers, The Economist, May 2012

18 With Wonga, your prosperity could count on an algorithm, The Guardian, October 2011

19 Personalised welfare: rethinking employment support and jobcentres, Policy Exchange, 2011

20 Watson, IBM

21 The power of information: putting all of us in control of the health and care information we need, Department of Health, May 2012

22 Big data across the Federal Government, Office of Science and Technology Policy, March 2012

23 Detecting influenza epidemics using search engine query data, Ginsberg et al, 2009

Opportunity	Summary	Precedent
Delivering more timely population estimates at lower cost	Replace the decennial census exercise with a rolling, real-time mosaic generated by combining a range of existing population datasets	<p>The 2011 census cost nearly £500 million to conduct – and if the historic schedule is kept, will not be updated until 2021.²⁴ Other datasets have partial coverage of the UK population – for example the council tax register, the electoral roll, NHS patient records, child benefit claims data and state pension entitlement data. Carefully combined and de-duplicated, these could be used to generate demographic information of comparable quality and superior timeliness.</p> <p>The Netherlands conducts a “virtual census”, where administrative registers are linked at a pre-determined time to establish an enumeration of the population. Results from other surveys provide individual characteristics that are not available in the registers, provided the information can be linked at unit level.²⁵</p>

Other opportunities that were suggested to us in the course of our review, and which could be explored further, include:

- Developing more detailed and/or interactive, personalised tax and benefits statements for individuals and households
- Adopting predictive policing tools and techniques to advance crime prevention and make best use of resources
- Accelerating smart grid investments to increase the overall efficiency of energy use by households and businesses.

²⁴ Public administration select committee looks at 2011 Census preparations, UK Parliament, 2009

²⁵ Census taking in Europe: how are populations counted in 2010?, Valente, 2010

Figure 2: Visualising public sector performance



5

The Size of the Prize

Beyond personalising the individual experience and delivering public services that are more responsive to individual users' needs, big data has the broader potential to enhance public sector efficiency.

The McKinsey Global Institute (MGI) identifies three areas where big data can be used to deliver savings for the public sector:

- Efficiency savings. By making smarter decisions about how departments are organised and what work gets prioritised, the direct cost of government operations can be reduced.
- Reductions in fraud and error. The application of big data tools can help identify sources of fraud and error in the welfare system, and target scarce enforcement resources where they will have the best payoff.
- Improvements in tax collection. Again the application of big data tools can be used to detect and prioritise actions that will help close the tax gap (the difference between theoretical tax liabilities and actual receipts collected by the government).

MGI estimate the potential savings for European public administrations at around €150 billion to €300 billion a year.²⁶

Following the MGI methodology, our estimates suggest that fully capturing the big data opportunity to drive up efficiency and cut out waste in the UK public sector could be worth a total of between £16 billion and £33 billion a year (see Figure 3 for a breakdown of the broad areas where these savings might fall).

This is equivalent to around £250 to £500 a year for each person in the country, or 2.5% to 4.5% of the government's total budget of around £700 billion.²⁷

These estimates are necessarily broad-brush. By way of comparison, HMRC estimate that the tax gap – the amount of theoretical tax liabilities that goes uncollected – is around £35 billion a year.²⁸ The Cabinet Office estimates that fraud in the public sector costs taxpayers around £21 billion a year, with a further £10 billion lost to errors and £7–8 billion in uncollected debts.^{29,30}

In the United States, a data linking and mining cooperation led by the Federal Bureau of Investigation has reduced Medicare fraud by \$4 billion.³¹

Public sector productivity in the UK has been pretty much flat for a decade or more.³² Better use of data and analytics may provide an opportunity to shift this situation, and to help meet the fiscal mandate by improving efficiency rather than reducing service levels.

26 Big data: the next frontier for innovation, competition, and productivity, McKinsey Global Institute, June 2011

27 Budget 2012, HM Treasury, 2012

28 Measuring tax gaps, HM Revenue & Customs, 2011

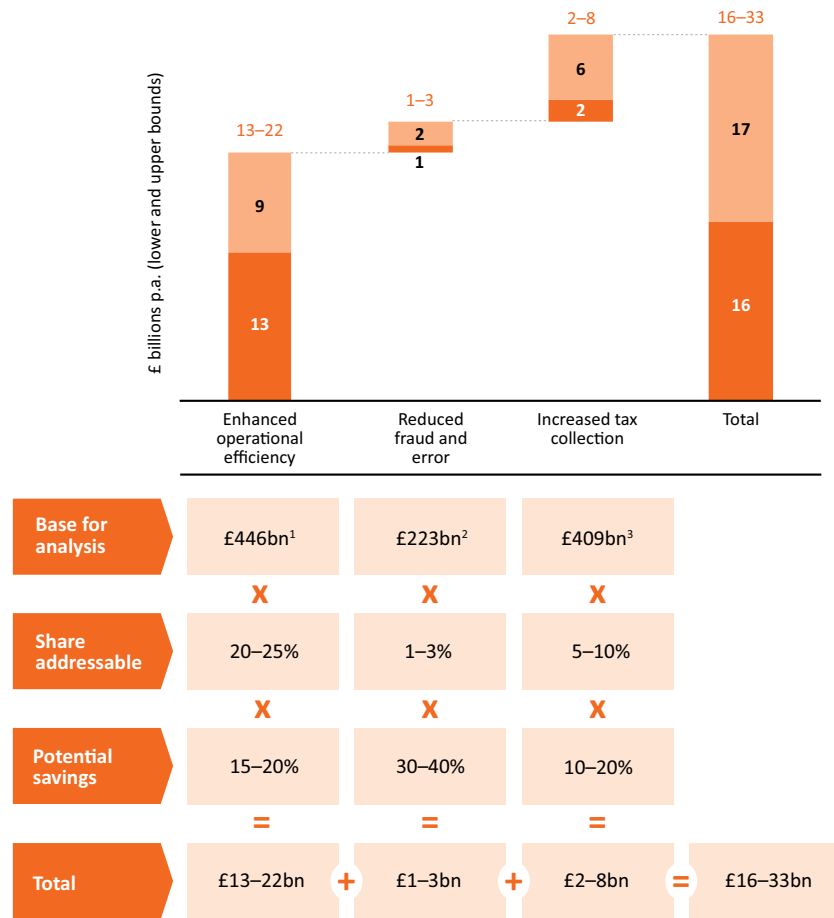
29 Tackling fraud and error in government, Cabinet Office, 2012

30 Tackling debt owed to central government, Cabinet Office, 2012

31 Health care fraud prevention and enforcement efforts result in record recoveries, Federal Bureau of Investigation, 2012

32 Total public service output, inputs and productivity, National Statistics, 2010

Figure 3: Big data potential for the UK public sector



1 Public sector expenditure on all functions excluding social protection 2009–10, PESA 2011 Table 4.2

2 Public sector expenditure on social protection 2009–10, PESA 2011 Table 4.2

3 Total tax receipts 2009–10, HMRC Tax and NICs Receipts

Source: HM Treasury and HM Revenue & Customs, McKinsey Global Institute, Policy Exchange analysis

6

Big Data Also Needs Sound Judgment

We believe that the application of big data tools and techniques has the potential to improve public service delivery and efficiency. An abundance of data and computing power does not, however, automatically guarantee good decision making.

It is important to recognise that big data is not a silver-bullet solution for all the difficult economic, statistical and analytical challenges that governments face. Leaders and analysts working with big data (or the outputs from big data exercises) must continue to pay attention to ensuring that advice is robust and properly grounded in the evidence. Particularly in the realm of public policy, we need to work with data sets that are representative, pay close attention to selection bias, and maintain a balanced perspective on the confidence we attach to data-driven inferences and the limitations of quantitative analysis. Campbell's law, Goodhart's law and the Lucas critique of policy prediction all remain relevant for practitioners.^{33,34,35}

Data quality and security are also important considerations – and again may be particularly relevant challenges for the public sector to overcome.

The government has made progress in driving up data quality, in part through its commitment to increased transparency. Nevertheless, for government to fully realise the benefits of bringing diverse data sets together, individual data owners must be incentivised to manage the quality of data collection and data capture across the board – and not just for the indicators they (or their managers) deem important.

Data security must be adhered to and maintained in line with technological developments. This is partly about hardware – modern data centres, encryption, secure government platforms and the like. It is also, again, about people. Where individuals and teams are working with sensitive data the culture and incentives must be compatible with good data security. Laptops and memory sticks left on trains and in bars – let alone unencrypted CDs lost in the post – must become a thing of the past.

Interoperability also matters for bringing diverse datasets together and implementing big data analytics in government. Open standards set a framework for data that can be joined across departmental boundaries (and shared with the public and business communities). Where standards are closed and analysis relies too heavily on proprietary data formats and software packages, there is a risk that some of the potential for imaginative use of data in the public sector will end

33 Campbell's law: the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor

34 Goodhart's law: once a social or economic indicator or other surrogate measure is made a target for the purpose of conducting social or economic policy, it may lose the information content that would qualify it to play that role

35 The Lucas critique: it can be naïve to try to predict the effects of a change in economic policy entirely on the basis of relationships observed in historical data (especially highly aggregated historical data)

up locked in to expensive solutions or be lost altogether. It will not be sufficient for public sector leaders simply to assume that a range of different data sets from different sources will be easy to connect together – in many cases they may need to force it.

All of these words of caution highlight the critical importance of the attitudes of those working with data and analytics in the public sector. The most effective individuals and teams will strike an even balance, combining their own judgment with data-driven analysis, listening carefully to different perspectives and being prepared to offer dissenting opinions. The Corporate Executive Board describes this group as “informed sceptics”, occupying a middle ground between visceral decision makers (who seldom trust analysis, and make decisions unilaterally) and unquestioning empiricists (who trust analysis over judgment and value consensus).³⁶

36 Good data won't guarantee good decisions, Harvard Business Review, April 2012

7

Capability

Capturing the big data opportunity will take more than recognition and aspiration. Beyond the vision required by public sector leaders to seek out big data opportunities, there is a real and significant challenge to meet around talent and capability.

The combination of skills and aptitudes required to excel in this field are wide and varied. In recent years the emergence of a new role within organisations – the data scientist – has begun to encapsulate the sort of capabilities that employers are looking for. The data scientist typically spans a number of disciplines; Gartner describes data scientists as having “three core data science skills: data management, analytics modelling and business analysis”. Beyond this they also point out the importance of soft skills such as communication, collaboration, leadership, creativity, discipline and passion – both for information and for finding the truth.³⁷

From this description it is abundantly clear that this is not a traditional role within most organisations. Moreover, combined excellence across these different domains is a rare occurrence. Taking the United States as a leading indicator, the number of data scientist roles across the economy looks set to rise rapidly (albeit from a low base – see Figure 4).

Numerous studies have estimated the potential jobs gap for data scientists and related disciplines – and put the figure for the shortage of staff with deep analytical talent in the United States in the low hundreds of thousands by the early 2020s. Perhaps ten times as many “data savvy” managers will be required to make sense of big data in business and other organisations.³⁸

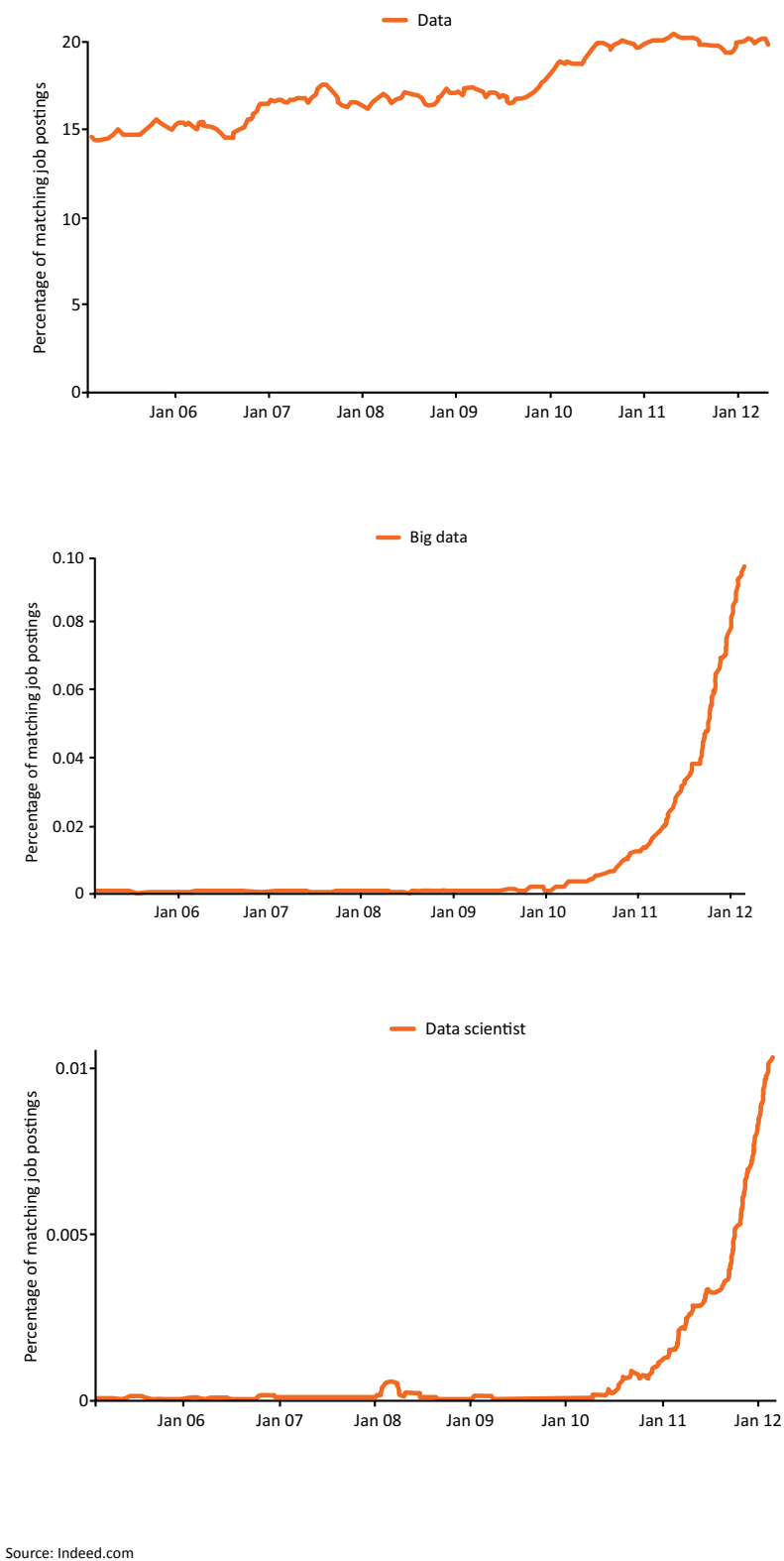
There is no formal data scientist career path in the UK public sector. Perhaps the closest parallel is in the Government Operational Research Service (GORS), which describes its remit as supporting decision making by understanding and structuring complex situations. Core competencies for its staff include management science, statistics, numerical/computing skills, project management, problem-solving and teamworking. GORS is a professional grouping; its analysts are embedded as civil servants in individual government departments. It recruits through the Civil Service Analytical Fast Stream, which also handles recruitment for the Government Economic Service (GES), Government Statistical Service (GSS) and Government Social Research Service (GSR).

So the challenge for the UK public sector is twofold: to further develop core professional expertise in data science, and to build out a more general capability to work with and consume data products across the mainstream civil service.

³⁷ Defining and differentiating the role of the data scientist, Gartner, March 2012

³⁸ Big data: the next frontier for innovation, competition and productivity, McKinsey Global Institute, 2011

Figure 4: Job trends



In thinking about how to kick-start this sort of agenda, the experience with two recent innovations at the centre of government is instructive.

The first is the Government Digital Service (GDS). This arose after Directgov was transferred from the Department of Work and Pensions to the Cabinet Office in 2010, and its remit expanded to reflect the recommendations made in Martha Lane Fox's report on government digital services.³⁹ The GDS are responsible for projects including digital engagement, Directgov and the single government domain. Their aim is bold: "to be the unequivocal owner of high quality user experience between people and government, through being the architect and the engine room of government digital service provision."⁴⁰ To help achieve this aim the GDS has recruited externally as well as drawing in existing civil servants; championed open and transparent ways of working; and moved into new offices at arm's length from Whitehall. A clear remit from Ministers to challenge departments and drive through change across government has helped significantly.

The second is the Behavioural Insights Team (BIT), also part of the Cabinet Office. BIT was created in the early months of the Coalition government to "find intelligent ways to encourage, support and enable people to make better choices for themselves."⁴¹ The team have again worked across government departments, with a clear mandate to challenge existing practices and pilot new approaches. Momentum is maintained by two innovative clauses in its terms of reference: to achieve a ten-fold return on the cost of the team, and a sunset review for the team in July 2012 (two years after it was established).

Elements from both of these models – including strong political backing to challenge the status quo, freedom to explore innovative new practices, and a rigorous specification of expected returns and timescales – are relevant for how government could build up similar momentum around data and analytics.

39 Digital by default proposed for government services, Cabinet Office, 2010

40 About the Government Digital Service, Cabinet Office

41 The Behavioural Insights Team, Cabinet Office

8

Power and Responsibility

Big data technologies – and predictive analytics in particular – can empower organisations with astonishing capabilities. Knowing a lot about an individual user, consumer or citizen makes it possible to forecast their specific needs and behaviour, sometimes with a very high degree of precision. Used responsibly, this capability to anticipate needs can be massively beneficial. Time can be saved, services personalised, and – in perhaps the most significant cases – decisions made or behaviours altered to avert undesirable outcomes.

We do all of these things already to some extent, from relying on linked data to speed up transactions to trying to eliminate or mitigate lifestyle factors that increase our risks of illness or disease.

We are increasingly encountering, however, situations where big data analytics collide with difficult issues around privacy and ethics.

To take just a few examples:

- Retailers can use data on purchases in conjunction with other publicly available information to infer intimate information about an individual's or household's circumstances. This can be extremely valuable for targeting promotions and building customer loyalty. Not all customers are happy, however, with the intensity of scrutiny that can be applied. In one famous example, the US retailer Target overstepped when it inferred that a teenager was pregnant (and mailed out coupons accordingly) before she had told her parents.⁴²
- Social networks, search engines and other online services can deploy sophisticated methods to track a user's behaviour online, infer what they will be interested in and predict changes in their personal circumstances. Fine-grained patterns in relationships can be discerned from scraping status updates from Facebook – and might be used to predict when individual relationships will form and end if combed with other information on a user's behaviour and connections in the social graph.⁴³ Google's recent changes to its privacy policy helpfully consolidate a number of different regimes together, but also increase the company's ability to target advertising across all of its services.⁴⁴
- Security and counter-terrorism agencies are working against a backdrop of massive growth in electronic communications and increasingly sophisticated adversaries. At the time of writing the government is consulting on proposals to extend the state's ability to monitor, intercept and store traffic and content data from electronic communications.⁴⁵ The arguments for an expansion of powers are in tension with rights to privacy and civil liberties. On the one hand, the more information the authorities have access to, the better their

⁴² How companies learn your secrets, New York Times, February 2012

⁴³ Amazing facts about Facebook and break-ups, Mikkelsen, 2010

⁴⁴ Privacy policy, Google, 2012

⁴⁵ Draft communications data bill, HM Government, 2012

ability to monitor and pre-empt threats. On the other, many citizens object to having their activities and communications closely monitored on a routine basis.

- Political parties are increasingly deploying big data analytics to enable micro-segmentation of voters. A detailed understanding of individual circumstances and priorities can enable campaign materials personalised on a household-by-household basis. This sort of activity is already commonplace in the United States. For political parties these sorts of tactics can make a crucial difference when margins of victory are small. Some commentators, however, argue that these developments are fuelling divisive and extreme partisan positions at the expense of reasoned public debate.⁴⁶

As government continues to explore and expand on its use of big data tools and technologies it will inevitably encounter more of these sorts of tensions. And from its position of authority it bears particular responsibility for executing analytics responsibly. How a government chooses to behave in this arena sets the standard for its peers and for other organisations that work with data in its jurisdiction.

Governments should have the utmost respect for civil liberties – and citizens themselves can and must hold their government to the highest ethical and moral standards.

46 The information arms, race, GOOD News, September 2011

9

Recommendations

Capitalising on the full potential of big data in the public sector is a major challenge and will not be achieved overnight. Nevertheless we believe that the prize at stake – better services for real people, and a leaner, smarter public sector – merits a renewed effort to make better use of the public sector’s data assets.

Our review of this topic has led us to two recommendations.

1. Advanced Analytics Team

A new Advanced Analytics Team should be established in the Cabinet Office, with responsibility for identifying big data opportunities and helping departments to realise them.

Building on the Government Digital Service and Behavioural Insights Team approaches, the team should be tightly focused, bring in external expertise, and have a mandate to seek out opportunities for improving policy regardless of how this might cross departmental boundaries.

Some suggested terms of reference are provided below.

Box 2: Draft Advanced Analytics Team terms of reference

The purpose of the team is to push the boundaries on how government uses its data assets to make efficiency savings and improve public services.

The team’s primary objectives are to:

- Work with departments to transform three major areas of public policy delivery by applying data and analytics in new, imaginative and/or more sophisticated ways
- Spread awareness, understanding and demand for cutting-edge data and analytical tools and techniques amongst senior leaders in the public sector
- Achieve savings and benefits for central government, over and above existing plans, worth at least £1 billion

To foster a sense of urgency the team will publish a progress review within one year, clearly identifying which savings it has identified and how they will be captured. After a further year to execute on these savings the team will be subject to sunset review.

The team budget might fall in the low millions of pounds (with enough flexibility to trade headcount against salaries if this is necessary to attract high-calibre candidates). By way of comparison, the new Open Data Institute, which

is researching open data, has government funding commitments of £10 million over five years.⁴⁷ The GDS budget for 2011–12 is set at £22 million.⁴⁸

If the initiative is successful then the government should consider formalising a data scientist career path in the public sector. This might be incorporated into the existing operational research service, or take the form of a new professional grouping for those working on a more regular basis with big data and analytics.

In any event the relevant professional group should also take responsibility for raising the level of competence on working with data across the public sector. In the majority of cases it will be non-specialists creating datasets and consuming the outputs of data analysis and visualisation. The best data scientists in the world will be no use if senior officials and Ministers have not had an opportunity to learn how to consume and use big data analysis effectively.

2. Code for Responsible Analytics

Government should adopt a Code for Responsible Analytics, to help it adhere to the highest ethical standards in its use of data and analytics.

Given the unique position of the state with respect to data, powers and citizens, this code should be the gold standard for ethical data use by big data owners. Other organisations would be free to follow the government's lead in adopting the code.

The components of such a code need further detailed consideration, and scrutiny from all parts of society – Parliament, government departments, citizens and civil society, businesses and beyond.

A first draft of a code is provided below – we urge government and all other interested parties to debate, challenge and improve on our proposal.

Box 3: Draft Code for Responsible Analytics

In the modern business of government, ethical data use is a matter of fundamental principle. This government will:

- **Put outcomes before capabilities.** Data and analytics capabilities will always be acquired on the basis of a clear and openly communicated public policy justification. Such capabilities will never be accumulated for their own sake, and when redundant will be surrendered promptly. Curiosity alone will never be a good enough reason.
- **Respect the spirit of the right to privacy.** Auxiliary data and analytics will not be used to infer personal or intimate information about citizens (including, but not limited to, reproductive status, sexual orientation, and political activity). Where this data is needed for public policy reasons, consent will be sought explicitly.
- **Fail in the lab, not in real life.** A sandbox environment and synthetic data will be used to test all major big data initiatives – after which they will be subject to intense scrutiny and peer review. Initiatives that are deemed to overstep the mark on ethics or privacy will be dropped. Only those that stand up to close scrutiny by Ministers will be implemented.

Across all of this agenda, we believe government should only execute on data and analytics where it is prepared to make an open and transparent case about the public policy benefits.

⁴⁷ Plans to establish Open Data Institute published, Cabinet Office, May 2012

⁴⁸ Electronic government: finance, Hansard, September 2011

The last element of our proposed code is particularly important given the difficulty we all have in drawing a clear line about how far it is acceptable for the government to go with big data and analytics. For the most part we are best at recognising policies that overstep the mark when we see them. The risk of a public backlash setting the entire data agenda back significantly puts a high premium on getting initiatives right. Proposals for sandboxing and lab work on synthetic data fits the spirit of “policy labs” announced in the recent Civil Service reform paper.⁴⁹ Scrutiny could be tightened further by providing an oversight role for Parliament in addition to that for Ministers.

Just because government can do something with big data, that doesn’t mean that it should do it. In the final analysis, if a Minister would not be comfortable putting themselves or their family under the sort of scrutiny required by a departmental big data initiative, then that initiative should not make it into government policy.

49 The Civil Service reform plan, Cabinet Office, June 2012

Postscript:

What Leaders Need to Ask

We recognise that the public sector is at the start of the big data journey – and that capturing the sorts of opportunities described in this report will take time and effort.

To help leaders in the public sector start the conversation within their organisations, we have assembled five key questions we believe they need to ask.

1. Do we know what we know?

This is the foundation for working with big data in any organisation. A comprehensive audit of big data assets – combined with a careful analysis of why the data is being collected and stored – will create a baseline for understanding what opportunities the organisation can seek out. Much of the data created by an organisation may be generated in the course of business-as-usual, but not routinely captured. An exercise in partitioning data and information into known knowns, known unknowns etc. may be helpful to expose gaps in knowledge and identify areas for improvement.

2. Are we using our data to drive decisions?

The importance of evidence-based policymaking is one of the most oft-repeated mantras in Whitehall. The quantity of data generated by modern organisations means it is theoretically possible to base many (if not all) decisions on at least some quantitative data. Of course in practice data may not be captured in a way that facilitates analysis, or be analysed in a way that helps decision makers. It is important to be able to distinguish which decisions are based on data and repeatable analysis, and those where judgement, estimates and rule of thumb take priority.

3. How quickly can we get answers?

Events often demand a rapid response – and this is just as true in the public sector as it is for commercial businesses. An organisation that is in control of its big data assets and has a well-developed analytical function should be able to answer questions and test hypotheses rapidly. This means knowing what is going on throughout the organisation in close to real time. If you have to wait days, weeks or months for the organisation to deliver answers then an audit of business processes and bottlenecks is in order.

4. What's our strategy for big data innovation?

As an organisation gets more comfortable with working with big data there will be an increasing desire and demand to realise the opportunities outlined at the start of this paper. Leaders will need to balance the need to create space for data

scientists and others to innovate and be creative, with the need to ensure talent is focused on the most important strategic questions. Organisations in the public sector will also need to decide how they are going to work with each other, and with experts in the private sector.

5. Who is responsible... and is this the right person?

Responsibility for generating value from big data and analytics sits in different places in different organisations. Typical domains holding responsibility include finance, IT, strategy, HR, operations and the C-suite. The important factor is that the locus of responsibility is coherent with what the top team are trying to achieve with big data. In many cases responsibility for big data may be more an artefact of the history of the organisation than a deliberate, strategic decision. Getting this right will have important benefits as an organisation's capability to use and extract value from big data matures.



The modern world generates a staggering quantity of data – and the business of government is no exception. The term *big data* has come to refer to the very large datasets involved, and big data analytics to refer to the process of seeking insights by combining and examining them.

This report is about a strategy for big data in the public sector. We provide an overview and examples to inspire policymakers around the opportunity for data and analytics to transform public service delivery. We also sound a note of caution about the challenges this agenda poses for the public sector, particularly around talent, capabilities and civil liberties.

Our recommendations show how government might begin to capture the opportunities of big data whilst meeting the challenges the public sector will face along the way.

£10.00
ISBN: 978-1-907689-22-2

Policy Exchange
Clutha House
10 Storey's Gate
London SW1P 3AY

www.policyexchange.org.uk