House of Commons

Science and Technology Committee

# The big data dilemma

## Fourth Report of Session 2015–16

HC 468

House of Commons

Science and Technology Committee

# The big data dilemma

## Fourth Report of Session 2015–16

*Report, together with formal minutes relating to the report*

*Ordered by the House of Commons to be printed 10 February 2016*

**HC 468**
Published on 12 February 2016
by authority of the House of Commons
London: The Stationery Office Limited
£0.00

## Science and Technology Committee

The Science and Technology Committee is appointed by the House of Commons to examine the expenditure, administration and policy of the Government Office for Science and associated public bodies.

### Current membership

Nicola Blackwood MP *(Conservative, Oxford West and Abingdon)* (Chair)

Victoria Borwick MP *(Conservative, Kensington)*

Stella Creasy MP *(Labour (Co-op), Walthamstow)*

Jim Dowd MP *(Labour, Lewisham West and Penge)*

Chris Green MP *(Conservative, Bolton West)*

Dr Tania Mathias MP *(Conservative, Twickenham)*

Carol Monaghan MP *(Scottish National Party, Glasgow North West)*

Graham Stringer MP *(Labour, Blackley and Broughton)*

Derek Thomas MP *(Conservative, St Ives)*

Valerie Vaz MP *(Labour, Walsall South)*

Matt Warman MP *(Conservative, Boston and Skegness)*

The following were also members of the committee during the parliament:

Liz McInnes MP *(Labour, Heywood and Middleton)*

Daniel Zeichner MP *(Labour, Cambridge)*

### Powers

The Committee is one of the departmental select committees, the powers of which are set out in House of Commons Standing Orders, principally in SO No 152. These are available on the internet via www.parliament.uk.

### Publication

Committee reports are published on the Committee's website at www.parliament.uk/science and by The Stationery Office by Order of the House.

Evidence relating to this report is published on the relevant inquiry page of the Committee's website.

### Committee staff

The current staff of the Committee are: Simon Fiander (Clerk); Marsha David (Second Clerk); Dr Grahame Danby (Science Clerk); Dr Elizabeth Rough (Committee Specialist); Darren Hackett (Senior Committee Assistant); Julie Storey (Committee Assistant); and Nick Davies (Media Officer).

# Contents

# Summary

We are living in the data age. Since Sir Tim Berners-Lee proposed his "vague but exciting" plan for a 'distributed information system' at CERN — and in the process inadvertently launched the information revolution — the amount of data we share has exploded.

- 90% of the data currently in the world was created in the last two years.
- In 2014 there were 204 million emails every minute and Google reported 4 million search queries.
- The total amount of global data is predicted to grow 40% year on year for the next decade.

The Data Centre for the Large Hadron Collider at CERN (the world's largest and most powerful particle accelerator) processes about one petabyte of data every day — the equivalent of around 210,000 DVDs, and distributes this data across the world via a grid which gives over 8,000 physicists near real-time access to LHC data. In the future, the Square Kilometre Array (the world's largest radio telescope, run from the UK's Jodrell Bank Observatory) will require supercomputers faster than any in existence in 2015, and network technology that will generate more data traffic than the entire Internet. The computer power it will need will be about three times more powerful than the most powerful supercomputer available in 2013, equivalent to the processing power of about 100 million 2013-era PCs.

Properly exploited, this data should be transformative, increasing efficiency, unlocking new avenues in life-saving research and creating as yet unimagined opportunities for innovation. But even existing datasets are nowhere near fully exploited. Despite data-driven companies being 10% more productive than those that do not operationalise their data, most companies estimate they are analysing just 12% of their data.

The stakes for the UK economy are massive. Big data is already a UK success story but it has huge unrealised potential, both as a driver of productivity and as a way of offering better products and services to citizens. An analysis in 2012 calculated that big data could create 58,000 new jobs over five years, and contribute £216 billion to the UK economy, or 2.3% of GDP, over that period. In the public sector, big data can increase the operational efficiency and targeting of service delivery.

Big data depends crucially on developing the necessary skills, providing infrastructure and setting parameters for sharing data to ensure valid privacy and security concerns are addressed. It is essential that the Government's forthcoming Digital Strategy sets a clear course to address these matters not only so that UK plc can capitalise on our world-leading data capabilities but also so our public sector can develop the sustainable solutions promised by big data within a secure regulatory and practical framework.

No Digital Strategy will succeed, however, without immediate action to tackle the crisis of our digital skills shortage. The Government should urgently commit to further supporting the development of 'data analytics' skills — a mix of technical skills, analytical and industry knowledge, and the business sense and soft skills to turn data

into value for employers — in businesses as well as in Government departments, and promoting more extensively the application of big data at local government level. But the Government must also address the wider context of its policies on apprenticeships and immigration control, including widespread concerns that these could jeopardise the necessary big data skills-base that the UK will increasingly need.

On infrastructure, the Government facilitates industrial access to academic infrastructure for research, and small business access to advanced software and hardware. Together with the Digital Catapult and the Open Data Institute, there is help for making datasets 'open' for researchers and analysts, or available as 'shared data'. The Government has a key role to play in making its own data 'open' and 'shared'.

Its work in this area has put the UK in a world-leading position, but there is still more to do, particularly in breaking down departmental data silos and improving data quality. The Government should examine how it can build capacity to deliver more datasets, increasingly in real-time, both to decision-makers in Government and to external users. It should map out how the Digital Catapult's work and the Government's plans to open and share its own data could be dovetailed. The Government should also consider the scope for giving the Office for National Statistics greater access both to Government departments' data and private sector data. It should charge the Government Digital Service, the Office for National Statistics or another expert body with auditing the quality of data within Government departments amenable for big data applications, and for proactively identifying data sharing opportunities to break departmental data silos. Healthcare interventions can be more precisely tailored to individual patients' circumstances using big data. The momentum for this was reduced, however, by the experience of bringing patient data together under the 'care.data' initiative. After the programme was delayed, the Spending Review has now raised the prospect of progress on this front, but the Government cannot afford a second failure from a re-launched scheme. It should take careful account of the lessons from a similar, successful, scheme in Scotland. In particular, to help bring patients onside and to streamline healthcare across different NHS providers — hospitals, GPs, pharmacists and paramedics — it should give them easy, online access to their own health records.

There are risks, as well as opportunities, from big data. Personal data is only a small proportion of big data, with huge potential from non-personal datasets for transport and weather forecasting, for example. Given the scale and pace of data gathering and sharing, however, distrust and concerns about privacy and security is often well founded and must be resolved by industry and Government if the full value of big data is to be realised. The benefits therefore have to be weighed against such potential loss of privacy and the risks of our data being lost or misused. Controls are covered by the Data Protection Act 1998, but will need to be overhauled within the next two years or so as a result of the agreement of an EU General Data Protection Regulation in December 2015.

The new Regulation will increase potential fines, but the Government should immediately go further by introducing a criminal penalty — already provided for in existing UK legislation — for serious data protection breaches. The Government and Information Commissioner should also ensure that the UK's already developed

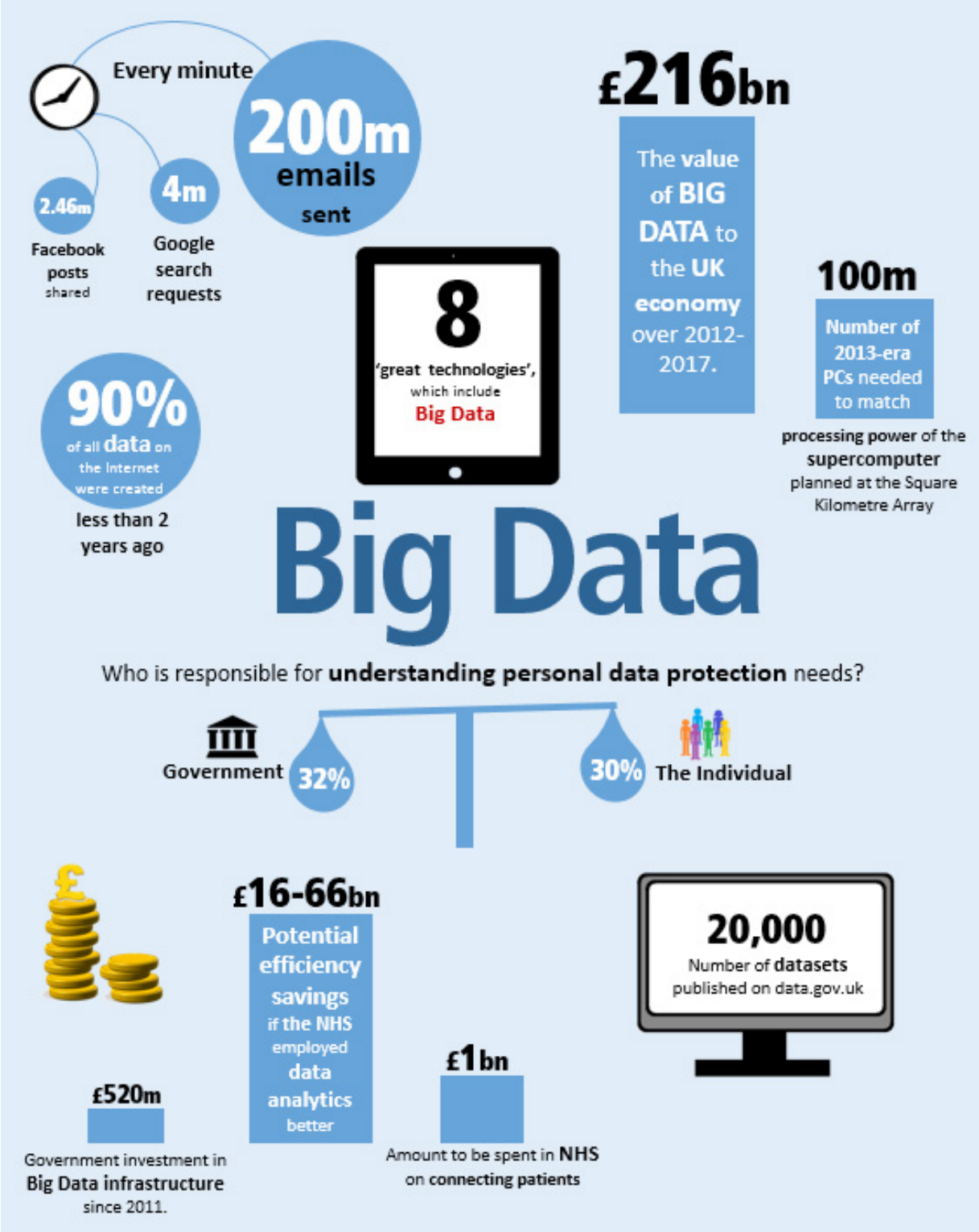
kitemark, to acknowledge and encourage good practice, is adopted as soon as possible along with a campaign to raise public awareness of it.

We do not share the Government's view that current UK data protections can simply be left until the Data Protection Act will have to be revised to take account of the new EU Regulation. Some areas need to be addressed straightaway — introducing the Information Commissioner's kitemark and introducing criminal penalties. And there remain concerns that big data techniques which 're-identify' individuals from previously anonymised data may be outside the scope of the current UK legislation. The way the new EU Regulation is framed appears to leave it open for data to be potentially de-anonymised if "legitimate interests" or "public interest" considerations are invoked. It is particularly important therefore that the Government set out its anonymisation strategy for big data in its upcoming Digital Strategy, including a clear funding commitment, a plan to engage industry with the work of the UK Anonymisation Network and core anonymisation priorities.

The anonymisation and re-use of data is becoming an issue that urgently needs to be addressed as big data becomes increasingly a part of our lives. There are arguments on both sides of this issue: Seeking to balance the potential benefits of processing data (some collected many years before and no longer with a clear consent trail) and people's justified privacy concerns will not be straightforward. It is unsatisfactory, however, for the matter to be left unaddressed by Government and without a clear public-policy position set out. The Government should clarify its interpretation of the EU Regulation on the re-use and de-anonymisation of personal data, and after consultation introduce changes to the 1998 Act as soon as possible to strike a transparent and appropriate balance between those benefits and privacy concerns.

Such clarity is needed to give big data users the confidence they need to drive forward an increasingly big data economy, and individuals that their personal data will be respected. The Government should establish a Council of Data Ethics as a means of addressing the growing legal and ethical challenges associated with balancing privacy, anonymisation, security and public benefit. Ensuring that such a Council is established, with appropriate terms of reference, offers the clarity, stability and direction which has so far been lacking from the European debate on data issues.

# Big data in numbers



# 1 Introduction

1. The term 'big data' is used to describe the collection and analysis of data on a scale or of a complexity that makes the use of such data challenging. According to the Information Commissioner, big data is a way of analysing data that "uses massive datasets", "often involves bringing together data from different sources" and may involve processing data "in real time".[1] Research Councils UK gives the following description:

> What constitutes 'big data' varies between disciplines and sectors. It goes beyond the extremely large and complex datasets generated by, for example, the Large Hadron Collider, DNA sequencing, Earth observation, government records and transactions, commercial or online interactions, to include data from new technologies. Smaller scale data of high complexity and variability, for example from environmental monitoring and the Internet of Things, where sensors capture and process large amounts of fast-moving (and often personal) data, is such technology. Regardless of the source, big data is about gaining value and insights from extremely large, complex, fast moving or combined data, across a range of sectors in innovative and beneficial ways.[2]

2. Big data has huge potential value to the UK, both as a driver of productivity and as a way of offering better products and services to citizens. It can help businesses and entrepreneurs "to identify areas of opportunity for innovation in new products, processes and services; improve customer engagement; identify inefficiencies; improve productivity, identify market trends; and use the UK Government's Open Data (data.gov.uk) to innovate and to create new companies".[3] In the public sector, "intelligent data analytics can help public service organisations to increase the operational efficiency of public service delivery, reduce expenditure and costs whilst delivering increasingly personalised services to citizens".[4] The European Commission's *Digital Single Market Strategy for Europe* states that:

> The growth of data is exponential—90% of all data circulating on the Internet were created less than 2 years ago.
>
> Only 1.7% of EU enterprises make full use of advanced digital technologies, while 41% do not use them at all. Digitisation also offers unprecedented opportunities to other economic sectors, such as transport (e.g. intelligent transport systems) or energy (e.g. smart grids, metering).[5]

3. In 2013, the Government identified big data as one of its 'eight great technologies', which it envisaged would contribute to future UK growth. The opportunities offered by big data, and the challenges involved in realising them, were highlighted, also in 2013, in the Government's *Information Economy Strategy*: "Business sectors across the economy are being transformed by data, analytics, and modelling. Data is increasingly being produced at a rate that means that current techniques are insufficient to fully exploit it."[6] The 'eight great technologies' were given £600 million in the 2012 Autumn Statement, of which £189 million was assigned to big data technologies, particularly for bioinformatics and environmental monitoring.

1 Information Commissioner's Office, *Big data and data protection* (July 2014)
2 Research Councils UK (BIG0057)
3 Research Councils UK (BIG0057)
4 Tech UK (BIG0039)
5 European Commission, *A Digital Single Market Strategy for Europe*, COM(2015) 192 (May 2015)
6 HM Government, *Information Economy Strategy* (June 2013)

4. Much coverage of big data focuses more on the risks than the opportunities, particularly in connection with the storage and processing of personal data. Our predecessor Committee examined the privacy issues arising from big data in the context of social media in their 2014 *Responsible Use of Data* report.[7] It cautioned that: "The UK is already a leading player on the global stage in using social media data and we are keen for this status to be maintained, but only if that can be achieved while ensuring the personal privacy of UK citizens."[8]

## Our inquiry

5. Since our predecessor Committee's earlier inquiry, £450 million has been allocated in the 2015 Spending Review for the Government Digital Service. The Data Protection Act 1998, which transposed the 1995 EU Data Protection Directive, will need to be overhauled within the next two years as a result of the agreement of an EU General Data Protection Regulation in December 2015.[9] The Government is also developing a Digital Strategy, which it expects to publish soon.[10]

6. We undertook this inquiry to follow up our predecessor's report in the light of these developments, and to begin what we aim to be a programme of inquiries looking at the 'great technologies'. We sought written evidence on the following terms of reference:

(1) the opportunities for big data, and the risks

(2) whether the Government has set out an appropriate and up-to-date path for the continued evolution of big data and the technologies required to support it

(3) where gaps persist in the skills needed to take advantage of the opportunities, and be protected from the risks, and how these gaps can be filled

(4) how public understanding of the opportunities, implications and the skills required can be improved, and 'informed consent' secured

(5) any further support needed from Government to facilitate R&D on big data, including to secure the required capital investment in big data research facilities and for their ongoing operation.

7. We received over 80 written submissions. We held three evidence sessions covering the health, direct marketing and financial technology sectors; the Information Commissioner and others concerned with the ethics of consent for data use; organisations tasked with assisting industry and other researchers use big data; and ministers for the digital economy (Ed Vaizey MP) and internet safety and security (Baroness Shields) as well as departmental officials. We are grateful for the assistance in our inquiry provided by Heather Reeve-Black from the National Audit Office.

7 Science and Technology Committee, *Responsible Use of Data*, Fourth Report of Session 2014–15, HC 245

8 Ibid, paragraph 35

9 European Commission, Press release, *Agreement on Commission's EU data protection reform will boost Digital Single Market*, 15 December 2015

10 DCMS, *UK Digital Strategy — the next frontier in our digital revolution*, News story, 29 December 2015

# 2 The opportunities for big data

8. We are living in the data age. Since Sir Tim Berners-Lee proposed his "vague but exciting"[11] plan for a 'distributed information system' at CERN—and in the process inadvertently launched the information revolution—the amount of data we share has exploded (paragraph 2). Properly exploited, this data should be transformative, increasing efficiency, unlocking new avenues in life-saving research and creating as yet unimagined opportunities for innovation across all sectors and industries.

9. As TechUK put it, "big data is a UK success story—underpinning the digital transformation across sectors and industries including retail, media and fintech, and is a key driver in enabling digital entrepreneurialism".[12] As Nesta highlighted, the UK—as 'the connected kingdom'[13]—is well placed to capitalise on this opportunity.[14] In the year to June 2015, eight UK tech businesses reached a valuation of at least $1bn—so called 'unicorns'.[15] But, as TechUK noted, "we are only at the beginning of the evolution of big data technologies". Existing datasets are nowhere near fully exploited, with most companies surveyed estimating that they are analysing just 12% of their data.[16]

10. The Centre for Economic and Business Research estimated in 2012 that big data could create 58,000 new jobs over the period 2012–2017, and contribute £216 billion to the UK economy, or 2.3% of GDP, over that period (Figure 1).[17] It expected two-thirds of this contribution to come from business efficiency (£149 billion), with the rest accounted for by business creation (£42 billion) and business innovation (£24 billion). The manufacturing sector was expected to be the biggest contributor, with £45 billion over the five year period (see Example 1). The opportunities for central government, as we discuss below, were also seen as significant (paragraph 34).

**Figure 1: Economic value of big data, by sector, 2012–2017**

| | |
|---|---|
| Manufacturing | £45.3 bn |
| Retail | £32.5 bn |
| Professional services | £27.6 bn |
| Central Government | £20.4 bn |
| Healthcare | £14.4 bn |
| Telecoms | £13.7 bn |
| Transport & logistics | £12.4 bn |
| Retail banking | £6.4 bn |
| Energy & utilities | £5.4 bn |
| Investment banking | £5.3 bn |
| Insurance | £4.6 bn |
| Other activities | £27.9 bn |
| **UK economy** | **£216.0 bn** |

11 CERN website, accessed February 2016
12 Tech UK (BIG0039)
13 Boston Consulting Group, *The Connected Kingdom: How the Internet Is Transforming the UK* (2010)
14 Nesta (BIG0047)
15 The Telegraph, It's taken years but the UK is finally building a great technology industry, 15 June 2015
16 Forrester Research, *The Forrester Wave: Big data hadloop solutions*, News item, 27 February 2014
17 Centre for Economics and Business Research, *Data equity: Unlocking the value of big data* (April 2012)

### *Example 1: Modelling in automotive design*

High-powered computing is being used to run computer simulations that model components of a product before the manufacturing process begins. Engineers from Bentley Motors used one such system to create virtual models of vehicles. This enabled faster product development times, decreased the number of prototypes required, reduced costs and eliminated the need for late-stage modification.[i]

i House of Commons Parliamentary Office of Science and Technology, Big data in business, POST Note 469 (July 2014)

11. Nesta told us that "data–driven companies are over 10% more productive than 'dataphobes'—firms that don't exploit their data". They estimate that if all such dataphobes were to make good use of data in driving their business decisions, it would produce a 3% increase in UK productivity.[18]

12. Big data is a "key driver in enabling digital entrepreneurialism".[19] The Government has sought to promote this through initiatives such as the Digital Catapult (which we discuss at paragraph 53) and Tech City UK. Tech City UK was launched in 2010 to support the East London tech cluster known as Silicon Roundabout, though the organisation has since extended its support to Greater London and other UK cities. Tech City's budget from Innovate UK is £2.2 million for 2015–16, which covers "programmes, policy informing and championing work".[20] Baroness Shields, minister for internet safety and security but also a former chief executive of Tech City, told us that its budget "does not sound much, but the impact of that initiative, putting a spotlight on technology in this country, has led to enormous investment and innovation around this area".[21] Indeed, the number of London's digital technology sector companies grew by 92% between 2010 and 2013,[22] and its Tech companies now employ over 250,000 people—a 17% increase on five years ago compared to a 7.8% rise in overall employment.[23]

13. The Government has identified the financial services technology—'fintech'—as a potential growth technology sector. In a 2014 report commissioned by UK Trade & Investment, consultants EY estimated that the UK fintech sector generates £20 billion in annual revenue.[24] Imram Gulamhuseinwala from EY gave evidence to us on the factors which were making "the UK market one of the most attractive markets in Europe" for fintech.[25] Experian believed the role of big data in financial services was to:

> detect patterns of financial or insurance fraud, to combine trader performance data, market data, unstructured news, user data, and general ledger data to gain previously impossible insights. This enables the 'real time' decision-making power that makes a difference between winners and losers in the financial markets.[26]

18 Nesta (BIG0047)
19 Tech UK (BIG0039)
20 Tech City website, accessed January 2016
21 Q239
22 Tech City UK, *Powering the digital economy 2015,* pp8, 16, 48
23 Oxford Economics, as reported by The Telegraph, *It's taken years but the UK is finally building a great technology industry*, 15 June 2015
24 EY and UK Trade & Investments, *Landscaping UK Fintech* (2014)
25 Ibid.
26 Experian (BIG0022)

James Meekings, the co-founder of Funding Circle, stressed the "huge potential that big data holds for the fintech industry" (see Example 2). He told us that:

> The more data we can access about the small businesses that come to us, and the better we can analyse it, the greater the benefit we can have. We already carry out extremely thorough checks on every business that applies for a loan, using the same systems as the banks. However, greater access to businesses' data would allow us to speed up these checks—and increase the overall number of businesses that we can assess.[27]

### *Example 2: Fintech*

Funding Circle is an online peer-to-peer lending network, identified by TechCity as one of its 'Future Fifty'[ii]—the UK's top 50 growth-stage digital companies. The Government's British Business Bank has partnered with Funding Circle as part of its Investment Programme, and has invested a total of £60 million in smaller businesses since 2012 via the company's lending platform.[iii]

ii Funding Circle, Future fifty

iii BIS, New £40 million investment by British Business Bank to support £450 million of lending to smaller businesses, News item, 25 February 2014

14. The potential benefits of big data are also significant in healthcare and medical research—both in terms of efficient delivery of services and discovery of more effective, personalised treatment of patients (see Example 3). Professor John Williams of the Royal College of Physicians illustrated the possibilities of big data for stratified medicine, using the example of targeted treatments for irritable bowel disorders. He concluded that:

> If we had large datasets, where we could analyse the physical and genetic make-up of the patient and their disease, we would be able to predict from the large dataset which of those treatments the patient was most likely to respond to. We would avoid putting the patient through a series of very dangerous treatments and end up precisely with the one most likely to benefit them.[28]

### *Example 3: Cancer diagnosis routes*

The National Cancer Intelligence Network 'Routes to Diagnosis' study examines different routes to cancer diagnosis, including delays in diagnosis, and their impacts on survival. It links data from Hospital Episode Statistics, cancer waiting times and cancer screening to data from the National Cancer Data Repository. Personal identifiers are used to link these datasets at patient level and to look at the effects of factors such as socio-economic status, age, gender and ethnicity on Routes to Diagnosis and patient outcomes. Results have informed public awareness campaigns, such as Public Health England's 'Be Clear On Cancer' campaign, seeking to help patients to spot symptoms of cancer earlier.[iv]

iv House of Commons Parliamentary Office of Science and Technology, Big data and public health, POST Note 474 (July 2014)

27 Funding Circle (BIG0081)

28 Q3

15. Other scientific disciplines such as experimental physics also make extensive use of big data techniques. The Data Centre for the Large Hadron Collider at CERN (the world's largest and most powerful particle accelerator) "processes about one petabyte of data every day—the equivalent of around 210,000 DVDs", and distributes this data across the world via a grid which "gives a community of over 8,000 physicists near real-time access to LHC data".[29] In the future, the Square Kilometre Array (the world's largest radio telescope, run from the UK's Jodrell Bank Observatory) will "require supercomputers faster than any in existence in 2015, and network technology that will generate more data traffic than the entire Internet".[30] The computer power it will need will be about three times more powerful than the most powerful supercomputer available in 2013, equivalent to the processing power of about 100 million 2013-era PCs.[31]

16. The Meteorological Office uses supercomputers to model climate change and its impacts. In a more everyday application, it collects and analyses a massive amount of data every day to produce weather forecasts, as well as advising energy and retail sectors, for example, about weather that might affect "consumer trends".[32] Others have used Met Office data, along with other datasets, to provide additional big data commercial outputs (Example 4).

*Example 4: Using big data to plan for extreme weather events*

UK company, KnowNow Information Ltd, provides information for emergency services to plan for, and respond to, extreme weather conditions. Using the big data analysis capabilities provided by the Science and Technology Facilities Council's Hartree Centre, it can predict the probability of certain types of emergency, based on location and weather conditions. Its flood event model combines existing open data generated by the emergency services, the Met Office, Ordnance Survey, the British Geological Survey and the Environment Agency.[v]

v STFC, Big Data predicts extreme weather blackspots for UK emergency services, news item (5 September 2015)

17. The Centre for Economic and Business Research estimated that there could be £20 billion of benefit from big data over a five-year period for central government (Paragraph 10). The Government described how big data can "help cut costs, increase productivity, and improve the delivery of services". By analysing 800 million monthly credit and debit card payments, for example, and matching these with other datasets, HMRC has been able "to more effectively target tax enforcement activity".[33] Paul Maltby, Director of data at the Government Digital Service, described the opportunities for better use of data:

> There is huge opportunity … [to use data] to segment audiences and think about predictive analytics and tailor interventions. That is one of the very large potential gains in this field for public services.[34]

29 CERN website, accessed January 2016
30 Square Kilometre Array website, accessed January 2016
31 Square Kilometre Array website, accessed January 2016
32 Met Office website, accessed January 2016
33 BIS and DCMS (BIG0069)
34 Q200

18. Elsewhere in the public sector there are opportunities for efficiency savings in transport (see Example 5). Transport for London (TfL) told us how they were using big data to "transform transport services":

> Twelve million daily public transport trips make Oyster and contactless payment cards a significant source of big data. Nineteen million daily 'taps' from these systems allow travel patterns to be studied, bringing a depth to our understanding of customer profile and behaviour. The additional 18 million car, cycling, and walking trips provide a phenomenal 30 million daily journeys on the TfL network that are fit for big data analysis.
>
> London's richly detailed travel data feeds into transport planning models to predict the impact of development in our city. TfL has long used station entry and exit data for network planning. We can now also infer where people are leaving a bus through a big data tool … that combines bus location and ticketing data to create origin and destination pairs. This creates a comprehensive picture of travel patterns which network planning teams can use to minimise the impacts of closures or diversions.[35]

*Example 5: Modelling the rail network assets*

Network Rail's £330 million 'ORBIS' programme aims to create a detailed digital model of the UK's rail network in order to improve the organisation's asset management. This uses geographical data collected by maintenance staff using tablets and smartphones to generate a spatial model of the railway infrastructure, containing information about how assets are used and their capability and performance.[vi]

vi Royal Academy of Engineering and Institution of Engineering & Technology, Connecting data: Driving productivity and innovation (November 2015

19. **The UK is a world leader in big data research across disciplines and our Tech sector, especially in London, dramatically outperforms the rest of the economy on growth and productivity indicators. By identifying big data as one of the Eight Great Technologies, and investing significant financing in large scale data infrastructure, the Government has signalled that realising the full potential of big data is a priority. However, investing in capital infrastructure projects alone will not deliver this. Urgent action on the digital skills crisis, overcoming public distrust over data sharing, further progress on 'open data' and greater clarity over prospective data protection legislative changes are essential if the UK is to set the pace on big data. We discuss these pre-requisites in the following chapters.**

35 GLA and TfL (BIG0067)

# 3 Skills and Infrastructure

## Skills

20. TechUK, like other witnesses, were clear that "the UK has a fantastic opportunity to be a world-leader in the development, adoption and exploitation of advanced big data analytics technologies, and is making steady progress to date." This progress will stall, however, without urgent action to address our digital skills crisis. Tech UK found in a recent survey that 93% of technology companies experienced digital skills gaps which affected their operations. They stressed, as others also did, that "the digital skills gap is a major concern for industry, and if not overcome will impede the UK's ability to be a world-leader".[36] Our predecessor Committee had similarly concluded in 2014 that:

> Data science is yet another skills area that urgently needs to be addressed if the UK is to be able to build an economy that can compete on the global stage. It is essential that the Government ensures that data science skills are promoted in educational institutions and within organisations that are able to provide data skills development.[37]

21. Nesta saw the key to achieving the full potential of big data to be "finding people with the right mix of skills—the data scientists who combine technical skills, analytical and industry knowledge, and the business sense and soft skills to turn data into value for employers".[38] In 2015, Nesta and Universities UK published *Analytic Britain: Securing the right skills for the data-driven economy* on how to "adapt, re-purpose and prioritise existing initiatives and programmes" to "remedy skills shortages in the short term, while ensuring a sustainable supply of excellent analytical talent in the longer term".[39] Recommendations for the school and university sectors were concerned with:

- Improving teaching and up-take of mathematics and statistics;
- Embedding data analysis across subjects, and fostering interdisciplinary research;
- Promoting analytical careers and role models among school and college students;
- Boosting the soft skills of data analytics graduates.

22. The University of Cambridge's Big Data Strategic Research Initiative emphasised that "data science is fundamentally interdisciplinary …. Collaborative partnerships, for interdisciplinary research and training … as well as between academic, commercial and policy stakeholders, must be supported and promoted."[40] Tech UK saw a need for "a strategy to boost the domestic big data talent pipeline and address the immediate skills shortage through a smart migration policy environment", addressing visa rules for overseas workers and students. They also wanted the Government's apprenticeship scheme to be "geared toward the high-value jobs of the future, such as big data" and the new schools Computing Curriculum to be delivered with sufficient numbers and quality of teachers.[41]

36 Tech UK (BIG0039)
37 Science and Technology Committee, Responsible Use of Data, Fourth Report of Session 2014–15, HC 245
38 Nesta (BIG0047)
39 Nesta (BIG0047)
40 University of Cambridge Big Data Strategic Research Initiative (BIG0016)
41 Tech UK (BIG0039)

23. Dr Paul Feldman, chief executive of JISC, told us that a cross-cutting taskforce envisaged in the *Analytic Britain* report (paragraph 21) would soon be convened—including TechUK, the Digital Catapult, the Tech Partnership, the UK Commission for Employment and Skills and others—to "discuss some of the skills issues and where we can try to plug skills gaps".[42] The Government told us that it was "working closely with Nesta, Universities UK and the British Academy on following-up recent studies into business demand for data analytics skills and how these skills are currently taught across different disciplines in [higher education] institutions".[43] The Government listed for us initiatives it had launched over recent years to develop education and training in data analytics:

- The new national curriculum in computing introduced in September 2014 to reform GCSEs and A levels in Computer Science and provide a strong foundation for students to progress to further education or employment.
- Reforming apprenticeships in areas like software development so that the standards are developed by employers.
- Over £40 million investment since Autumn 2014 in nine centres for doctoral training in different aspects of data.
- £18.4 million of business and Government co-funding announced in July 2014 for the employer-led Tech Partnership. This funding will give employers a greater say in skills provision, including the delivery of over 2,700 industry approved apprenticeships.[44]

### *Skills in government*

24. In the 2015 Spending Review the Government announced £450 million for the Government Digital Service,[45] which would "continue to act as the digital, data and technology centre for government".[46] Paul Maltby, Director of Data at the GDS, highlighted elements of the Government's data programme. This included "improving the Government's data infrastructure" (paragraph 28), "continuing our work on open data" (paragraph 34), and "introducing data science more at scale across the government system … in a way that will be transformative in how we are able to use data for digital services" (paragraph 38) and developing common policy objectives across government.[47]

25. In terms of skills, Paul Maltby told us that the Government has not had difficulty recruiting data scientists at an early stage in their careers but that "affording people later in their career is somewhat difficult", and that retention could be a problem.[48] However, he pointed to a development package for existing analysts in which they had "learned to use some of the [data scientists'] tools and techniques".[49]

42 Q63
43 BIS and DCMS (BIG0069)
44 BIS and DCMS (BIG0069)
45 HM Treasury, *Spending Review and Autumn Statement 2015*, Cm 9162 (November 2015), para 1.287
46 Q187
47 Q187
48 Q196
49 Qq192, 196

26. As the Government Digital Service has noted, "the world of data does not stop neatly at administrative borders", and has emphasised that its work "will need to speak across to local authorities, the NHS and the devolved administrations".[50] When pressed on what this would entail, however, Paul Maltby told us:

> For the moment [our data accelerator programme] has been a central government thing. … Certainly central government would not want to dictate an approach and tell people what to do. There are great analysts all over local government, and people have a different way of doing things in each area. … As this new programme starts, the data steering group has representation from senior leaders in local government. It will be a question of how we see things developing. Things like data standards and data infrastructure do not easily stop at administrative borders, so we are keen to think about how we make the most of that.[51]

The ultimate aim for Government was to:

> spread [a data science] capability, knowledge and skillset very broadly indeed. A world where Government departments and public agencies had already fixed their data infrastructure in a way that made it inter-operable and fluid, where it should be fluid, so that they had fabulous data sites and capabilities integrated within the decision-making processes and services for citizens as a matter of course would be a fabulous thing, and it is something we are aiming towards.[52]

27. **The digital skills gap is approaching crisis levels and this not only has economic implications but also puts the quality and security of this data at risk. There is a range of Government initiatives to help develop computing and digital skills, but the wider set of 'big data' skills is not being strategically addressed. This risks UK business being unable to grow the big data sector at the pace it should. In the meantime, this skills gap is forecast to grow exponentially as big data reaches further into the economy. The evidence we received on the digital skills crisis was so concerning that we have launched a further inquiry specifically into this issue on which we will report shortly. *In the meantime, the Government should commit to (a) a continuing substantial role in developing data analytics skills in businesses, with others already working in this field; (b) increasing big data skills training for staff in Government departments; and (c) promoting more extensively the application of big data at local government level. But the Government must also address the wider context of its policies on apprenticeships and immigration control. As it develops its approach in these areas, it should explicitly address widespread concerns that these could jeopardise the necessary big data skills-base that the UK will increasingly need. The Government should also set out in detail how the Government Digital Services' budget, including the additional funding announced in the Spending Review, will be spent.***

50 Paul Maltby, The work of the government data programme, Gov.uk blog, 24 September 2015
51 Qq198-199
52 Q201

## Infrastructure

28. In addition to education and training, the Government has recognised the need for infrastructure development in the research sector. It told us that:

> Since 2011, the Government has invested over £520 million in developing the UK's big data and high-performance computing capital infrastructure. This programme includes major investments in specialist centres such as the Farr Institute of Health Informatics Research, the Higgs Centre for Innovation, and the Hartree Centre at the Daresbury Science & Innovation Campus. The recent Government investment of £113 million in the Hartree Centre will develop cognitive computing and data intensive technologies to lower the skills barriers to using and deriving benefit from data.[53]

The Alan Turing Institute, the UK's national institute for data science, was launched in November 2015 with a £42 million grant announced in Autumn Statement 2014.[54] The Institute aims to:

> enable knowledge and predictions to be extracted from large-scale and diverse digital data. It will bring together the best people, organisations and technologies in data science for the development of foundational theory, methodologies and algorithms. These will inform scientific and technological discoveries, create new business opportunities, accelerate solutions to global challenges, inform policy-making, and improve the environment, health and infrastructure of the world in an 'Age of Algorithms'.[55]

29. Dr Paul Feldmann from JISC, which provides digital technology and resources for higher and further education and researchers, told us how the JANET network gives researchers in universities "access to the high-performance computing they will need to process that data".[56] JISC noted however that a lack of access to such facilities can be a block for big data exploitation among businesses—"the barriers to [analysing] big data sets can be significant, particularly where expertise and access to high performance computing … facilities are concerned".[57] In response to this, JISC have established a process which allows small and medium-sized enterprises and universities to "purchase access" to infrastructure and expertise in "£60 million of publicly-funded high-performance computing facilities".[58]

30. The minister for the digital economy, Ed Vaizey MP, emphasised that:

> The Government's role is to invest in big projects that can help companies analyse big data or invest in skills, because we need data scientists who can help companies crack those big datasets, and also work with companies to tell them about the opportunities and give them a route map to engage in big data.[59]

53 BIS and DCMS (BIG0069)
54 HM Treasury, *Autumn Statement 2014,* Cm 8961 (December 2014)
55 The Alan Turing Institute website, accessed January 2016
56 Q39
57 JISC (BIG0027)
58 JISC (BIG0027)
59 Q237

The Government told us that infrastructure was part of that approach:

> The Government recognises the importance of creating a coordinated infrastructure for our data capability. This includes offering industrial access to academic infrastructure for specific needs such as pre-competitive collaborative research, and small business access to advanced software and hardware.[60]

The Minister described this in more detail:

> The digital catapult is particularly focused on small businesses. It encourages small businesses and is designed to be a space for them to try out new business models and have the resources available to do that, which only a big company might have. That is the first issue. There are big companies as well. … We are doing [a collaboration] with IBM in terms of their health technology. They can work with the big data projects that the Government are funding. There are hundreds of millions of pounds going into these projects, and even big companies could not necessarily match that kind of research, so it is important that we collaborate with business.[61]

31. ***While investment for big data research is welcome, we believe that the Government should explore further ways of making publicly-funded infrastructure and expertise available to more businesses. The Digital Catapult is a good start but it is essential that ongoing resource investment in the Catapult is maintained so that it can consolidate and expand its work. As big data becomes an increasingly significant part of our economy, the Government should set out its strategy for longer term big data infrastructure development and how it will work with industry to provide a coherent programme of business support.***

60 BIS and DCMS (BIG0069)
61 Q244

# 4 Open data and data sharing

32. One of the distinguishing features of big data is that it often brings together data from multiple sources. Big data can make use of datasets which are 'open'—licensed for anyone to access and use. Many real-time datasets, such as live travel and weather feeds, fall into this category. Alternatively, access to external datasets can be facilitated by data-sharing agreements. 'Shared data' is typically stored in a secure setting, with restrictions on with whom the data is shared and for what purpose, as, for example, in the sharing of administrative data between government departments (paragraph 39). Gavin Starks, chief executive of the Open Data Institute, explained the role of his organisation in promoting each of these modes of data sharing:

> When we talk about an 'open licence' we are referring to data that anyone can access, use and share. … The more complex area, where we have significant questions to ask, is the 'shared data' category. … We have very strong views that core data infrastructure should be open and owned by the state … In terms of what we see as the open remit, we should help to stimulate open innovation. The kind of work we can do to get the roles, policies and liabilities sorted out around the shared part of the data spectrum will help to unlock a huge amount of innovation and value in the country.[62]

33. Gavin Starks thought that "the processes, policies, standards and so on" of open and shared data were "much harder" than the data analysis itself.[63] Government has a key role to play, nevertheless, in making its own data 'open' and 'shared', to enable the value in those data to be realised, whether for research, service delivery or commercial purposes. It can do this by making its data available for outside bodies to use, or make full use of its data itself to improve the cost-effectiveness of the public services it provides.

## Government data

### *Open data*

34. By publishing its data the Government can stimulate business and innovation, provide transparency and accountability, empower citizens to make informed choices about products and services, and improve data quality through its wider and more frequent use. The 2013 Shakespeare Review estimated the 'direct value' of public sector information at £1.8 billion, with 'wider social and economic benefits' worth up to £6.8 billion.[64] Transport for London (TfL) described how it has more than 5,000 developers registered to receive its data sets and how this has stimulated the creation of 360 transport information apps for mobile devices.[65] The Shakespeare Review estimated the value of time saved as a result of access to real-time travel data from TfL at £15-58 million a year, at a cost to TfL of £1 million a year.[66]

62 Q202
63 Q209
64 Stephan Shakespeare, An Independent Review of Public Sector Information (May 2013)
65 GLA and TfL (BIG0067)
66 BIS, Deloitte, *Market Assessment of Public Sector Information* (May 2013)

35. Gavin Starks from the Open Data Institute saw data as public infrastructure:

> We should really be thinking about data as infrastructure in the same way as we think about roads as infrastructure. Roads help us navigate to places; data help us navigate to decisions. Those decisions need to be made by everyone. There is a lot of work to be done to work out what we would classify as data infrastructure for the country—for example, our geo-spatial information. [The Department for Environment, Food & Rural Affairs] has just released its dataset called LiDAR, which is very detailed environmental mapping. That has helped local businesses, citizens and Government make better decisions about their built environment.[67]

36. The Government has introduced a number of open data initiatives. In 2010 it launched the 'data.gov.uk' data portal, which includes data from central government, local government, agencies and arms-length bodies, NHS bodies and the police—typically licensed under the Open Government Licence.[68] The datasets are prioritised for publication according to criteria in the National Information Infrastructure,[69] set up by the Government in 2013 in response to the Shakespeare Review.[70] In 2012, the Government established the Open Data Institute, with £10 million of funding over five years from Innovate UK. Cabinet Office minister Matt Hancock MP recently noted the extent of international recognition for Government open data:

> So far we've published over 20,000 datasets [on data.gov.uk], covering almost £200 billion of public spending. This approach has won us plenty of plaudits. For the second year running, we've topped the World Wide Web Foundation's Open Data Barometer. Last year we were number one in the Global Open Data Index.[71]

37. However, the Greater London Authority argued that current progress on open data does not go far enough:

> The internationally acclaimed London Datastore contains over 600 datasets, but few among these can be described as big data. Transport data feeds aside, it is a very high quality but largely static data catalogue.
>
> The public sector's strategy of opening up proprietary data in machine readable form so that third parties can develop products or analysis to benefit stakeholders and the wider digital economy, has been an undoubted success. … However, without further incentives to encourage more consistent, higher quality and higher volume 'big' data from a wider set of suppliers, cities like London will fail to capitalise on the potential of big data to tackle the complex questions with which cities have historically grappled and deliver potentially transformative innovation.[72]

67 Q215
68 data.gov.uk website, accessed January 2016
69 Cabinet Office, The National Information Infrastructure (March 2015)
70 HM Government, *Government Response to Shakespeare Review of Public Sector Information* (June 2013)
71 Open Data Institute summit 2015, Matt Hancock MP speech, 3 November 2015
72 GLA and TfL (BIG0067)

Experian considered that "whilst part of Government have been embracing [open data], there is a lack of a clear joined up directional policy around open data and the technology to facilitate and deliver value from this".[73]

### *Government administrative data*

38. Paul Maltby from the Government Digital Service saw big data being used increasingly in providing Government services:

> The world has changed and is changing utterly the way we enjoy services in our everyday lives that are powered from data analytics and the way data work. We want to bring that transformation to government.[74]

Cabinet Office minister Matt Hancock MP recently stated that:

> Openness is a means to an end. The end is to make government work better for the people of this country. That means better decision-making within government: policy based on data and evidence, not dogma and theory.[75]

39. The Government has initiatives in place to provide researchers with access to its administrative data. The Administrative Data Research Network, established by the Economic and Social Research Council as part of its Big Data Network,[76] is a "UK-wide partnership between universities, Government departments and agencies, national statistics authorities, the third-sector, funders and researchers". It securely provides administrative data to researchers wishing to carry out social and economic research which "has the potential to benefit society".[77] Similarly the HMRC 'Datalab' allows researchers from academic institutions and other Government departments to access anonymised data from HMRC.[78] Both of these programmes require researchers to apply for access to the data, and projects are approved on a case-by-case basis, rather than facilitating real-time access to Government data.

40. Where such schemes are not available, we heard an example of Government taking the initiative and sharing the benefit of its administrative data with external organisations:

> [The Ministry of Justice] hold great datasets in government, but we have to hold them very securely because they include very sensitive data. We are trying to explore ways of making that data available to academe in a way that is safe and in accordance with the law, and also bears in mind the important ethical and privacy issues academics take very seriously. … [The Ministry of Justice] developed a very interesting and novel way of helping charities work out who is and is not reoffending, by allowing charities to send their data to the Ministry of Justice. The Ministry of Justice did the matching and analysis and sent back the results. That was hugely successful.[79]

73 Experian (BIG0022)
74 Q192
75 Open Data Institute summit 2015, Matt Hancock MP speech, 3 November 2015
76 ESRC website, accessed January 2016
77 Administrative Data Research Network website, accessed January 2016
78 HMRC (gov.uk) website, accessed January 2016
79 Qq264-265

Digital economy minister Ed Vaizey MP believed, nevertheless, that at the moment data sharing between departments does not go far enough, and that facilitating legislation may be needed:

> We have set up the Government Data Taskforce with the chief scientist and others to try to get Government departments to take big data seriously, to see the opportunity and also to share it. Mindful of the ethical concerns … surrounding things like 'care.data' [paragraph 45], this provides massive opportunities. I think we need to look at potential future legislation to allow that sharing to be made easier between Government departments.[80]

41. In a similar vein, Hetan Shah from the Royal Statistical Society worried about data remaining in departmental silos and saw potential benefits if the Office for National Statistics were able to collect Government administrative data from across departments:

> It does not seem to me that variability of data quality is the key issue in terms of stopping the sharing of data within Government and making it open. Francis Maude always used to make the argument that if you open up datasets the quality will increase … One of the big problems is that there is a silo mentality within Government, and different datasets are held and not shared across departments.
>
> The single biggest opportunity is to move where other countries have gone—Canada, New Zealand and Ireland—in giving the statistical office a broad right to data access across departments. At the moment, the Office for National Statistics cannot easily get hold of HMRC, BIS and DWP data. If it could, we would have more real-time access to what is going on around the country. … You would not have the privacy issues, because the ONS is interested only in aggregate data; they do not care about us as individuals.[81]

42. **There are enormous benefits in prospect for the economy and for people's lives from making the nation's core data infrastructure 'open'. The Government's work in this area has put the UK in a world-leading position. But there is more to do to breakdown departmental data silos, to bring data together in order to further improve public services, as well as to improve data quality (as we discuss in the healthcare context below). *The Government should set out how it can build capacity to deliver more datasets, increasingly in real-time, both to decision-makers in Government and to external users and, in particular, should work to establish a right of access to data for the Office for National Statistics. The Government should also establish a framework—to be overseen by the Government Digital Service, the Office for National Statistics or another expert body—for auditing the quality of data within Government departments amenable for big data applications, and for proactively identifying data sharing opportunities to break departmental data silos.***

80 Q264
81 Q203

### Healthcare data

43. An area where the potential benefits of big data has been particularly significant, but also where data quality constraints have been evident, has been in healthcare and medical research. In 2014, Volterra and EMC consultants found that the NHS was "considerably behind other industries in terms of its use of data analytics", and identified potential efficiency savings of between £16 billion and £66 billion a year if the NHS employed data analytics more successfully.[82] The potential benefits could be better quality healthcare, with interventions more precisely tailored to individual patients' circumstances (as illustrated at paragraph 14 above) if their medical and other data can be matched to extensive datasets. Aisling Burnand from the Association of Medical Research Charities highlighted how this would help research on rare diseases:

> Up and down the country there may be only a handful of people with a particular condition. So the ability to join up public patient datasets to find those people and then use the data for research purposes will, we hope, lead to improvements in treatments and, hopefully, cures and life-saving advances.[83]

44. Such big data benefits depend however on the quality of the datasets being brought together. Professor John Williams of the Royal College of Physicians was concerned about the quality of hospital data because the data collection process is "out of date and no longer appropriate for [big data analysis]. … There is no requirement for a feedback-loop for clinicians to validate the data centrally so that we get richer and more accurate data."[84] He worried that current analysis of, for example, the mortality of patients admitted at the weekend was based on some available datasets but it still lacked quality data on other key statistics, so that "premature conclusions … are being drawn from the data because it is not rich enough".[85] Similarly, Professor Montgomery, chair of the Nuffield Council on Bioethics, told us:

> There are major problems of data quality. The further away the interpretation of the health data gets from the person who produced it, the more scope there is for it being misinterpreted. There is work going on to try to improve standardisation and the way we record things, which would make it more possible to translate those things. … Extracting them as if they can be understood without reference to context is problematic in health data, because people record things in so many different ways.[86]

45. As our predecessor Committee reported in 2014, the momentum for using big data to improve health services was reduced by the experience of bringing patient data together under the 'care.data' initiative. They stated that:

> Members of the public do not appear to be wholly against the idea of their data being used by Government institutions, but support for data usage is highly dependent upon the context within which the data is collected.

82 Volterra Partners for EMC, *Sustaining universal healthcare in the UK: Making better use of information* (September 2014); and EMC (BIG0046)
83 Q2
84 Q9
85 Q10
86 Q127

> The Government should have learned from the experience with care.data and we recommend that the Government develop a privacy impact assessment that should be applied to all policies that collect, retain or process personal data.[87]

46. The 'care.data' programme was introduced in 2013 by NHS England and the Health and Social Care Information Centre (HSCIC)—a system to "extract and link large amounts of patient data, collected as part of NHS care, in order to improve the delivery of healthcare and to benefit researchers inside and outside the NHS". However, before the system could be launched, the programme was delayed to "allow GPs more time to notify their patients and for NHS England to conduct a public awareness campaign".[88] Dame Fiona Caldicott told us that the system was "put on hold because there was loud and extensive protest, not least from the general practitioners who were being called upon to download patients' data from their health records about the patient to the HSCIC, in terms that GPs were not content about".[89]

47. To regain patient trust, in 2014 the Department of Health established a National Information Board to "put data and technology safely to work for patients, service users, citizens and the professionals who serve them".[90] In the same year, the Secretary of State for Health appointed Dame Fiona Caldicott to a new role as National Data Guardian for Health and Care—"the patients' champion on security of personal medical information".[91] Following a pathfinder stage, the care.data programme had been expected to be re-launched in September 2015, but Dame Fiona told us that:

> The Secretary of State for Health decided that new work should be done on the question of patients being able to opt-out of how their data were taken from one place to another and used, so there is currently another pause. Were [care.data] to be restarted, I think it would be on the lines of much improved communication with both GPs and patients. … One thing that might be worth considering for the future is whether we should look at a more general question about data flows for a list of purposes, rather than the rather narrow purpose as publicised.[92]

48. Dame Fiona first addressed the flow of patient data "from NHS organisations to other NHS and non-NHS organisations" in the 1997 *Report on the review of patient-identifiable information*.[93] The 'Caldicott Report' introduced the 'Caldicott principles'—key principles of good practice for using patient data. In a 2012 review of these principles, they were extended to include the principle: "The duty to share information can be as important as the duty to protect patient confidentiality".[94] Reflecting on her review, Dame Fiona told us that she was "very disappointed when we revisited the new 'Caldicott principles' … to find that the culture in the NHS of sharing information had not moved in the way we hoped".[95] Today, the benefits of sharing patient data have yet to be realised. Dame Fiona noted, for example, that "there is a real issue for the public about why the ambulance service cannot see key aspects of the [patient] record when they go to collect an unconscious patient."[96]

87 Science and Technology Committee, Responsible Use of Data, Fourth Report of Session 2014–15, HC 245, para 29
88 House of Commons Library, Care.data, Standard Note SN06781 (October 2014)
89 Q11
90 National Information Board (gov.uk) website, accessed January 2016
91 Gov.uk, National Data Guardian appointed to safeguard patients' healthcare information, News item, 13 November 2014
92 Qq12-13
93 Department of Health, *Report on the review of patient-identifiable information* (December 1997)
94 Williams Lea for the Department of Health, Information: To share or not to share?, (March 2013)
95 Q25
96 Q23

49. The November 2015 Spending Review has, however, now raised the prospect of progress on this front:

> The Government will invest £1 billion in new technology over the next 5 years to deliver better connected services for patients and ensure that doctors and nurses have the information they need at their fingertips. By September 2018, 80% of clinicians in primary, urgent and emergency care will have digital access to key patient information. By 2020 integrated care records will give every health and care professional concerned with an individual's care the information they need to provide safe and prompt care.[97]

50. The success of a scheme similar to care.data in Scotland demonstrates that patients and healthcare professionals are not against the sharing of patient records if that sharing is performed with due care, and the benefits are clearly articulated. According to Professor John Williams of the Royal College of Physicians, the programme in Scotland "put together the infrastructure and the process with patients … It is because of that engagement that they have done better".[98]

51. Another part of securing individuals' consent for the sharing and use of their data is allowing them to see their data record and amend it. Dame Fiona told us that "within the next year or two, access to their records will be available to patients".[99] Aisling Burnand from the Association of Medical Research Charities believed that:

> This might even help to drive up the quality if they are able to see what is in the record. They might be able to add to the record, at least to say, "Well, that is not our recollection." They would still have to have the health professional involved, but they may even help with the quality of the data. We would certainly welcome greater openness from a patient perspective.[100]

52. **Patients and GPs are more likely to be content for their personal data to be used for healthcare and medical research if the benefits—both to the individual and to society—are clearly explained and adequate safeguards are in place. But the track-record of 'care.data' shows that this cannot be taken for granted. The Government cannot afford a second failure from a re-launched scheme. *The Government should take careful account of the lessons from the pathfinder projects as well as the experience of the similar, successful, scheme in Scotland. To help bring patients onside and to streamline healthcare across different NHS providers—hospitals, GPs, pharmacists and paramedics—it should give them easy online access to their own health records.***

## Private sector data sharing

53. The Royal Academy of Engineering stated in its 2015 *Connecting Data* report that:

> Much potentially valuable data remains locked away in corporate silos or within sectors, although some data is already traded within the supply chains of individual sectors. The next step, in areas that do not impinge on the privacy of personal data, should be the creation of platforms to enable proprietary datasets to be traded within a framework that promotes trust and practicality.[101]

97 HM Treasury, *Spending Review and Autumn Statement 2015*, Cm 9162 (November 2015), para 1.101
98 Q28
99 Q23
100 Q23
101 Royal Academy of Engineering and Institution of Engineering & Technology, Connecting data: Driving productivity and innovation (November 2015)

The Digital Catapult is intended to facilitate this. It is one of Innovate UK's expanding network of Catapult Centres, which are designed to support innovation in specific technology areas by providing access to expert technical capabilities, equipment, and other resources. The Digital Catapult aims to "help UK businesses unlock new value from sharing proprietary data in faster, better and more trusted ways".[102]

54. Chirdeep Chhabra of the Digital Catapult was concerned that "we have yet to see enough sharing of data … between different silos. … We need to start off by enabling sharing of data between organisations".[103] He told us that "there is much more to be done" in taking forward the Government's work on developing the UK's big data capability. That included more work to facilitate "safe" data sharing:

> Mixing data from different sources, silos of data, … that is where I think we need more initiatives … 'data-sharing labs'.[104]
>
> Some of the things we are doing are to create safe havens where organisations can bring their datasets together. They are not giving data to each other; of course, they will not do that for governance and business reasons, but organisations are realising more and more that they can only get benefit from their data by mixing it with other data … . There is a lot of friction in data sharing, in terms of legal governance and so on. That is where we need to take a lead.[105]

55. Funding Circle believed that in their fintech sphere the Small Business, Enterprise and Employment Act 2015 could improve data sharing:

> [The Act imposes] a duty on designated banks to provide information about their small and medium sized business customers to designated credit reference agencies, and a duty on designated credit reference agencies to provide that information to finance providers. This … will be incredibly helpful as businesses currently have to provide bank statements themselves. … The new Act will mean this process is now automated (with businesses' consent), allowing us to speed up the process and enable creditworthy businesses to access finance faster.[106]

Hetan Shah from the Royal Statistical Society suggested an area where further legislation could be beneficial, to give the Office for National Statistics access to privately held datasets:

> The very interesting thing about Canada and New Zealand is that they have also mandated private sector data to be open to their statistical offices, and the private sector has said, "We are glad we are being put on a level playing field, because if I was volunteering my data to you I would be at a competitive disadvantage, but if we all have to give our telecoms or supermarket data it does not matter." That sort of legislation, in the mould that other more forward-thinking countries are taking, would be the right way forward.[107]

102 Digital Catapult website, accessed January 2016
103 Q37
104 Q44
105 Q54
106 Funding Circle (BIG0081)
107 Q203

56. **While the private sector is making great strides in identifying opportunities for bringing different datasets together, it is understandably more challenging for businesses in a competitive market to share valuable data with one another or with Government. The Government's Digital Catapult therefore plays a vitally important role in facilitating private sector data sharing in a 'safe', trusted environment.** ***The Government should map out how the Catapult's work and its own plans to open and share Government data could be dovetailed. The Government should also consider the scope for giving the Office for National Statistics greater access both to Government departments' data (paragraph 41) and private sector data.***

# 5 Data protection

57. Big data, as we have discussed, has enormous potential to improve the services we receive and the way businesses operate. When big data deals with personal or commercially sensitive data, however, those benefits have to be weighed against a potential loss of privacy and the risks of our data being lost or misused. The Information Commissioner highlighted that "every week there seems to be a horror story at the top of my pile, and my investigation team and my enforcement team, which is being expanded, has more and more work to do."[108] Chris Combemale from the Direct Marketing Association saw "a constant battle between data security and hackers".[109]

58. Big Brother Watch found in a 2015 poll that 79% of adults in the UK were "concerned" about their privacy online, and 46% believed that they were "being harmed by the collection of their data by large companies".[110] The Direct Marketing Association found that 60% of people were "happy with the amount of personal information that they shared with companies", and 47% considered that "the exchange of personal data is essential for the smooth running of modern society".[111] Such surveys might appear inconsistent, or else point to internal conflicts in people's attitudes towards big data. As Baroness Shields, minister for internet safety and security, pointed out:

> There is a chasm in terms of what people feel about trusting data. If you talk to teenagers, they do not care; they have given up privacy and decided that they are happy to share absolutely everything in their lives and have it catalogued. … We have some responsibility to look out for their interests, especially in terms of their rights online.[112]

59. A further 2015 survey from the Digital Catapult on people's attitudes towards sharing data found that the Government was the most trusted user of personal data: 44% of respondents named it as the most trusted sector, with financial services in second place with 29%. However, this trust comes with strings attached: 32% considered it the Government's responsibility to educate people about protecting or controlling personal data, while 30% thought it the responsibility of the individual.[113]

60. **It is important to note that personal data is only a small proportion of big data—there is huge value still to be realised from novel uses of non-personal datasets like transport data, weather data, etc. Nevertheless, given the scale and pace of data gathering and sharing, distrust arising from concerns about privacy and security is often well founded and must be resolved by industry and Government if the full value of big data is to be realised. We recommend below the establishment of a Council of data Ethics to help address these issues (paragraph 102).**

108 Q148
109 Q105
110 Big Brother Watch, Com Res, UK public research: Online privacy (2015)
111 Direct Marketing Association (BIG0014)
112 Q239
113 Digital Catapult, Trust in personal data: a UK review (2015)

## Data protection regulation and consent

61. Controls on the storage and processing of personal data are covered by the Data Protection Act 1998, which transposed the 1995 EU Data Protection Directive. The 1998 Act set out a number of 'data protection principles':

(1) Personal data shall be processed fairly and lawfully, with the subject's consent, by necessity or for the data controller's legitimate interests.

(2) Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.

(3) Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed.

(4) Personal data shall be accurate and, where necessary, kept up to date.

(5) Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes.

(6) Personal data shall be processed in accordance with the rights of data subjects under this Act.

(7) Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.

(8) Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.[114]

The Data Protection Act also makes special provisions for 'sensitive personal data', such as patients' health data.

62. The Information Commissioner's Office has emphasised that big data is subject to existing data protection legislation:

> We do not accept the argument that data protection principles are not fit for purpose in the context of big data. Big data is not a game that is played by different rules.[115]

63. Consent is one of the conditions which allow an organisation to process personal data—the first 'data protection principle' under the 1998 Act. Of course, as many witnesses pointed out, there are circumstances which make it more challenging to secure consent in big data. The Information Commissioner acknowledged this:

> If an organisation is collecting personal data to use in big data analytics, and it is relying on consent to legitimise this, then it has to make people aware of all the intended uses of the data, including, for example, whether it is going to share

114 Information Commissioners Office, Data protection principles
115 Information Commissioners Office, *Big data and data protection* (July 2014)

> the data with other organisations. Similarly, if an organisation is acquiring data from elsewhere, it has to satisfy itself that the original consent covers that further use of the data. Given the complex and sometimes unforeseen uses of data in big data analytics, this can of course be problematic.[116]

However, he was clear that these challenges could be overcome in most circumstances and emphasised that "the consent must be freely given, specific and informed … Furthermore, 'freely given' means that people can also withdraw their consent".[117]

64. Along with the Digital Marketing Association,[118] the Digital Catapult,[119] Big Brother Watch[120] and others, the Information Commissioner was clear that informed consent started with a comprehensive Privacy Impact Assessment to identify and mitigate privacy risks followed by innovative solutions for delivering meaningful, transparent 'privacy notices'.[121]

65. Terms & conditions and privacy notices are the primary mechanism for obtaining consent but many are so dense and opaque that they actively prevent rather than enable informed consent. In offering advice to a company seeking to obtain genuine consent, the Information Commissioner had some straightforward suggestions. Privacy notices, he said, should be:

- In English;
- Not written by lawyers;
- Not the length of a 'short Shakespeare play'.[122]

A combination of consumer demand, reputational concerns and legislative pressure are beginning to have effect. Google is trialing a layered approach to privacy notices, while Facebook has updated its privacy policy, cutting the word count by 70%.[123] Simplenote, a company which provides digital note-taking services, has ensured that its terms and conditions are only 140 words written in plain English.[124]

66. **Businesses and governments that communicate most effectively with the public, giving the citizen greater control in their data transactions by using simple and layered privacy notices to empower the consumer to decide exactly how far they are willing to trust each data-holder they engage with, will gain a huge commercial and societal advantage. Although the length of a privacy notice will be dictated by the service or data application involved, it should be best practice to draft them as simply as possible. Furthermore, if informed, freely given consent must be the bedrock of a trusting relationship between a consumer and a data-holder, then it must always be part of that deal that consent freely given can also be freely withdrawn.**

116 Information Commissioners Office (BIG0033), para 27
117 Ibid.
118 Direct Marketing Association (BIG0014)
119 Digital Catapult (BIG0080)
120 Big Brother Watch (BIG0006)
121 Q155; Information Commissioners Office (BIG0033)
122 Q155
123 Q155
124 As reported by Big Brother Watch (BIG0006)

## Anonymisation and penalties

67. Nevertheless, the Information Commissioner, Christopher Graham, raised concerns with us about big data techniques which 're-identify' individuals when previously anonymised data are combined with other datasets. The Wellcome Trust worried that "as datasets become more sophisticated … the technical possibility of undertaking 'jigsaw' re-identification of individuals increases, even from data that has been through a process of anonymisation".[125] Professor Jonathan Montgomery, chair of the Nuffield Council on Bioethics, noted that "in the context of health data, the difficulty is magnified by the fact that in order to be useful the data has to be quite rich about your health: The richer it is, the more possible it becomes to use those techniques to correlate."[126] He also told us that work on data breaches by the Nuffield Council on Bioethics showed that "most of the breaches are human as opposed to technological. That suggests that … some of the techniques we have used previously to safeguard privacy and confidentiality remain important, [such as] personal integrity and the quality of staff."[127]

68. The Information Commissioner was clear that more work needs to be done on researching anonymisation and raising industry standards.[128] The UK Anonymisation Network (UKAN) was set up by the Information Commissioner's Office in 2012 as a means of establishing best practice in this area. It aims to maximise the value of data, minimise the risks to privacy and preserve public confidence by collating best practice in anonymisation from a wide range of experienced practitioners.[129] Our witnesses were clear that there are technical options available to improve the effectiveness of anonymisation protocols; one method seen as more tried and tested than the others being 'differential privacy'.[130]

69. Differential privacy aims to ensure that the results derived from a dataset would look the same whether a person's data was included in the dataset or not. This is achieved by adding noise to the dataset in a way which does not interfere with the accuracy or outcome of results. Microsoft noted in 2011 that:

> Differential privacy thrives because it is natural, it is not domain-specific, and it enjoys fruitful interplay with other fields. … This flexibility gives hope for a principled approach to privacy in cases, like private data analysis, where traditional notions of cryptographic security are inappropriate or impracticable.[131]

Despite the academic enthusiasm for differential privacy, it is a system that is rarely deployed. Whilst differential privacy may not be a silver bullet, it requires greater research and further exploration to establish it as a method aimed at addressing privacy concerns.

125 Wellcome Trust (BIG0058)
126 Q132
127 Q99
128 Information Commissioner oral evidence
129 UKAN website, accessed February 2016
130 Big Brother Watch (BIG0006)
131 Cynthia Dwork, Microsoft research, *A Firm Foundation for Private Data Analysis* (2011), as reported by Big Brother Watch (BIG0006)

70. Furthermore, the Information Commissioner was concerned that 're-identification' of individuals may be outside the scope of the current legislation:

> Section 55 [of the Data Protection Act] just deals with the unauthorised obtaining or disclosure of personal information without the knowledge of the data controller. I would not like to test a de-anonymisation or re-identification case against that.[132]

Ed Vaizey told us:

> The original intention of section 55 was to address the problem of third parties obtaining personal data by deception and most prosecutions under this provision have dealt with these types of offences. It is unlikely that it was intended for the purposes of dealing with the de-anonymisation, which was not thought to be a major issue at the time.[133]

71. The Information Commissioner described the range of methods he uses to ensure compliance with the data protection requirements covered by the regulations. These included data protection audits which, though generally voluntary, may be conducted compulsorily on government departments[134] (including on NHS authorities since February 2015).[135] The Information Commissioner believed that compulsory audits "ought logically to apply to local government as well",[136] and possibly even to some commercial sectors.[137]

72. The Information Commissioner can also impose civil monetary penalties of up to £500,000.[138] He was primarily "interested in a way of doing enforcement that gets a result that is of benefit to consumers, rather than just getting off on civil monetary penalties".[139] But he was nevertheless frustrated that, in cases of serious or malicious breaches of data protection, the current penalties were not sufficient and there was a lack of the clout from criminal sanctions:

> I take cases before magistrates' courts and I weep when the fine is £250 and a £100 community sentence … It does not provide the deterrent that I want. We will continue, but it would make our investigations much easier if I could require people to attend for interview rather than asking them nicely. It would make the investigations quicker. If the offences were recordable and on the police national computer, that itself is a disincentive.[140]

He wanted serious breaches of the Data Protection Act to become criminal offences. This, he told us, could happen "in very short order" if sections 77 and 78 of the Criminal Justice and Immigration Act 2008 were commenced.[141]

132 Q177
133 Letter from Minister Ed Vaizey MP (BIG0089)
134 Data Protection Act 1998, Section 41A
135 Information Commissioners Office, ICO given new powers to audit NHS, News item, 2 February 2015
136 Q184
137 Q185
138 Qq149, 152, 156, 169
139 Q156
140 Q179
141 Qq161, 176-178

73. Such a recommendation has been made previously by the 2012 Joint Committee on the Communications Data Bill, the Home Affairs Committee in 2012, the Justice Committee in 2013, and a number of witnesses to our current inquiry.[142] The Wellcome Trust, for example, advocated criminal sanctions for misusing personal data, especially "unauthorised and unwarranted deliberate re-identification of individuals through big data technologies"[143].

74. The Association of Medical Research Charities was more cautious:

> We need to bear in mind that we do not want to make a system that is already very risk-averse even more risk-averse … We must not leave out the education piece if we go down the criminal sanction route.[144]

Minister Ed Vaizey also advised caution. When we asked him if there were any particular reasons why he might not introduce criminal penalties, he replied:

> The fact that it is a criminal offence maliciously to access data with the intent, effectively, to misuse that data covers what perhaps the ordinary person in the street would regard as a criminal act. I would not want criminal legislation inadvertently to catch people who have been negligent, however much they might be condemned for their negligent behaviour in allowing your data to become available. We would have to think very hard, if we were to introduce criminal penalties, about what kind of behaviour they would catch.[145]

Subsequently, the Minister wrote to us, saying:

> The forthcoming [EU] General Data Protection Regulation will give us an opportunity to stress test the existing sanctions available in relation to the misuse of personal data to make sure they are fit for purpose for the digital age. In particular, we will review current penalties for data protection breaches and aim for sanctions that act as effective deterrents against the misuse of personal data in all contexts.[146]

75. **As citizens' personal data is being used in ever increasing volumes and for ever changing purposes, it is vital that the Information Commissioner's Office has the powers it needs to help ensure data protection.** ***With a new EU data protection regulation now agreed (paragraph 83), we welcome the Government's commitment to review current penalties for data protection breaches. The Government should nevertheless introduce as soon as possible a criminal penalty for serious data protection breaches by commencing sections 77 and 78 of the Criminal Justice and Immigration Act 2008. The Government should not regard the two-year implementation period of the recently agreed EU data protection regulation, which will provide for bigger fines (paragraph 83), as a reason for delaying this.***

142 Joint Committee on the Draft Communications Data Bill, *Draft Communications Data Bill,* HC (2012–13) 479, para 226; Home Affairs Select Committee, *Private Investigators*, HC (2010–12) 100, para 47; Justice Committee, *The functions, powers and resources of the Information Commissioner*, HC (2012–13) 962, para 39.

143 Wellcome Trust (BIG0058)

144 Q30

145 Q259

146 Letter from Minister Ed Vaizey MP (BIG0089)

76. ***The Government should agree to the Information Commissioner's request to perform compulsory data protection audits on local government.***

77. ***The Government should set out its anonymisation strategy for big data in its upcoming Digital Strategy, including a clear funding commitment, a plan to engage industry with the work of the UK Anonymisation Network and core anonymisation priorities.***

## Kitemark

78. The Information Commissioner has concluded that, regardless of whether citizens feel threatened by big data, "they feel they have lost control over their personal information".[147] It is vitally important, therefore, that individuals give their informed consent to the use and sharing of their personal data—one of the 'data protection principles' in the Data Protection Act 1998 (paragraph 61). Our predecessor Committee concluded in 2014 that people's ability to provide informed consent was undermined by the "opaque, literary style" of terms and conditions documents, which "renders them unsuitable for conveying an organisation's intent for processing personal data to users".[148]

79. In our current inquiry, Chris Combemale from the Direct Marketing Association believed that a business's desire to maintain its brand reputation provided a key incentive for good practice.[149] James Meekings from Funding Circle, in a similar vein, emphasised the importance of following best practice and transparency in establishing a trustworthy brand.[150] The Nuffield Council on Bioethics argued that data users had a responsibility to protect personal data that addressed the 're-use' of data for purposes not envisaged when consent was originally obtained—another 'data protection principle' (paragraph 61):

> Where a person providing data about themselves cannot foresee or comprehend the possible consequences of how their data will be available for linkage or re-use, consent at the time of data collection cannot, on its own, protect all of their interests. … The changing context and potential for data re-use means that compliance with the law is not enough to ensure a data initiative is ethically appropriate. Those who manage data initiatives therefore have a continuing duty to promote and protect the legitimate rights and interests of those who have provided data about themselves irrespective of the terms of any consent given.[151]

80. While the specific duties of the Information Commissioner are set out in legislation, including those concerned with consent, Hetan Shah of the Royal Statistical Society saw a need for an oversight body to help ensure good practice with big data more generally:

> Regulation is always lagging behind new technology and developments … What we are hoping for is that the Alan Turing Institute will take a lead in thinking through the ethics around big data. In the US there is a council of

147 Q145
148 Science and Technology Committee, Responsible Use of Data, Fourth Report of Session 2014–15, HC 245, para 49
149 Q99
150 Qq70-71
151 Nuffield Council on Bioethics (BIG0015)

> ethics on big data, and I wonder whether the UK needs something similar to take forward this agenda.[152]

81. Good practice can be encouraged by acknowledging it through the use of kitemarks, as our predecessor Committee recommended.[153] In their inquiry, the Government highlighted the work on the collection of personal information by the British Standards Institution and the Information Commissioner's 'privacy seal' programme.[154] The Information Commissioner told us in our current inquiry that:

> The idea of a privacy seal is that it is beyond the ISO standard; it is something people can recognise as a good housekeeping seal of approval on sites that sign up to doing things properly, and are prepared to be audited for doing that. … I think that will give consumers a way of recognising that this is a serious player that understands privacy and is committed to looking after their data. That will give those companies a competitive advantage.[155]

82. **The Information Commissioner has developed a data protection kitemark, ready for use now. The use of such kitemarks, acknowledging good behaviours, would complement the greater sanctions of criminal penalties for bad behaviours that we have recommended. *The Government and Information Commissioner should work with industry to ensure that the UK's already developed kitemark is adopted as soon as possible, and initiate a campaign to raise public awareness of it.***

## New EU regulations

83. The European Parliament, Commission and Council agreed a General Data Protection Regulation in December 2015.[156] It will now require changes within the next two years[157] to the UK's Data Protection Act 1998, which transposed the existing 1995 EU Data Protection Directive. The Commission proposed the new Regulation, as part of a larger 'Data Protection package', in January 2012.[158] The provisions in the EU Regulation include changes to, or restatements of, requirements in the existing Directive:

- A 'right to be forgotten', whereby individuals have the right to withdraw their consent and request their data be deleted—a new right;
- A requirements for consent to be explicit, rather than implied, and to be requested in clear and plain language;
- Access to one's own data and the right of data portability, whereby customers can transfer a copy of personal data from one service provider to another—a new right;
- A requirements for organisations to notify serious data breaches within 24 hours—a new requirement;

152 Q190
153 Science and Technology Committee, Responsible Use of Data, Fourth Report of Session 2014–15, HC 245, para 69
154 Science and Technology Committee, Responsible Use of Data, Fourth Report of Session 2014–15, HC 245, paras 46-47
155 Q171
156 European Commission, Agreement on Commission's EU data protection reform will boost Digital Single Market, Press release, 15 December 2015
157 Implementation will be required within two years after the Regulation is formally published by the Commission.
158 European Scrutiny Committee, *Eleventh Report,* 2015–16, HC 342xi, para 2.1

- A 'one-stop-shop' mechanism, where an information commissioner in one state could deal with data processing across borders—a new provision; and
- Tougher penalties for breaching the regulations, with the maximum fine increasing from £500,000 (the current fining limit of the UK Information Commissioner) to €100 million or 5% of annual worldwide turnover, whichever is greater.

84. The Regulation has taken four years to be finalised between the three EU institutions under the trilogue 'ordinary procedure' negotiation process. In 2012 the House of Commons Justice Committee concluded that the proposed Regulation in its original prescriptive form would not produce a "proportionate, practicable, affordable or effective system of data protection".[159] The Council's position, in putting forward its own revisions, was "broadly more pro-business than the European Parliament".[160]

85. The UK Government's approach appears to have supported the Council's position during the negotiations. The European Scrutiny Committee stated in December, before the Regulation was agreed, that it was "unconvinced" that the Government's "unusual approach of supporting a [draft] text with 'serious reservations' would lead to greater negotiating influence over the text in trilogues [negotiations]". These reservations were about the provisions dealing with 'the right to be forgotten', the 'one-stop-shop mechanism, and the liability and sanctions faced by data controllers.[161]

### *The scope of permitted data collection and processing*

86. Ed Vaizey MP told the European Scrutiny Committee in 2015 that a strand of the UK's negotiating approach on the Commission's Digital Single Market package, would focus on

> Encouraging innovation: We should support an approach that is light-touch and flexible enough to respond to rapid technological changes; in particular, we do not want regulation to close down innovation and the potential of fast-moving technologies such as big data and cloud computing.[162]

87. The European Parliament had generally sought to strengthen the meaning of consent in the draft Regulation; the Council to weaken consent and to widen some grounds legitimising processing. For example, the provision on data minimisation was weakened by the Council's draft which deleted the requirement that personal data "shall only be processed if, and as long as, the purposes could not be fulfilled by processing information that does not involve personal data". [163] The Council draft introduced data collection and processing exceptions for 'statistical', 'scientific' and 'historical' purposes—regimes "which might describe some big data operations".[164] The 1995 Data Protection Directive only addressed statistical and historical processing, and "only incidentally".[165] The European Parliament's proposals started with a prohibition on processing unless it would satisfy a number of conditions; the Council's version (that was subsequently included in the

159 Justice Committee, *The Committee's opinion on the European Union Data Protection framework proposals*, HC (2012–13) 572, para 105, as reported also in European Scrutiny Committee, *Fifth Report*, 2015–16, HC 342v, para 5.2
160 Prof Lorna Woods (BIG0082)
161 European Scrutiny Committee, *Eleventh Report,* 2015–16, HC 342xi, para 2.10
162 European Scrutiny Committee, *Fifth Report*, 2015–16, HC 342v, para 4.21
163 Prof Lorna Woods (BIG0082)
164 Prof Lorna Woods (BIG0082)
165 Prof Lorna Woods (BIG0082)

final Regulation) started from a position allowing personal data processing for scientific, statistical or historical purposes but subject to safeguards.[166]

88. Ed Vaizey had told us that the sticking points in the negotiations were about "the level of burden on business. We do not want to place too many onerous reporting requirements on business. We want to make sure we get that balance absolutely right."[167] It appears from the recently agreed Regulation text that the Government's and the European Council's concerns have been met. A number of provisions allow data to be collected, retained or processed for "scientific research" purposes or more generally "in the public interest". Article 5 stipulates, for example, that:

> Personal data must be … collected for specified, explicit legitimate purposes and not further processed in a way incompatible with those purposes; [but] further processing of personal data for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes, shall … not be considered incompatible with the initial purposes …
>
> Personal data may be stored for longer periods insofar as the data will be processed solely for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes.

89. The agreed Regulation allows processing of data if "necessary for the purposes of the legitimate interests pursued by the [data] controller".[168] More fundamentally, the Regulation includes a clause inserted by the Council that allows states to "maintain or introduce more specific provisions to adapt the application of the rules of this Regulation with regard to the processing of personal data" to meet a national legal obligations (such as those for investigatory powers)[169] or to perform tasks carried out in the public interest.[170]

90. There may be a debate still to be had about whether data collection and processing that satisfies 'the public interest' would include private sector activities. The Regulation leaves the term undefined.

## Consent

91. During the negotiations of the Regulation there had been arguments over the nature of the consent that people would have to give. The European Council's assessment of the final agreed Regulation concluded that "the way in which consent is to be given by data subjects remains 'unambiguous' for all processing of personal data, with the clarification that this requires a 'clear affirmative action', and that consent has to be 'explicit' for sensitive data."[171] Article 7 requires that the data controller will have to be able to demonstrate that consent was given by individuals to the processing of their personal data, and that seeking consent must be presented "in an intelligible and easily accessible form, using clear and plain language".[172]

166 Article 83; Prof Lorna Woods (BIG0082)
167 Q254
168 Article 6(1)
169 Science & Technology Committee, *Investigatory Powers Bill: Technology issues*, HC (2015–16) 573
170 Article 6(2a)
171 European Council, Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) [first reading]: Analysis of the final compromise text with a view to agreement, 15 December 2015
172 Ibid.

92. There had been particular concerns about possible restrictions on the use of medical records for big data processing. Before the EU regulation was agreed, some witnesses were concerned that consent requirements might become too onerous. TechUK, for example, warned that:

> It is important that the [Regulation] discussions do not result in the introduction of a narrow consent requirements that are not adaptable to citizens' expectations nor to their online behaviour …. Such a move would risk 'consent fatigue' or worse 'meaningless consent' whereby overly burdensome requirements on consumers … could undermine the willingness of consumers to navigate preferences and understand how their data is being used.[173]

93. The medical research community, in particular, was concerned that more stringent consent requirements would be extremely restrictive in a sector where data is often re-used and re-purposed as techniques develop:

> When the regulations started out, it was clear that a separate case was made for the research piece, and then it got amended. We are very worried that, if it goes ahead, medical research will be damaged and become unworkable, which will not benefit us at patient/public level; nor does it help all the investment that has been made in this particular area. It will all be for nothing. It is a matter of great concern at the moment.[174]

The Information Commissioner seemed unconvinced that exceptions should be made for medical research. He commented that:

> The Information Commissioner is sufficiently imaginative to see the power of big data in medical research and in the delivery of health services … but we want to see things done in the right way so that people's fundamental rights and privacy are not trashed in the name of some higher obligation to efficiency and the onward march of science.[175]

94. The Minister told the European Scrutiny Committee, before the Regulation was agreed in December, that the Council draft text:

> does not prevent the processing of NHS medical records data for research purposes … The European Parliament's text however, would appear to significantly restrict processing for research purposes. The UK has been very clear that the position under [Council's] General Approach must be preserved …[176]

In the event, the final agreed text retained the 'public interest' and 'scientific and historical research' exemptions; meeting the minister's (and, it would appear, the medical research community's[177]) concerns.

173 Tech UK (BIG0039)
174 Q29
175 Q158
176 European Scrutiny Committee, *Eleventh Report,* 2015–16, HC 342xi, para 2.15
177 Academy of Medical Sciences, Positive outcome for research from EU data laws, News item, 23 December 2015

### 'Profiling'

95. The EU Data Protection Regulation provides some safeguards where data processing profiles people according to particular characteristics. It essentially prohibits 'profiling' of people according to characteristics which would normally be discriminatory, except for 'public interest' or legal requirements.[178] It allows someone to object to the processing of their personal data even when processed according to those conditions, and the data controller would only be able to continue to process the personal data if able "to demonstrate compelling legitimate grounds".[179] The agreed EU Regulation gives a "right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her", unless concerned with contracts or satisfying national laws.[180]

### Data portability

96. The agreed EU Regulation includes provisions on 'data portability'—allowing individuals to re-use their personal data. There is no equivalent in the 1995 Data Protection Directive.[181] The Regulation stipulates that the data would have to be provided to the individual "in a structured and commonly used and machine-readable format, and [they would] have the right to transmit those data to another controller without hindrance".[182]

97. On data portability, the Government has already introduced the 'midata' initiative,[183] under which a Current Account Switch Service has been established.[184] Midata introduced a portable data format which allows consumers to use their own consumption or transaction history to compare products and services in the energy, finance and telecoms sectors.[185] The Current Account Switch Service facilitates the automatic transfer of all credit and debit instructions associated with an account. Imran Gulamhuseinwala from EY consultants, a member of the steering committee of the Open Bank Working Group, told us that:

> There is a broad feeling that midata has been a very interesting, robust first step in enabling consumers to understand that they have transaction level data; it belongs to them and they can also use it for their own benefit. Midata has a very narrow use case, which is about trying to shop around for the best current account. None the less, it feels that it is beginning to move in that direction.[186]

98. Under the new Regulation these UK initiatives will acquire a statutory footing, with enforceable rights for individuals.

178 Article 9
179 Article 19
180 Article 20
181 Prof Lorna Woods (BIG0082)
182 Article 18
183 HM Treasury, Bank account switching service set to launch, News item, 10 September 2013
184 BIS, The midata vision of consumer empowerment, News item, 3 November 2011
185 BIS, New plans to make switching suppliers easier for consumers, News item, 22 October 2015
186 Qq82-83

### *Implementing the EU regulation*

99. The new Regulation must be implemented within two years of it being formally published by the Commission—expected soon. We asked the minister, Ed Vaizey MP, whether in the meantime national data protection safeguards were sufficient. He felt that:

> We can live with them as they are. … I do not think it would be sensible to have any kind of interim measures between our current regulations and the future regulation because that would be confusing for business. … You have to have a conversation and dialogue with business about the best way of implementing the regulation.[187]

100. **The Data Protection Act will have to be revised to accommodate the recently agreed EU Data Protection Regulation, which will come into force with the next two years or so. We do not share the Government's view that current UK data protections can simply be left until then. Some areas in particular need to be addressed straightaway—introducing the Information Commissioner's kitemark (paragraph 78), and introducing criminal penalties (paragraph 72) rather than relying only on the prospective greater fines envisaged by the new EU Regulation.**

101. **The new EU Regulation appears to leave it open for data to be re-used, and potentially de-anonymised, if "legitimate interests" or "public interest" considerations are invoked. This is an issue that urgently needs to be addressed as big data becomes increasingly a part of our lives. There are arguments on both sides of this issue: Seeking to balance the potential benefits of processing data (some collected many years before and no longer with a clear consent trail) and people's justified privacy concerns will not be straightforward. It is unsatisfactory, however, for the matter to be left unaddressed by Government and without a clear public-policy position set out.** ***The Government should therefore clarify its interpretation of the EU Regulation on the re-use and de-anonymisation of personal data, and after consultation introduce changes to the Data Protection Act 1998 as soon as possible to strike a transparent and appropriate balance between the benefits of processing data and respecting people's privacy concerns.***

102. ***Given the UK's leading position in big data and the Government's stated commitment to capitalise on the potential innovation and research opportunities it promises, the Government should establish a Council of Data Ethics within the Alan Turing Institute as a means of addressing the growing legal and ethical challenges associated with balancing privacy, anonymisation, security and public benefit. Ensuring that such a Council is established, with appropriate terms of reference, offers the clarity, stability and direction which has so far been lacking from the European debate on data issues.***

187 Q255

# Annex

## Committee seminar, University of Oxford

On Thursday 19 November 2015, the Bodleian Library, in collaboration with the Department of Politics and International Relations at the University of Oxford, hosted a one day event for Oxfordshire Sixth Formers to learn more about parliamentary representation and the work of select committees. It formed part of 'Parliament Week', a programme of events and activities that connect people across the UK with Parliament and democracy.

Members of the Committee joined the seminar in the Divinity School of the Bodleian Library to discuss the big data with two experts. The sixth formers present were also able to ask questions. (The seminar also discussed the Ebola outbreak, and we included notes on that subject in our *Science in emergencies: UK lessons from Ebola* report.)[188]

Members present:

- Nicola Blackwood MP, Chair
- Chris Green MP
- Carol Monaghan MP
- Derek Thomas MP
- Valerie Vaz MP

The panel comprised:

- Helen Margetts, Director of the Oxford Internet Institute, and Professor of Society and the Internet
- Dr Neil Geddes, Director at the Science and Technology Facilities Council.

The panel noted that the benefits of big data cannot be easily valued or costed, in part because it overlapped to some degree with other long-established analytical techniques and processes. 90% of data was created in the past two years. New skills are needed for big data, beyond those usually captured under the 'STEM' heading, concerned with how to store, access and protect and share data, as well as the processing of it.

Ethical issues arose where big data involves dealing with personal data. In particular, issues around obtaining 'informed consent' from 'data subjects' were difficult to deal with because big data often involves bringing diverse datasets together to bring out new insights and in the process re-analysing personal data that may have collected a long time beforehand. It could be difficult to trace those who had given the original consent to use the data when first collected, to be able to ask them for consent for the new use of the data. The bringing together of different datasets which have been 'de-anonymised' can also potentially make it possible to 're-identify' individuals. We need to establish a balance between security and privacy, and the trade-offs made between the two should be made much more explicit.

188 Science & Technology Committee, *Science in emergencies: UK lessons from Ebola*, HC (2015–16) 469

Privacy issues may relate to individuals, but also groups of people collectively who share particular traits or circumstances. The Enigma decodings in the Second World War presented the Prime Minister with a dilemma about whether to use the acquired data to evacuate Coventry before a bombing raid—new unforeseen dilemmas might also arise from the results of big data analysis.

The meaning of 'informed consent' may need to be rethought. It will depend to some degree on different cultures. Scandinavians were generally more amenable to their data being used by the State, which is more open with the public about what consent means and what the government will and will not do with public data. Penalties with appropriate repercussions for improper use of big data need to be part of a consent regime. Consent is generally also likely to be easier to obtain in cultures where the big data benefits for society are clearly seen, where processes and results are transparent, and where people can readily see the policy trade-offs being contemplated from those results. People need a vehicle for have an intelligent discussion about how data will be used to allow them to give consent which is 'informed'. Consent should be built in to everything we do with data.

Problems arise when people think their data is being sold and feel powerless to stop this. Particular care is needed where data relates to minors, who may need additional protections about how their data are used and before their consent is obtained.

A danger is that technologies are how data can be used are advancing more quickly than the skills are evolving which will be needed to use the data safely and to protect privacy. It is people rather than technologies that go wrong in most systems.

Some big data are not about personal data, but commercial or research data. CERN is collecting massive volumes of non-personal data from the Hadron Collider.

# Conclusions and recommendations

## The opportunities for big data

1. The UK is a world leader in big data research across disciplines and our Tech sector, especially in London, dramatically outperforms the rest of the economy on growth and productivity indicators. By identifying big data as one of the Eight Great Technologies, and investing significant financing in large scale data infrastructure, the Government has signalled that realising the full potential of big data is a priority. However, investing in capital infrastructure projects alone will not deliver this. Urgent action on the digital skills crisis, overcoming public distrust over data sharing, further progress on 'open data' and greater clarity over prospective data protection legislative changes are essential if the UK is to set the pace on big data. We discuss these pre-requisites [below]. (Paragraph 19)

## Skills and Infrastructure

2. The digital skills gap is approaching crisis levels and this not only has economic implications but also puts the quality and security of this data at risk. There is a range of Government initiatives to help develop computing and digital skills, but the wider set of 'big data' skills is not being strategically addressed. This risks UK business being unable to grow the big data sector at the pace it should. In the meantime, this skills gap is forecast to grow exponentially as big data reaches further into the economy. The evidence we received on the digital skills crisis was so concerning that we have launched a further inquiry specifically into this issue on which we will report shortly. *In the meantime, the Government should commit to (a) a continuing substantial role in developing data analytics skills in businesses, with others already working in this field; (b) increasing big data skills training for staff in Government departments; and (c) promoting more extensively the application of big data at local government level. But the Government must also address the wider context of its policies on apprenticeships and immigration control. As it develops its approach in these areas, it should explicitly address widespread concerns that these could jeopardise the necessary big data skills-base that the UK will increasingly need. The Government should also set out in detail how the Government Digital Services' budget, including the additional funding announced in the Spending Review, will be spent.* (Paragraph 27)

3. While investment for big data research is welcome, we believe that the Government should explore further ways of making publicly-funded infrastructure and expertise available to more businesses. The Digital Catapult is a good start but it is essential that ongoing resource investment in the Catapult is maintained so that it can consolidate and expand its work. As big data becomes an increasingly significant part of our economy, the Government should set out its strategy for longer term big data infrastructure development and how it will work with industry to provide a coherent programme of business support. (Paragraph 31)

## Open data and data sharing

4. There are enormous benefits in prospect for the economy and for people's lives from making the nation's core data infrastructure 'open'. The Government's work in this area has put the UK in a world-leading position. But there is more to do to breakdown departmental data silos, to bring data together in order to further improve public services, as well as to improve data quality. *The Government should set out how it can build capacity to deliver more datasets, increasingly in real-time, both to decision-makers in Government and to external users and, in particular, should work to establish a right of access to data for the Office for National Statistics. The Government should also establish a framework—to be overseen by the Government Digital Service, the Office for National Statistics or another expert body—for auditing the quality of data within Government departments amenable for big data applications, and for pro-actively identifying data sharing opportunities to break departmental data silos.* (Paragraph 42)

5. Patients and GPs are more likely to be content for their personal data to be used for healthcare and medical research if the benefits—both to the individual and to society—are clearly explained and adequate safeguards are in place. But the track-record of 'care.data' shows that this cannot be taken for granted. The Government cannot afford a second failure from a re-launched scheme. *The Government should take careful account of the lessons from the pathfinder projects as well as the experience of the similar, successful, scheme in Scotland. To help bring patients onside and to streamline healthcare across different NHS providers—hospitals, GPs, pharmacists and paramedics—it should give them easy online access to their own health records.* (Paragraph 52)

6. While the private sector is making great strides in identifying opportunities for bringing different datasets together, it is understandably more challenging for businesses in a competitive market to share valuable data with one another or with Government. The Government's Digital Catapult therefore plays a vitally important role in facilitating private sector data sharing in a 'safe', trusted environment. *The Government should map out how the Catapult's work and its own plans to open and share Government data could be dovetailed. The Government should also consider the scope for giving the Office for National Statistics greater access both to Government departments' data and private sector data.* (Paragraph 56)

## Data protection

7. It is important to note that personal data is only a small proportion of big data—there is huge value still to be realised from novel uses of non-personal datasets like transport data, weather data, etc. Nevertheless, given the scale and pace of data gathering and sharing, distrust arising from concerns about privacy and security is often well founded and must be resolved by industry and Government if the full value of big data is to be realised. We recommend below the establishment of a Council of data Ethics to help address these issues. (Paragraph 60)

8. Businesses and governments that communicate most effectively with the public, giving the citizen greater control in their data transactions by using simple and layered privacy notices to empower the consumer to decide exactly how far they are willing to trust each data-holder they engage with, will gain a huge commercial and societal advantage. Although the length of a privacy notice will be dictated by the service or data application involved, it should be best practice to draft them as simply as possible. Furthermore, if informed, freely given consent must be the bedrock of a trusting relationship between a consumer and a data-holder, then it must always be part of that deal that consent freely given can also be freely withdrawn. (Paragraph 66)

9. As citizens' personal data is being used in ever increasing volumes and for ever changing purposes, it is vital that the Information Commissioner's Office has the powers it needs to help ensure data protection. *With a new EU data protection regulation now agreed, we welcome the Government's commitment to review current penalties for data protection breaches. The Government should nevertheless introduce as soon as possible a criminal penalty for serious data protection breaches by commencing sections 77 and 78 of the Criminal Justice and Immigration Act 2008. The Government should not regard the two-year implementation period of the recently agreed EU data protection regulation, which will provide for bigger fines, as a reason for delaying this.* (Paragraph 75)

10. *The Government should agree to the Information Commissioner's request to perform compulsory data protection audits on local government.* (Paragraph 76)

11. *The Government should set out its anonymisation strategy for big data in its upcoming Digital Strategy, including a clear funding commitment, a plan to engage industry with the work of the UK Anonymisation Network and core anonymisation priorities.* (Paragraph 77)

12. The Information Commissioner has developed a data protection kitemark, ready for use now. The use of such kitemarks, acknowledging good behaviours, would complement the greater sanctions of criminal penalties for bad behaviours that we have recommended. *The Government and Information Commissioner should work with industry to ensure that the UK's already developed kitemark is adopted as soon as possible, and initiate a campaign to raise public awareness of it.* (Paragraph 82)

13. The Data Protection Act will have to be revised to accommodate the recently agreed EU Data Protection Regulation, which will come into force with the next two years or so. We do not share the Government's view that current UK data protections can simply be left until then. Some areas in particular need to be addressed straightaway—introducing the Information Commissioner's kitemark, and introducing criminal penalties rather than relying only on the prospective greater fines envisaged by the new EU Regulation. (Paragraph 100)

14. The new EU Regulation appears to leave it open for data to be re-used, and potentially de-anonymised, if "legitimate interests" or "public interest" considerations are invoked. This is an issue that urgently needs to be addressed as big data becomes increasingly a part of our lives. There are arguments on both sides of this issue: Seeking to balance the potential benefits of processing data (some collected many years before and no longer with a clear consent trail) and people's justified privacy concerns will not be straightforward. It is unsatisfactory, however, for the matter to be left unaddressed by Government and without a clear public-policy position set out. *The Government should therefore clarify its interpretation of the EU Regulation on the re-use and de-anonymisation of personal data, and after consultation introduce changes to the Data Protection Act 1998 as soon as possible to strike a transparent and appropriate balance between the benefits of processing data and respecting people's privacy concerns.* (Paragraph 101)

15. *Given the UK's leading position in big data and the Government's stated commitment to capitalise on the potential innovation and research opportunities it promises, the Government should establish a Council of Data Ethics within the Alan Turing Institute as a means of addressing the growing legal and ethical challenges associated with balancing privacy, anonymisation, security and public benefit. Ensuring that such a Council is established, with appropriate terms of reference, offers the clarity, stability and direction which has so far been lacking from the European debate on data issues.* (Paragraph 102)

# Formal Minutes

## Wednesday 10 February 2016

Members present:

Nicola Blackwood, in the Chair

| | |
|---|---|
| Victoria Borwick | Carol Monaghan |
| Chris Green | Valerie Vaz |
| Dr Tania Mathias | Matt Warman |

Draft Report (*The big data dilemma*), proposed by the Chair, brought up and read.

*Ordered*, That the draft Report be read a second time, paragraph by paragraph.

Paragraphs 1 to 102 read and agreed to.

Annex and Summary agreed to.

*Resolved*, That the Report be the Fourth Report of the Committee to the House.

*Ordered*, That the Chair make the Report to the House.

*Ordered*, That embargoed copies of the Report be made available, in accordance with the provisions of Standing Order No. 134.

[Adjourned till Tuesday 23 February at 2.00 pm

# Witnesses

The following witnesses gave evidence. Transcripts can be viewed on the Committee's inquiry page at www.parliament.uk/science.

| Tuesday 28 October 2015 | *Question number* |
|---|---|
| **Aisling Burnand MBE**, Chief Executive Officer, Association of Medical Research Charities, **Professor John Williams**, Director, Health Informatics Unit, Royal College of Physicians, and **Dame Fiona Caldicott**, National Data Guardian for Heath and Care | Q1-34 |
| **George Windsor**, Senior Policy Researcher, Creative and Digital Economy, Nesta, **Sue Daley**, Head of Big Data, Cloud and Mobile, techUK, **Dr Paul Feldman**, Chief Executive, Jisc, and **Chirdeep Chhabra**, Senior Manager, Data Programmes, Digital Catapult | Q35-66 |
| **Richard Woolhouse**, Chief Economist and Head of Research and Strategy, British Bankers Association, **James Meekings**, Co-Founder and UK Managing Director, Funding Circle, and **Imran Gulamhuseinwala**, Partner, Ernst & Young | Q67-97 |
| **Tuesday 17 November 2015** | |
| **Chris Combemale**, Chief Executive Officer, Direct Marketing Association, **Professor Jonathan Montgomery**, Chair, Nuffield Council on Bioethics, and **Renate Samson**, Chief Executive, Big Brother Watch | Q98-144 |
| **Christopher Graham**, Information Commissioner | Q145-186 |
| **Tuesday 1 December 2015** | |
| **Hetan Shah**, Executive Director, Royal Statistical Society, **Paul Maltby**, Director of Data, Government Digital Service, and **Gavin Starks**, Chief Executive Officer, Open Data Institute | Q187-235 |
| **Mr Edward Vaizey MP**, Minister of State for Culture and the Digital Economy, Department for Culture, Media and Sport and the Department for Business, Innovation and Skills, **Rebecca Endean OBE**, Director, Research Base, Department for Business, Innovation and Skills, **Baroness Shields**, Parliamentary Under-Secretary of State for Internet Safety and Security, Department for Culture, Media and Sport, and **Peter Knight**, Deputy Director, Research and Development Directorate, Department of Health | Q236-265 |

# Published written evidence

The following written evidence was received and can be viewed on the Committee's inquiry web page at www.parliament.uk/science. BIG numbers are generated by the evidence processing system and so may not be complete.

1 Academy of Medical Sciences (BIG0079)

2 Alzheimer's Research UK (BIG0063)

3 ANS Group (BIG0042)

4 Association of Medical Research Charities (BIG0030)

5 Barclays Bank Plc (BIG0062)

6 BBA (BIG0078)

7 BCS, The Chartered Institute for IT (BIG0038)

8 Big Brother Watch (BIG0006)

9 British Academy (BIG0064)

10 British Heart Foundation (BIG0023)

11 British Library (BIG0051)

12 British Pharmacological Society (BIG0009)

13 British Standards Institution (BIG0076)

14 British Vehicle Rental and Leasing Association (BIG0049)

15 British Veterinary Association (BIG0056)

16 CompanyNet (BIG0007)

17 Consumer Data Research Centre, UCL (BIG0013)

18 Department for Business, Innovation and Skills, and the Department for Culture Media & Sport (BIG0069)

19 Department for Culture Media & Sport (BIG0089)

20 Department for Environment Food and Rural Affairs (BIG0002)

21 Digital Catapult (BIG0080)

22 Digital Science (BIG0048)

23 Direct Marketing Association (BIG0014)

24 Dr Kevin Macnish (BIG0003)

25 E-infrastructure Leadership Council (BIG0040)

26 EMC (BIG0046)

27 Ernst & Young (BIG0055)

28 ESRC Big Data Network, Phase 2 (BIG0021)

29 Experian (BIG0022)

30 Farr Institute of Health Informatics Research (BIG0074)

31 Food Ethics Council (BIG0019)

32 Food Standards Agency (BIG0037)

33 Fujitsu Services Ltd (BIG0045)

34 Funding Circle (BIG0081)
35 GARNet (BIG0020)
36 GBG (BIG0077)
37 Genetic Alliance UK (BIG0034)
38 Greater London Authority (GLA) (BIG0067)
39 ICAEW (BIG0001)
40 IET and the Royal Academy of Engineering (BIG0065)
41 Information Commissioner's Office (BIG0033)
42 Innovate UK (BIG0026)
43 Insurance Law Research Group, University of Southampton (BIG0052)
44 Institute of Cancer Research, London (BIG0071)
45 Internet Advertising Bureau UK (BIG0054)
46 Involve (BIG0010)
47 Jen Persson (BIG0085)
48 Jisc (BIG0027)
49 John Innes Centre (BIG0061)
50 Lloyd's Register Foundation (BIG0068)
51 medConfidential (BIG0088)
52 Mediatrac (BIG0005)
53 Medical Schools Council (BIG0036)
54 Microsoft UK (BIG0060)
55 Mydex CIC (BIG0032)
56 National Oceanography Centre (BIG0012)
57 National Physical Laboratory (BIG0041)
58 Nesta (BIG0047)
59 Nuffield Council on Bioethics (BIG0015)
60 Open Data Assurance by B Longo (BIG0004)
61 Open Data Institute (BIG0084)
62 Open University (BIG0070)
63 PHG Foundation (BIG0073)
64 Professor Lorna Woods (BIG0082)
65 Pure Storage (BIG0050)
66 Research Councils UK (RCUK) (BIG0057)
67 Rise.global (Leaderboarded Ltd) (BIG0090)
68 Royal College of Physicians (BIG0029)
69 Royal Society (BIG0035)
70 Royal Statistical Society (BIG0044)
71 SAP (BIG0059)

72 Science and Engineering South Consortium (BIG0075)

73 Smart Energy GB (BIG0024)

74 South London and Maudsley NHS Foundation Trust (BIG0018)

75 Susan Hall (BIG0017)

76 Swanbarton Ltd (BIG0028)

77 techUK (BIG0039, BIG0086)

78 UK Computing Research Committee (UKCRC) (BIG0011)

79 UK Data Archive (BIG0008)

80 University of Cambridge Big Data Strategic Research Initiative (BIG0016)

81 University of Edinburgh (BIG0043)

82 University of Essex (BIG0053)

83 Wellcome Trust (BIG0058)

84 Wellcome Trust Sanger Institute (BIG0066)

# List of Reports from the Committee during the current Parliament

All publications from the Committee are available on the Committee's website at www.parliament.uk/science.

The reference number of the Government's response to each Report is printed in brackets after the HC printing number.

## Session 2015-2016

| | | |
|---|---|---|
| First Report | The science budget | HC 340 (HC 729) |
| Second Report | Science in emergencies: UK lessons from Ebola | HC 469 |
| Third Report | Investigatory Powers Bill: technology issues | HC 573 |
| First Special Report | Royal Botanic Gardens, Kew: Government Response to the Committee's Seventh Report of Session 2014–15 | HC 454 |
| Second Special Report | Current and future uses of biometric data and technologies: Government Response to the Committee's Sixth Report of Session 2014–15 | HC 455 |
| Third Special Report | Advanced genetic techniques for crop improvement: regulation, risk and precaution: Government Response to the Committee's Fifth Report of Session 2014–15 | HC 519 |
| Fourth Special Report | The science budget: Government Response to the Committee's First Report of Session 2015–16 | HC 729 |