HM Government

Seizing the data opportunity

A strategy for UK data capability

October 2013

Contents

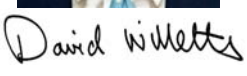
Ministerial foreword

One of the greatest opportunities and challenges facing policymakers today is the ever increasing significance of data. Data underpins our businesses and our economy, providing new insights into consumer needs and enabling new products and services to be developed. The next generation of scientific discovery and innovation will be data-driven, from modelling and simulation, to handling massive data traffic. Governments around the world must change the way they engage with citizens, the way they develop policy and deliver services, and the way they are held to account. The benefits for citizens are many, from revolutionising the information available at our fingertips with clever apps, to taking back control of our own data to understand energy use or spending habits, data can change the way we live our lives. The UK government is determined to position the UK to make the most of the data revolution.

Big data is one of our Eight Great Technologies, and for good reason – its potential impact is so significant that it could transform every business sector and every scientific discipline. We are supporting our data infrastructure, most recently with £189 million of funding for big data in last year's Autumn Statement. We have also established the E-infrastructure Leadership Council to advise the government on the computing infrastructure and skills we need to take advantage of this opportunity. The UK's open data agenda is world-leading, with over 10,000 public datasets published on data.gov.uk, and the ground-breaking Open Data Institute.

But there are still important steps we need to take to ensure that we are at the forefront of developments in data science and analytics. The challenge of meeting the demand for skilled people, from both industry and academia, is one that is globally recognised. It is a challenge that cannot be tackled by government in isolation, which is why we will work with industry and academia to come up with solutions. There is still plenty more to do on making more data accessible and usable, with the right protections in place to safeguard privacy. For example, through the recently-established Research Sector Transparency Board, and working with the research community, we are developing an open research data regime.

This is a real opportunity for the UK. We have some of the best universities in the world, and some truly innovative small businesses. We have some of the richest historic datasets of any country. We have, in short, the building blocks of a data success story. Working together with industry, academia, and citizens, we can make it a reality.



Rt Hon David Willetts MP
Minister for Universities and Science

Matthew Hancock MP
Minister for Skills and Enterprise

Information Economy Council foreword

The UK has the potential to lead one of the defining developments of the 21st century – the data revolution. This will enable the UK to capture the inherent value within increasing volumes of data to create wealth and social benefits that can transform the lives of all. This grand ambition will only be achieved through trust and engagement, within a philosophy that empowers the individual, where data is owned by and under the control of the citizen and consumer, protected appropriately and maintained in a transparent and accessible manner.

The UK has considerable capabilities in people, infrastructure, and data itself. If appropriately coordinated and supported these capabilities have the potential to secure unique insights from our assets. Additionally, the UK has great data intensive institutions such as the NHS, the BBC, our universities, and the Met Office. The UK has world-class business sectors such as aerospace, agri-tech, automotive, healthcare, life sciences, media and telecoms. These sectors can be leaders in understanding the context of data within their own domains. All of this holds enormous promise if set within a strategic framework that enables a network effect of capability, infrastructure and data to take hold.

The Information Economy Council - a collaboration across business, academia and government - will work to promote a strategic approach to data capability in the UK. The government also has an important role to play. It can encourage with incentives and appropriate procurement, it can support open standards, it can mandate public sector data releases and as a last resort it can regulate. Additionally, government can and should support the development of capabilities that underpin the emerging world of data (education, skills, infrastructure) to help generate immediate and future value for the UK as a whole.

We believe that this strategy establishes the need for a strategic approach to data capability within the UK. We must encourage organisations, both public and private, large and small, to invest and to focus resources and efforts in order to produce a winning outcome that delivers wealth creation in the UK and wider benefits for society.

Professor Sir Nigel Shadbolt
Chairman, Open Data Institute
Professor of Artificial Intelligence,
University of Southampton

Phil Dawson
Chief Executive Officer,
Skyscape Cloud Services

Executive summary

In the information economy, the ability to handle and analyse data is essential for the UK's competitive advantage and business transformation. The volume, velocity and variety of data being created and analysed globally is rising every day, and using data intelligently has the potential to transform public sector organisations, drive research and development, and enable market-changing products and services. The social and economic potential is significant, and the UK is well placed to compete in the global market for data analytics. Through this strategy, the government aims to place the UK at the forefront of this process by building our capability to exploit data for the benefit of citizens, business, and academia. This is our action plan for making the UK a data success story.

Working in partnership with business and academia, the government has developed a shared vision for the UK's data capability, with the aim of making the UK a world leader in extracting insight and value from data for the benefit of citizens and consumers, business and academia, the public and the private sectors. The Information Economy Council and the E-infrastructure Leadership Council will oversee delivery of the actions in this strategy, and continue to develop additional plans to support this vision.

Data capability

This strategy focuses on three overarching aspects to data capability. The first is human capital – a skilled workforce, and data-confident citizens. The second covers the tools and infrastructure which are available to store and analyse data. The third is data itself as an enabler – data capability is underpinned by the ability of consumers, businesses and academia to access and share data appropriately.

The UK requires a strong **skills** base, able to manage, analyse, interpret and communicate data, in order to extract insight and value for the benefit of citizens, business and academia. As a nation, we have strengths in some of the key subjects relating to data capability. We need to continue to build on these and ensure a strong pipeline of skilled people, ready and able to enter the workforce and academia, in order to meet future demand. We also need business leaders to be aware of the potential benefits of data analytics, as well as citizens who understand how data can be interpreted and used.

The growth in data is a global phenomenon; so are the efforts to research and develop new tools and techniques, and innovative ways to generate and use data. As emerging products and technologies create increasing volumes of data, and business and consumer demands become more sophisticated, it is vital that the UK's **data infrastructure** remains agile, resilient and responsive to changing trends and behaviour and that there is continued research and development (R&D) into new tools, techniques and applications of data analytics.

Data itself also underpins and enables our capability. It can be a valuable asset, and a vital part of infrastructure. In the future, it will become an integral part of the way in which economies and societies function. To best use data, we need to be able to handle, protect, guarantee and validate it, to **share data** across sectors and disciplines, and to link datasets from various sources.

Data skills

A shortage of skilled workers in the data analytics market is cited as one of the key barriers to further data analytics activity. Work is required across the skills landscape, from schools through to post-graduate study, to embed core data analytic skills and open up pathways into the discipline. The government is also taking action to improve its own capabilities in data analytics, working through the Government Statistical Service and Government Operational Researchers Service.

There is no single and universally-recognised definition of a data analytics job – these typically involve a wide range of skills, but are generally built on three core skills of data management, data analysis (including numeracy and statistics) and business insight (an awareness of commercial and wider contextual issues, and some subject specialisation). In order to best target efforts, the government will work with employers, e-skills UK, Nesta, Universities UK and the Open Data Institute to **explore the skills shortages in data analytics and set out clear areas for government and industry collaboration**.

In order to encourage people to pursue a career in analytics, the government will work with the Information Economy Council and other bodies to **bolster the image of data science as a discipline, and illustrate different career pathways in data analytics**.

While the precise skills required for a data analyst are not fully established, it is clear that computing and mathematical skills are core components and action is required across all skill levels. Recent changes to the school curriculum will provide a stronger basis for producing the data analysts of the future, and government reforms to the apprenticeship system will put employers in the driving seat to develop training programmes which focus on the skills needed for big data and business analytics.

In higher education, there is the opportunity to build on our strong relationships between universities and businesses, and the government will hold a **workshop in November 2013 bringing together representatives from universities and businesses to discuss computer science graduates and how to get the right skills to meet current and future needs.**

Given the cross-cutting benefits of data use across disciplines, Universities UK will **review how data analytics skills are taught across different disciplines and assess whether more work is required to further embed these skills across disciplines**.

Infrastructure, software, and collaborative research

Data capability is underpinned by access to the infrastructure and tools which enable data to be generated, stored, analysed and managed effectively. The exact requirements will vary according to the task at hand. For example, software tools which enable businesses and researchers to manipulate more easily large volumes of data are transforming what is possible in data analytics, particularly tools such as Hadoop and NoSQL which allow analysis of large-scale data which is not neatly structured.

Access to computing power and connectivity is also important, as well as the capability to store data safely and in a way which is accessible on demand. The UK has a strong and established data centre industry, and the government will work with the Data Centre Alliance, Intellect and UK Trade & Investment to **promote the UK data storage market overseas.**

The UK is developing its e-infrastructure, including through substantial investment in broadband, and with 24 high performance computing systems – the fourth highest share in the world – linked together by the Janet high-bandwidth network. Developments in cloud computing mean that organisations of any size can have access to shared resources and fast computing on demand. There is scope to increase awareness of the UK's e-infrastructure, particularly for SMEs, so the E-infrastructure Leadership Council will monitor a **programme of activity to drive awareness, support, and access to e-infrastructure for businesses across six key sectors, as well as a separate campaign specifically aimed at SMEs**.

The UK's excellent research base in computing, mathematical and physical sciences stands us in good stead to be a world leader in data science, pushing the boundaries of what is possible with data. The Research Councils, the Technology Strategy Board, and the European Union's Horizon 2020 programme all have roles to play here. As a first step, the Engineering and Physical Sciences Research Council is developing a **proposal for a national network of centres in big data analytics**. These centres would be just one way of encouraging collaboration and working across disciplines, which is where the greatest potential for innovative uses of data lies.

Sharing and linking data securely and appropriately

Data analytics depends on a range of enablers, including the availability of data, the ability of data from different sources to be combined, and confidence that personal or sensitive data will be protected. Some data can be shared freely, and some can be aggregated or have sensitive information removed in order to make it usable. Some data can not, and should not, be made available openly, and so must be handled appropriately and securely.

The UK government is a world leader in open data and since 2010 has released over 10,000 datasets through its data.gov.uk site. The government is also creating a National Information Infrastructure, which will contain the data held by government which is likely to have the broadest and most significant economic and social impact if made available and accessible outside of government.

Administrative data, which is the data collected by public bodies for registers and other administrative purposes, can be a valuable source of information, but by its nature can be sensitive and contain personal information. To enable the safe and secure linking of administrative data, the government has accepted the recommendations of Sir Alan Langlands' Administrative Data Taskforce report and announced the creation of four new innovative administrative data research centres to form the Administrative Data Research Network.

The UK is improving access to publicly funded research data and published research findings, to enable other researchers to make new and innovative use of that data, and promote collaboration. The government has established the Research Sector Transparency Board to advise on the strategic approach for open research data, and has also implemented the recommendations of the Finch report on Open Access with an initial £10 million of funding available, being followed up by the Research Councils UK's additional allocation of £37 million over the period 2013/14 to 14/15 to more than 100 universities. The research community, led by the Royal Society, will convene an **Open Science Data Forum next year to develop proposals to support the access to, and use of, research data.**

The government is also **supporting non-commercial research through enabling text and data mining** which will allow researchers to undertake large scale analysis of journal articles and other works.

Personal data needs to be carefully protected, but citizens can also be empowered by access to data about them and their behaviours. The midata initiative is enabling consumers to access data that companies such as energy providers or banks hold about them, which they can use for their own benefit. The government will **convene a working group on widening the midata programme**.

Data can most effectively and safely be shared and linked through good data standards, and a strong data protection framework. The research sector and the UK's standards bodies are working on developing data standards across disciplines, as well as best practice for collecting and cataloguing data. The Information Commissioner's Office produces a range of guidance on best practice in data sharing, and the UK government is working in the European Commission to achieve a data protection framework which protects the civil liberties of individuals while allowing for economic growth and innovation. The government will look at **options to promote guidance and advice on the rights and responsibilities of data users.**

The data opportunity

We live in a world where the volume, velocity and variety of data being created and analysed each day is ever increasing. It is estimated that 90% of the world's data was created in the last two years[1], and each day 2.5 billion gigabytes of data is created – enough to fill over 27,000 iPads per minute[2].

Data has been likened to the "new oil" of the 21st century[3], but unlike oil, we are not going to run out of it. On the contrary, we will continue to amass more and more data. As set out in the Information Economy Strategy[4], business sectors across the economy are being transformed by data, analytics, and modelling. New and emerging technologies will fuel the growth of data: as access to computing and the internet becomes ever more mobile, data will be transmitted and analysed continuously; and the development of the Internet of Things could mean that by 2020 sensor data will be created by as many as 50 billion connected devices across the globe[5]. As Smart Cities and the Internet of Things become a reality, real-time data will be created and require instant analysis as it is continually fed back into the network of connected objects and devices, providing ground-breaking services for citizens and businesses.

The challenge arises not only from the volume of data, but in its variety, and the speed at which it is created and processed. Data can be streamed continuously, such as from seismic or medical sensors, for real-time analysis and interpretation. Data is increasingly generated through experimental modelling and simulation. As well as the familiar, structured forms of datasets, in neat rows and machine readable formats, there is a significant rise in unstructured datasets, such as blog posts, video and audio content, and social media feeds. The vast majority of new data being generated is predicted to be unstructured[6].

Benefits of data

From increased transparency and accountability through open data, to new scientific discoveries, and market-changing products and services which can be developed using modelling and simulation, big data analytics and data-driven science, the opportunities – and challenges – are significant.

The Centre for Economics and Business Research estimates that the big data marketplace could create 58,000 new jobs in the UK between 2012 and 2017[7], whilst a recent report from Deloitte estimates that the direct value of public sector

[1] IBM, What is big data?
[2] Calculation based on IBM figures and 64GB iPad
[3] For example, Marketing Week, Why we believe that data is the new oil
[4] HM Government, Information Economy Strategy, 2013
[5] Ericsson, More than 50 billion connected devices, 2011
[6] EMC, The digital universe in 2020, 2012
[7] CEBR, Data equity: Unlocking the value of big data, 2012

information alone to the UK economy is around £1.8 billion per annum, with wider social and economic benefits bringing this up to around £6.8 billion[8].

Research by Nesta shows that UK data-driven firms are 40% more likely to report launching products and services ahead of their competitors than those who aren't data-driven firms[9]. In a forthcoming report, Nesta will also show a strong connection between online data use and a firm's productivity and profitability.

Data can also empower consumers to get deals based on their real habits and use of goods and services. For example, Billmonitor estimates that 76% of UK mobile users, who have a monthly contract, spend on average £194 more than they need to every year, equating to almost £5 billion a year in unnecessary costs[10]. With greater access and understanding of their usage data, consumers could choose the best deal for their needs.

The wider social and economic benefits are manifold: such as better health outcomes for NHS patients as a result of analysis of clinical trials data; consumers receiving personalised marketing of goods as businesses analyse spending habits; and greater transparency and accountability of government to citizens with the release of more open data. There are already many examples of the transformative impact that data is having on the way we live and work.

Formula One in the NHS

McLaren Electronic Systems are using their real-time data system expertise, developed as part of their Formula One work, to help Birmingham Children's Hospital improve the monitoring of children in intensive care.

The project, which started in 2010, enables the hospital to benefit from the advanced telemetry commonly used in Formula One to assess a car's performance whilst on the track. In 2011, the real-time data system, which is used by every team in Formula One, was installed onto the hospital's computer network. This allows physiological data to be streamed live from bedside monitors in intensive care and from the specialist child transport ambulance, which can then be processed in real-time to understand the shifting patterns that can tell medical staff when a patient's condition is changing.

Image courtesy of McLaren

[8] Based on 2011 prices. Deloitte, Market assessment of public sector information, commissioned by BIS, 2013
[9] Nesta, Rise of the Datavores, 2012
[10] Billmonitor.com, National mobile report, 2012

RailSimulator.com

RailSimulator.com is a Kent-based videogames developer and exporter. Established in 2009, the company has grown rapidly from a team of six and now employs more than 50 people across three UK studios, including a full-time data analyst. The core product is the successful Train Simulator series, a PC train driving simulation which launches annually with paid-for downloadable add-ons released each week.

Image courtesy of RailSimulator.com

Through data analytics, RailSimulator.com can continue to improve not only the service provided to customers, but also enhance business performance. Using data analytics systems, the company can identify the times each day when users in different territories engage with sites, and target them with relevant, localised messaging. This data also helps the company establish the relative significance of feedback expressed by their user community. Customers can post their feedback on brand new products, so the extent of any issues can be reviewed immediately by analysing the number and frequency of their comments, enabling them to tailor products in line with user expectations.

The company recently began gathering telemetry data showing how players spend their time while using the software. Assessing the popularity of different game features allows RailSimulator.com to craft the user experience to match user behaviour, increasing customer satisfaction and enhancing their service in the process.

Case for action

The increasing importance of data is a global phenomenon. Countries around the world are acting now to position themselves to take advantage of the data opportunity. From Japan's Growth Strategy which allocates nearly £90 million for big data R&D, to India's new Fourth Paradigm Institute focusing on large scale modelling and computation, to France's new industrial strategy which includes big data as a key element, global activity here is significant. Furthermore, government leaders at the October meeting of the European Council agreed that Europe should boost digital, data-driven innovation across all sectors of the economy as part of its growth strategy.

The UK is well placed to compete. In terms of data analytics, we have many areas of strength on which to build. One of these strengths is undoubtedly the richness of the UK's historical datasets, where we have some of the world's best and most complete

data in several areas. In healthcare, for example, the National Health Service has outstanding clinical data from a large and diverse population, and trusted procedures to maintain patient confidentiality, which can provide opportunities for research to advance patient care. In environmental studies, the UK has a wealth of data including meteorological records dating back as far as 1659 in some areas, and weather reports in captains' log books, all of which provide a valuable source of insight and evidence.

Data use covers a wide spectrum of activities, including big data analytics, and some organisations are already performing sophisticated analytics with data to commercial benefit. There are also significant opportunities for government. Better use of existing data and taking advantage of the potential of big data could make government services more efficient, more tailored or even transform them.

However, since this is a global market, we must continue to develop and plan for the future if we are to maintain the UK's national competitiveness. The UK's ability to access and provide insights into data is crucial for success across all economic sectors.

So we have the opportunity to make the best use of the growth of data: knowing how to sort through the volume, bring different datasets together, and gain new knowledge and insight. The UK needs to act now to build the capability within the UK to be at the forefront of extracting knowledge and value from data for the benefit of citizens, business, academia and government. The onus cannot be all on government – we need to work in partnership with business and with academia, to realise all the benefits. This is our action plan for making the UK a data success story.

Vision

The UK is a world leader in dealing with the volume, velocity and variety of data created and analysed every day – extracting insight and value for the benefit of citizens and consumers, business and academia, the public and the private sector – through:

- A strong skills base, able to manage, analyse, interpret and communicate data
- A strategic plan for our data infrastructure across the country
- World-leading research and development, pushing frontiers and driving innovation in data science and analytics
- Ensuring that data can be accessed and shared securely, as appropriate

This is underpinned by a strong policy framework which protects and empowers citizens and supports innovation and growth – with government an exemplar of best practice and with science as a key driver of this capability.

This is our vision, developed and shared with industry and academia, for the UK's data capability over the next decade. At the same time, we need to ensure that in all areas, we balance the drive for innovation and growth with the rights and privacy of the individual.

Where next?

The next chapter explores what data capability means and the key components of it. Then, the strategy sets out, for each component, a clear set of actions to help us achieve this. These actions are highlighted and summarised in a table at Annex A.

Implementation of this strategy will be overseen by the Information Economy Council and the E-infrastructure Leadership Council, who will work in partnership across business and academia to ensure delivery of these actions, but also consider what more can be done to support the strategy's overarching vision. This strategy is only the start. Regular updates on progress will be given at the Information Economy Council meetings and updates on delivery and new actions will be available on the www.gov.uk website.

What is data capability?

This strategy considers three overarching aspects to data capability. The first is human capital – a skilled workforce, and data-confident citizens. The second covers the tools and infrastructure which are available to store and analyse data. The third is data itself as an enabler – data capability is underpinned by the ability of consumers, businesses and academia to access and share data appropriately.

Human Capital – a skilled workforce and data-confident citizens

The UK requires a strong skills base, able to manage, analyse, interpret and communicate data, in order to extract insight and value for the benefit of citizens, business and academia. As a nation, we have strengths in some of the key subjects relating to data capability, but we need to continue to build on these and ensure a strong pipeline of skilled people, ready and able to enter the workforce and academia, in order to meet future demand. We also need business leaders to be aware of the potential benefits of data analytics, as well as citizens who can interpret data.

While technology will continue to develop apace, and big investments have been made in infrastructure by the government over the last few years, it is the availability of skilled people that will determine the future capability of the UK to be a world leader in data analytics.

Data analytics entails a wide range of skills. Typical data-related analytical roles are built on three core areas: data management; data analysis; and business and policy insight. The UK needs to develop an all-educational-level approach from equipping school children with basic mathematics and analytics skills, through ensuring the wider workforce remain abreast of developments in data use, to funding doctoral students working at the cutting-edge of data analytics.

Data scientists typically require knowledge in both analytic techniques and the sector in which they work. Some will have a university education in mathematics, informatics or computer science and then gain sector knowledge on the job, but many more will have undertaken first degrees in other science, technology, engineering and mathematics subjects. Some will have a background in social and behavioural sciences where quantitative data analysis is used to understand a wide range of phenomena. Others will have studied web and internet science with its requirement to understand the structure and dynamics of the largest data repository on the planet. Many will need to gain advanced analytic skills at a higher level, either by undertaking a relevant masters degree, through on the job training, or during doctoral study, such as at a Centre for Doctoral Training (CDT).

The potential of data analytics will not be met unless business managers not only understand the technology, but also how to integrate this effectively into their business model, so it is also important that business leaders are aware of data analytics and its capacity to transform businesses.

As well as ensuring that we have a pipeline of skilled workers trained to work with data, all UK citizens should have an understanding of data – how it is created and stored – and confidence in its use and their rights.

Infrastructure, software and collaborative R&D

The growth in data is a global phenomenon, as are the efforts to research and develop new tools and techniques, and innovative ways to use data. As emerging products and technologies such as modelling and simulation create increasing volumes of data, and business and consumer demands become more sophisticated, it is vital that the UK's data infrastructure remains agile, resilient and responsive to changing trends and behaviour and that there is continued R&D into new tools, techniques and application of data analytics.

The UK needs a data infrastructure whereby companies can access their data swiftly from secure and resilient storage systems, which can adapt to new developments, such as 5G spectrum and the Internet of Things, and that provide sufficient and sustainable high-end computing capacity.

Computer power and storage is vital to ensure that the UK has the tools and capacity to interrogate and store the growing amounts of data being created. With increasing energy prices and global action to reduce carbon emissions, energy efficient computing – using smart algorithms to get the same result with less effort – is also a key technology.

To form an effective e-infrastructure ecosystem, individual components such as high performance computing systems (HPCs) and distributed systems such as cloud computing need to be linked together via a high-performance, very high bandwidth network designed to meet requirements of those using e-infrastructure in leading-edge research and development. Good internet connectivity creates opportunities for organisations large and small to transmit and collect data, supporting economic growth across the country.

Being able to handle large volumes of unstructured data will be key to UK competitiveness, since the vast majority of new data being generated is predicted to be unstructured[11]. Additionally, data analysis will be increasingly dynamic, with datasets updating constantly or being streamed real-time from experimental sources. This could include video footage, embedded medical devices such as sensors measuring heart rate, or data from entertainment and social media.

Continuing R&D into new tools, techniques and infrastructure will be important to keep the UK at the forefront of data analytics. There are opportunities involving massive volumes of data which require sustained development of, and investment in, new tools and techniques for data analytics, as the case studies below demonstrate.

[11] EMC, The digital universe in 2020, 2012

The Square Kilometre Array

Image courtesy of SKA

The Square Kilometre Array (SKA), which will be the world's largest and most sensitive radio telescope, is a global project in which the UK is looking to play a major role. The project design office of the SKA organisation is located at Jodrell Bank Observatory near Manchester and a number of UK research organisations and companies are involved in the project.

When the SKA is operational, the data it produces will be on a scale far greater than our current capabilities can handle. The SKA could produce 100 times the global internet traffic and the SKA supercomputer will also need to perform 10^{18} operations per second – equivalent to the number of stars in three million Milky Way galaxies – in order to process all the data that SKA will produce. This computing power, developed by pushing the boundaries of current capability, will enable important advances in other fields.

100,000 genome project

The government is also supporting a cutting-edge project to map 100,000 whole genomes over five years, with £100 million of funding and a pioneering partnership between Cancer Research UK and government-owned Genomics England, as part of our ambition for the UK to lead the world in unlocking the power of DNA data.

Image courtesy of Genome Research Ltd

When the first human genome was fully sequenced in 2000, the project cost upwards of £500 million – whereas with developments in health and information technology, we may soon be able to sequence a human genome for less than £1,000, which will revolutionise the data available to researchers and the care available to patients.

Data as an enabler

Data itself also underpins and enables UK capability. It can be a valuable asset, and a vital part of infrastructure. In the future, it will become an integral part of the way in which economies and societies function. To best use data, we need to be able to handle it, to share data across sectors and disciplines, and to link datasets from various sources.

The availability and accessibility of data is a key part of capability. There is not simply one type of data – as a resource, it can come from vastly different sources, have entirely different attributes, and be manipulated in different ways.

Stephan Shakespeare's review of public sector information identified the importance of 'core reference data' - the most important data held by each government department and other publicly funded bodies, as identified by an external body[12]. This could be, for example, locations of schools or hospitals, as well as data of critical national importance, such as responding to national emergencies.

The public sector also holds data which, while not necessarily 'core reference data', can provide useful information to citizens and opportunities for researchers and organisations to innovate. Other organisations can also provide open data; for example, publishing research data in a reusable form, which supports research findings, and makes the data more widely usable.

Not all data can be published or made available for free. Some is commercially confidential, and it is important that companies and researchers who invest time and effort into gathering data have the opportunity to derive economic benefit from their work. Some data is sensitive personal information, which cannot be shared for reasons of privacy and security. However, where it is appropriate to do so and the right protections have been taken, such as removing personal identifiers or aggregating data, sharing or linking data can bring both social and economic benefits.

For example, Mastodon C, a start-up big data company, worked with Open Healthcare UK and Dr Ben Goldacre to look at the regional prescribing patterns of statins by NHS doctors in England. Their analysis of this aggregated prescription data highlighted a variation in prescription rates and the team estimated that up to £200m could have been saved if the research had been conducted a year ago. Transparent sharing of information can help clinicians understand whether they are over or under prescribing and allow NHS funding to be invested in other areas.

One of the most important factors in enabling data to be shared and linked is the wide adoption and use of good data standards. Ensuring that data is curated and archived correctly, with appropriate descriptions and metadata about the source and method of collection, will allow data to be reused most effectively. It will also ensure that data can be reused with confidence in the analysis and that conclusions can be drawn from data which someone else has collected.

[12] Stephan Shakespeare, Independent review of public sector information, 2012

Another factor is a clear and proportionate legal framework for sharing and handling data. This ensures that those who handle data know their duties and obligations to protect it, while individuals understand how their data may be used and what rights they have. In the UK, the Data Protection Act 1998 is the main guiding legislation, and itself is based on the EU Data Protection Directive 1995.

Skills

This chapter looks at three key areas: ensuring that the UK education system produces people with the skills necessary to work with data; improving the profile of careers in data analytics; then finally considering how the government ensures that it continues to upskill and use data as a key part of policy development.

Securing a pipeline of talent

A shortage of skilled workers in the overall data analytics market is cited as one of the key barriers to further data analytics activity for businesses both globally[13] and in the UK[14], with one global survey finding that 46% of respondents quoted staff shortages as the most common barrier to implementing data analytics[15]. On top of this, e-skills UK predicts an increase of between 13% and 23% per annum in demand for big data staff between now and 2017[16].

Demand for big data specialists, including data scientists, will continue to increase over the next five years, so concerted effort is required to ensure that the flow of school leavers and graduates can meet this demand. Work is required across the skills landscape, from schools through to post-graduate study, in order to embed core data analytic skills and open up pathways into the discipline.

As stated in the previous chapter, there is no single, internationally recognised definition of data analytics and data analytics jobs involve a wide range of activities and skills. Typical data-related analytical roles are usually built on three core skills: data management (data storage and linking); data analysis (including numeracy, statistics and coding knowledge); and business and policy insight (subject specialism, context awareness and entrepreneurial spirit).

As a follow up to their 'The Rise of the Datavores'[17] report last year, Nesta will publish 'Skills of the Datavores' in 2014, which will delve down into the data science skills that businesses need and also the attributes of a multi-disciplined data team. As part of their upcoming report, Nesta will also look at which universities already offer specific data analysis and data science undergraduate and postgraduate courses, as well as which universities are already working closely with industry on curriculum design. In order that government and industry can best target their efforts, it is essential to understand fully the skills that industry needs, in particular where the gaps are, and match these against the areas of current UK strengths and weaknesses.

[13] McKinsey Global Institute, Big data: The next frontier for innovation, competition and productivity, 2011
[14] Deloitte, Market assessment of public sector information, commissioned by BIS, 2013 and e-skills UK/SAS, Big data analytics: An assessment of demand for labour and skills, 2013-2017, 2013
[15] TDWI, Best practices report: Big data analytics, 2011
[16] e-skills UK/SAS, Big data analytics: An assessment of demand for labour and skills, 2013-2017, 2013
[17] Nesta, Rise of the Datavores, 2012

The e-skills UK and SAS report 'Big Data Analytics' contains detailed analysis of the skillsets of the key big data specialist roles, including data scientists. The report also further references the critical technical skillsets that are required for each role e-skills UK and SAS will release the next phase of their work into skills issues for big data specialists in early November this year. Based on primary evidence from UK employers, this report will provide estimates of current employment in big data specialist roles and amongst big data users, and then forecast projected employment levels of both through to 2017.

Action: Building on the findings of these reports, government will work with employers, e-skills UK, Nesta, Universities UK and the Open Data Institute to explore the skills shortages in data analytics and set out clear areas for government and industry collaboration.

Whilst there is not currently a complete understanding and definition of the requisite skillsets of a data scientist or big data specialist, it is clear that computing and mathematical training are core component skills of a range of data scientist jobs. The rest of this section looks at these skills throughout the education system to ensure diverse routes into the data analytics profession.

Schools

Government reforms to primary and secondary education will ensure that pupils leave school with a good level of ability in mathematics, computing and science, which are the essential skills needed for further study and careers in data analytics.

From September 2014, the renamed computing national curriculum will require that students aged between 5 and 16 are given the skills they need to build apps and write computer programs. The curriculum will cover theoretical concepts in computing and practical problems, software and hardware systems. Students will be given a thorough understanding of logic and set theory, and they will need to study algorithms, programming languages and the architecture of the internet.

As well as changes to the curriculum, the government is also introducing tough new requirements that computer science teachers should be able to write code. To encourage exceptional graduates to become teachers, the government has launched new bursaries and scholarships of up to £25,000 for computer science graduates who become teachers, as well as £2 million of funding to build a network of 400 'master teachers', who will help thousands of other computing teachers acquire skills that they in turn can pass on to their students.

These changes address important problems in the teaching of programming at schools, as identified in the Livingstone and Hope independent review of skills for the video games and visual effects industries, 'Next Gen', written by Nesta in 2011[18].

[18] Nesta, Next Gen., 2011

The mathematics curriculum is also being reformed to ensure that students are fluent in the fundamentals of analytic techniques. The new curriculum will emphasise that mathematics is a creative and highly inter-connected discipline which is essential to everyday life, critical to science, technology and engineering, and necessary for the financial literacy required in most forms of employment.

There is a lack of suitable qualifications for the 40% of students who achieve at least a grade C at GCSE but for whom A level or AS level mathematics is unsuitable. To counteract this, the government has announced the development of Core Maths qualifications for these students. Core Maths qualifications will include topics such as statistics, probability, advanced calculation and modelling, and will develop students' mathematical thinking and problem-solving skills. The qualifications will also count as the maths element of the new Technical Baccalaureate, to be introduced from September 2014 to recognise high performance by students in vocational education.

The new qualifications will be available for first teaching in 2015 and the government will spend £20 million from 2014 to 2016 on a programme to support schools and colleges to develop teaching for the new courses.

Make Things Do Stuff

Image courtesy of Nesta

Make Things Do Stuff is a campaign and website designed to mobilise the next generation of digital makers, bringing together big business and start-ups, formal education and after school clubs, mentors and young people to create new opportunities to improve our digital skills by delivering practical tools, clubs, and online digital making opportunities.

Launched by the Chancellor in May 2013 and led by Mozilla, Nesta and Nominet Trust, the Make Things Do Stuff campaign is supported by Facebook, O2 Telefonica, Virgin Media, Caffè Nero, BlackBerry, Samsung, Microsoft, BFI and grassroot organisations like Code Club, Apps for Good, Young Rewired State and Freeformers. It aims to inspire young people to become creators, not just consumers of technology, helping them make their own games, apps, websites and inventions using programming and other technology-led creative skills.

Since the campaign launched, over 100,000 new digital making opportunities have been taken up by young people across the UK, including over 1000 code clubs reaching over 15,000 primary school kids with Code Club.

Putting employers at the heart of skills development

The government is clear that employers should be closely involved in designing the right training to ensure that their future skills needs are met.

Following 'The Richard Review of Apprenticeships' published in 2012[19], the government has recently published 'The Future of Apprenticeships in England: Implementation Plan'[20], which sets out policy, process and timescales for reforming apprenticeships in England.

The reforms will put employers in the driving seat by basing apprenticeships on standards designed by employers to meet their needs. The standards will be easy to understand and will describe the skills and knowledge that an individual needs to master in order to be fully competent in their occupation. We will also improve the quality of apprenticeships through introducing higher expectations on English and maths, grading and an increased focus on assessment at the end of an apprenticeship.

The first new apprenticeship standards will be developed through Trailblazer projects, led by employers and professional bodies. One of the first eight Trailblazer projects announced alongside the Implementation Plan will focus on digital industries and it will develop the apprenticeship standards for software development and networking. The project will be led by organisations including Accenture, BT, BCS, the Chartered Institute for IT, CapGemini, Cisco, IBM, Microsoft and the Test Factory.

As well as apprenticeships, the government has introduced the new Employer Ownership of Skills pilot, which encourages companies to come forward with robust bids to co-invest in a range of training initiatives, and allow business to take end to end responsibility for skills development and design the vocational training programmes required.

Industry leaders have come together with a wide range of small, high growth digital companies, supported by e-skills UK, to develop a bid focused on:

- Investing in the strategic technology skills for global leadership in massively growing new markets
- Enabling faster growth of geographic digital clusters and tech entrepreneurs
- Creating new routes into the sector that deliver quicker results for business and accelerate job creation

The bid includes a particular emphasis on big data and business analytics, as well as new post-graduate training and employer-designed apprenticeships with greater flexibility, based on global industry qualifications.

[19] The Richard review of apprenticeships, 2012
[20] BIS, Future of apprenticeships in England, 2013

getstats

getstats launched in 2010 to improve how we all handle numbers – the practical numbers of daily life, business and policy. Statistics turn data into useful information. They give numbers meaning and help to decode complexity. The Royal Statistical Society believes that a society where more of us are comfortable handling data would be more prosperous and better informed.

getstats has reached journalists, politicians, policymakers, school children and teachers, and the wider community, with support from funders including the Nuffield Foundation and the Department for Business, Innovation and Skills. Over 500 journalists and student journalists have been trained in statistical literacy, and we have been helping journalism school curricula to respond to the challenges of reporting in a data-heavy world.

Parliamentarians and policy-makers have benefited from regular events in parliament exploring the statistical underpinnings of different policy agendas such as crime, health and education. Improving quantitative education, at school, post-16 and higher education level, has been a key focus for getstats.

getstats continues to campaign on the importance of statistical literacy and quantitative skills for all, with more developments planned in 2014 and beyond.

Higher Education

Typically there are two types of person working in a data analytics job: someone with a deep analytical education background (usually computing, mathematics or statistics), who acquires specialist knowledge and policy skills through courses and on the job training; or someone with a sector specialism who then acquires data analytic skills, whether through a conversion course or on the job training.

Data on the number of new graduates shows that enrolments on full-time first degrees in analytical subjects (mathematics, physics and engineering) have been increasing between 2002/03 and 2011/12, although enrolments on full-time first degree computer science courses declined during this period. The UK also has the highest proportion of computing, mathematics and statistics doctoral graduates among comparator countries[21], which puts the UK in a strong position on which to build.

With elements of coding, data analysis and programming as part of most computing degrees, computing graduates will be a key source of skilled people moving into data analytics jobs.

[21] BIS analysis of OECD data

Action: The government will hold a workshop in November 2013 bringing together representatives from universities, businesses and other relevant bodies to discuss computer science graduates and how to get the right skills to meet current and future needs.

Action: Universities UK will review how data analytics skills are taught across different disciplines and assess whether more work is required to further embed these skills across disciplines.

It is also important that industry and universities continue to work together to ensure that graduates leave university with the skills that industry needs. There are already good examples of where this is happening. For example, SAS has launched the SAS Student Academies, which give educational institutions the ability to train students in real-life big data skills. There are now sixteen academies in universities across the UK, which are designed not only to equip students with big data analytics skills, but also to help UK businesses take advantage of the innovation and efficiency big data can deliver.

At post-graduate level, the Engineering and Physical Sciences Research Council (EPSRC) currently supports around 80 Centres for Doctoral Training (CDTs) with around 4,000 students, in line with the future skills needs of the UK as described by UK-based businesses. CDTs offer a 4 year programme which includes taught elements, including technical and transferrable skills, alongside doctoral research. The benefits of this approach include producing students with both breadth and depth of knowledge, and enabling new methods of collaborative and multi-disciplinary research. Thus, CDT students develop both advanced scientific and analytical abilities alongside a thorough grounding in the relationship between research and application.

EPSRC is currently refreshing the CDT landscape, with the intention of investing £350 million and an expectation of leveraging a further £250 million from universities and industry. In setting priorities, the EPSRC is very aware of the skills needs identified by the E-infrastructure Leadership Council and the Information Economy Council. Around two thirds of the applications currently being considered for funding will provide skills that are relevant to managing big data and the results will be announced later in 2013.

Q-Step

The Nuffield Foundation, the Economic and Social Research Council (ESRC) and the Higher Education Funding Council for England (HEFCE) have launched Q-Step, a £19.5 million initiative designed to promote a step-change in quantitative social science training.

Over a five year period from 2013, 15 universities across the UK are delivering specialist undergraduate programmes, including new courses, work placements and pathways to postgraduate study. Expertise and resources will be shared across the higher education sector through an accompanying support programme, which will also forge links with schools and employers.

Q-Step was developed in response to the shortage of quantitatively-skilled social science graduates, which had led to employers across all sectors being unable to recruit people with the skills to apply quantitative methods when evaluating evidence and analysing data.

Data as a career

As well as ensuring that we equip school leavers and graduates with the key skills, we also need to ensure that data analytics is considered an exciting and rewarding career to pursue – amongst schoolchildren and graduates, but also parents and the wider media.

Career pathways and progression routes

In 2011, the Harvard Business Review published an article referring to the data scientist as the sexiest job of the 21st century[22], yet as described earlier in this chapter, this is an area which is currently experiencing skills shortages.

One of the potential underlying issues here is information asymmetries, as students do not realise or understand the potential career paths and benefits from a career in data analytics. In some cases, this is due to the fact that career routes are less well established than for other professions, such as law or medicine, where information is more readily available and the skillsets of these professions do not change as rapidly as in data analytics. For example, recognition in industry is often financial and in academia it is through publication. As yet, data scientists, whose role is to support and enable research without directly driving it, have no clearly defined recognition pathway.

[22] Harvard Business Review, Data scientist: the sexiest job of the 21st century, 2012

STEMNET

The government funds STEMNET to run the STEM Ambassador programme, which is a nationwide network of over 26,000 volunteers from science, engineering and technical companies or academia, who work with schools across the UK, supporting teachers and inspiring and enthusing young people about STEM (science, technology, engineering and maths) subjects and the range of careers that STEM qualifications offer. STEMNET's network of sub-regional contractors matches schools with STEM Ambassadors, who can respond to a range of requests and opportunities, from providing stimulating scientific activities to engaging careers talks.

STEM Ambassadors come from a whole range of sectors and STEMNET also partners with industry to promote various STEM careers. For example, STEMNET has partnered with BT to set up the IT Ambassadors scheme and Ukie to create a Video Games Ambassadors programme.

Image courtesy of STEMNET

With a wide variety of data analytics job and also a range of different ways to develop a career, possible entrants into the field need to understand the career paths into a job in data analytics, as well as the various career progression routes. Industry has a role to play, by highlighting the opportunities and career pathways within the discipline. Many employers are already working together on this, for example in developing apprenticeships, and creating content about data careers for e-skills UK's www.bigambition.co.uk careers website.

Action: The government will work with the Information Economy Council, e-skills UK and Intellect on a communications plan to bolster the image of the discipline by spring 2014. As part of this, government will work with the Information Economy Council, Research Councils, and the relevant professional bodies, including BCS and IET, to collate career profiles of people working in data analytics and the different career pathways.

Data Centre Alliance bootcamp

In a bid to solve the job shortages in the data centre industry, the Data Centre Alliance (DCA) ran a bootcamp in August 2013 to highlight the opportunities in the sector to unemployed graduates to work in the industry.

In total 21 people attended the 10-day intensive course, the majority of which were out-of-work graduates, as well as forces leavers. At the end of the bootcamp, all of the participants were successful in gaining an interview at a data centre.

Image courtesy of the Data Centre Alliance

The government's analytical capabilities

With increasing amounts of data, the government must also make sure it is well positioned to store and analyse the data it owns, and maximise its use during policy development.

Evidence collected by Deloitte highlighted a lack of skills and familiarity of civil servants to work effectively with data [23]. In particular, the concerns were about reluctance to use data not produced within government, as well as a lack of the necessary skills to combine and manipulate big data and linked data. Lack of skills to prepare and publish data was also one of the most commonly cited barriers by Local Authority respondents to the Local Government Association's survey (27% of respondents)[24].

The emphasis on developing professional capability set out in the Civil Service Reform Plan will help. The analytical professions within government, particularly the Government Statistical Service (GSS) and the Government Operational Researchers Service (GORS), will lead on developing the internal capability to use these new

[23] Deloitte, Market assessment of public sector information, commissioned by BIS, 2013
[24] Deloitte, Market assessment of public sector information, commissioned by BIS, 2013

techniques and, later this year, the GSS will launch their data strategy, which will look at key skills, IT and legislative needs, as well as setting out clear next steps and actions.

As well as the GSS and GORS, the Government Economic Service and the Government Social Research Service are also ensuring that they continue to upskill their members. At their joint conference in September 2013, the GES and GSRS held a session looking at how the government can make better use of big data.

The Government Chief Scientific Adviser, Sir Mark Walport, will also be working with departmental Chief Scientific Advisers and the network of scientists and engineers across government to increase the use made of data and analytics in analysing and tackling big policy challenges.

Specifically, as announced at the Spending Review, Sir Mark and the Government Office for Science will work with the GSS to identify the datasets which are important at times of national emergencies. This data helps organise the response to disasters as well as help plan mitigations and defences.

Data in policy making and service delivery

However, it is not just the government analytical professions that need to be confident in handling and analysing data. In a recent survey of public sector staff, 72% of respondents agreed that it is becoming increasingly important for all civil servants to know how to access, share and use data[25]. Policy makers need to be confident in using data, and understand how and when to work with the professional analytical specialists.

The Cabinet Office's Government Innovation Group is leading new priority work on open policy making. This is considering new policy tools and techniques, including the effective use of data in policy and service delivery. The Cabinet Office is working with the Government Digital Service (GDS) to develop and make available context specific tools for policymakers to help them in their roles.

Furthermore, the GDS Service Design Manual, which contains guidance to improve public services and ensure they are digital by default, highlights the importance of public service managers harnessing data analytics to provide improved services. Data about who is accessing government services and how they are using them can inform better design and delivery of services.

[25] Policy Exchange, Smaller, Better, Faster, Stronger, 2013

What Works Networks

The What Works Network is a new initiative to improve the use of high quality evidence when the government makes decisions about public services. It is also a world first - the first time any government has prioritised evidence to inform policy and practice through a national approach.

The What Works Network is made up of six evidence centres covering health and social care, education attainment, ageing better, local growth, crime reduction and effective early intervention. Their role is to summarise and share research about what works, and crucially what doesn't work, with central government and local decision-makers including commissioners, head teachers, and police and crime commissioners. The centres will not only disseminate research findings, but also support people to critically engage with research and apply the findings to their own contexts.

Together, the What Works centres cover areas with public spending of more than £200 billion. Ensuring rigorous, high quality, independently-assessed research shapes decision-making at every level will mean services that deliver better outcomes for citizens and better value for money for taxpayers.

The six centres are independent of government and are supported by a national adviser who leads the cross-government effort to improve the integration and generation of evidence in policy.

Infrastructure, software and collaborative R&D

As well as human capital, the UK's data capability is underpinned by access to the infrastructure and tools which enable data to be stored, analysed, and managed effectively. This chapter looks at the key software and infrastructure required for data analytics, and includes actions to ensure these tools are accessible by all. Research and development into new infrastructure and software is vital to enable the UK to be at the forefront of extracting knowledge and value from data, and this can be supported through bringing together academia, businesses and other key stakeholders, to push the boundaries of what is possible with data analytics.

Software

Data analytics software and services provide organisations of all sizes with the opportunity to make more of the data they own and use. Software tools can allow the most effective use of computing power and e-infrastructure. Recent years have seen the development of new software tools and database systems, such as Hadoop, HBase and NoSQL, which allow the analysis of not only structured datasets, but also unstructured datasets, such as blog posts, social media, and video and audio content. In addition, commercial tools have been developed which make it possible to mine structured data sources alongside unstructured ones, merging datasets that were previously incompatible. The UK has a comparative strength in developing software, with some of the most innovative products and uses coming from small businesses.

Infrastructure

Data storage and cloud computing

It is crucial that the UK maintains a sufficient level of data storage capacity to ensure that organisations can store and access increasing amounts of data. There are good reasons to expect that this need will be met. The UK currently has the third largest share of the global data centre market[26], and during 2011-12, it is estimated that UK-based companies invested £2.1 billion in data centres – the second highest spending of any country[27]. It is not just the size of our data storage market which makes the UK a global competitor. In 2012, the UK was named the second lowest risk destination in the world for data centre locations[28] and the UK is also positioned third in the overall Data Centre Development Index, which ranks markets according to their state of development[29]. This ranking is predominantly due to the UK's high internet bandwidth capacity and a high rating for the ease of doing business.

[26] New Statesman, Going digital: Where is our data?, 2013
[27] DataCenter Dynamics, The 2011 census, 2011
[28] Cushman Wakefield, Data centre risk index 2012, 2012
[29] Techworld article, UK data centre market booms, 2011

Achieving an increase in energy efficiency is important as energy costs are estimated to comprise between 25% and 60% of data centre operating costs[30]. Through increases in energy efficiency, data centres can achieve greater effectiveness as more work is done per unit of energy used. Without continued advances in energy efficiency computing, energy costs could limit the increase in big data analytics.

Another response to the growth of data is the increasing levels of virtual storage provision offered by data centres. The availability of cloud-based solutions has dramatically lowered the cost of storage, which has lead to more UK companies being able to work with big data analytics[31]. Over coming years, the data storage market will continue to grow, with global spending on data centres predicted to reach $149 billion next year[32]. This will present opportunities for UK companies to attract new customers and further investment, and to increase their share of the market.

Action: The government will work with the Data Centre Alliance, Intellect and UK Trade & Investment on options to attract overseas investment and customers to the UK data storage market, and how UK providers can expand their offer in overseas markets.

Connectivity

As well as the ability to store and manage data, a high level of connectivity and good connection speeds are vital for academia and businesses to be able to transmit and collect data throughout the economy. Greater connectivity can also create new commercial opportunities. As more consumers shop on-line, companies are able to analyse browsing and spending patterns and adapt their offer to customers' needs and preferences.

The UK's broadband infrastructure has received substantial commercial investment in recent years. Compared to 2008, average broadband speeds have quadrupled[33], and large parts of the country will have access to high speed 4G mobile broadband services, with 98% coverage expected by 2015.

Alongside this, the government is boosting superfast broadband access across the UK with an investment of £1.2 billion from central government, local authorities and the devolved administrations. The government has recently announced an additional £250 million to extend superfast broadband coverage to 95% of the UK by 2017 and are also exploring with industry how to expand coverage further, using more innovative fixed, wireless and mobile broadband solutions, to reach at least 99% of premises in the UK by 2018. Through the Super Connected Cities Programme, the government is investing up to £150 million to deliver faster and better broadband to businesses and residents in major cities, in order to help them develop the digital

[30] Intellect UK, Data centres and power: Fact or fiction, 2013
[31] NESSI, Big data – A new world of opportunities, 2012
[32] Guardian article, Technology firms to spend $150bn on building new data centres, 2013
[33] Ofcom, Average UK broadband speed continues to rise, 2013

infrastructure capability they need to remain internationally competitive and attractive places to invest, visit and do business. The government is also considering further options for how people and businesses in hard-to-reach areas can access superfast broadband.

Overall business access to the internet is strong, but with business internet connectivity very high across the globe, small differences in adoption rates set countries far apart and the UK is below the OECD average of 96%[34]. This is mainly due to slightly lower subscription rates by micro and small businesses. However, the World Economic Forum's 2013 Networked Readiness Index, which measures the preparedness of an economy to use ICT to boost competitiveness and well-being, places the UK seventh - ahead of countries like the USA, Canada, Japan, Germany, France and Italy[35].

E-infrastructure and computing capability

E-infrastructure refers to the ecosystem of resources that allows: distributed collaboration and computation; large-scale simulation and analysis; fast access to data collections; and analytical and visualisation services and facilities[36].

Central to the UK's e-infrastructure ecosystem is an integrated computing capability, which includes high performance computing (HPC), cloud computing and other emerging technologies, to provide the means to tackle computational problems that are too large, complex or would take too long on standard machines.

The UK now has 24 HPC systems, the fourth highest share in the world[37]. There has been significant investment into HPC resources in the UK: in 2011, the government invested £158 million in high performance computing and networks; followed by further investment of £189 million in big data and energy efficient computing in 2012.

HPCs in the UK are supported by the Janet national research and education network – a high performance, very high bandwidth network for e-infrastructure. Janet has supported research and education since 1984, and now supports a wide range of activities including collaborative research and technology transfer activity. Going forward, the Janet network will increasingly support industry through the nine Catapult centres, and other structures, to enable industrial research and development to benefit from access to the rich set of HPC and data developed and exploited by academic and public funding.

The UK has a growing cloud computing market, which over the next few years is expected to grow at a compound annual rate of 6.4%[38]. Cloud computing takes the concept of having shared computer resources to any organisation of any size at any time, providing fast computing on demand, and enabling users to access facilities and services on a needs basis, rather than having to own everything themselves. Cloud

[34] OECD, Internet economy outlook 2012, 2012
[35] World Economic Forum, The networked readiness index 2013, 2013
[36] BIS, E-infrastructure: ecosystem for innovation – one year on, 2013
[37] Top500.org, Top 500 list, 2012
[38] IBISWorld, Cloud computing in the UK, 2013

users are therefore able to benefit from greater flexibility, access to improved services, and energy savings through resources being used more efficiently.

The importance of computing capability for business development was illustrated in a recent BIS survey where the acquisition of computer software and hardware remain the most commonly reported activities that UK business engaged in during their innovation process[39]. However, there is still more that can be done to increase business understanding and awareness, particularly amongst SMEs, of the different elements of the UK's e-infrastructure resources and how these might support commercial projects.

Action: There will be a programme of activity to drive awareness, support, and access to e-infrastructure for businesses across six key sectors, as well as support for SMEs. This initiative will be monitored by the E-infrastructure Leadership Council and will be supported by organisations including the Technology Strategy Board and relevant Catapult centres, the Hartree centre, regional HPC consortia, and the Open Data Institute. For SMEs in particular, there will be a campaign in 2014/15 to support them to access not only HPC hardware and skilled staff, but also licensed software that is often out of their reach due to cost.

Government infrastructure

The government is a significant holder of data and it is therefore important that its own infrastructure meets appropriate storage and security requirements.

The G-Cloud Programme is a cross-government initiative being taken forward as part of the government's ICT Strategy. The initial focus of the programme is on introducing cloud services into government departments, local authorities and the wider public sector. To do this, a G-Cloud procurement framework has been established and central government departments now have to consider public cloud first in any IT procurement. The wider public sector is also strongly recommended to take the same approach.

As a result of this programme, the public sector will achieve large, cross-government economies of scale; have IT systems that are flexible and responsive to demand; be able to take advantage of new technologies and reduced costs; meet environmental and sustainability targets; and procure in a way that encourages a responsive supplier marketplace and supports emerging suppliers.

[39] BIS, First findings from the UK innovation survey 2011, 2012

Research and development (R&D)

The UK has the opportunity to be a world leader in research and development of big data and data science. We have an excellent research base in computer and mathematical sciences, and in 2011 the UK was among the top five countries in the world in terms of published articles in mathematics and computer science[40]. The UK also shows strengths in big data research publications. Between 2008 and 2012, around 40% of the publications related to big data from two UK universities - University of Leeds and Cardiff University - were in the top-cited 10% of publications globally in this subject area[41]. The UK has a strong environment for innovation, with R&D tax credits, a skilled workforce, and support for science and research.

ELIXIR

Image courtesy of EMBL-EBI

The volume of life sciences data is increasing exponentially, principally as a result of high-throughput gene sequencing technologies that have been developed in the UK over the past decade. This data is of little value unless it is efficiently collected, annotated, archived, analysed and made available to scientists, but this task is so great that no single company, organisation or country can tackle it alone.

Opened by David Willetts on 28 October 2013 with £75 million of government funding, the ELIXIR Technical Hub will be the nerve centre for bioinformatics based in the European Bioinformatics Institute in Cambridgeshire. The Hub will coordinate the delivery of services from several centres of excellence across Europe and also establish a robust computing infrastructure that can handle the rising tide of life science data, which will in turn ensure that the research data produced can be translated into innovations that meet global challenges in food security, energy and health.

[40] BIS, International comparative performance of the UK research base, 2011
[41] Sir Andrew Witty Review, Encouraging a British invention revolution, 2013

The Council for Science and Technology's report 'The Age of Algorithms'[42] emphasised the UK's strengths in both research into, and the business of, algorithms - the mathematical processes that underpin our ability to analyse, manage and improve processes through the use of data. The report made recommendations relating to the need for better links between business and the UK's outstanding research base in these disciplines.

Through R&D into data science and analytics, we can push the boundaries of what it is possible to do with data – new tools, new techniques, and new innovations. In the UK, the seven Research Councils invest around £3 billion in research across the full spectrum of academic disciplines, supporting the excellent research that maintains the UK's strong global position. The Research Councils work in partnership with other research funders such as the Technology Strategy Board, the UK Higher Education Funding Councils, and other bodies. Data analytics skills are important across all of the Research Councils, and work here is coordinated through the UK government and through Research Councils UK.

The UK is also working to maximise opportunities for UK companies and organisations to access EU funding for research and innovation, focusing on the forthcoming Horizon 2020 programme[43] for 2014 to 2020. With a budget of over €70 billion, Horizon 2020 aims to raise the level of excellence in Europe's science base and ensure a steady stream of world class research to secure Europe's long-term competitiveness. It also aims to make Europe a more attractive location to invest in research and development. Through both of these activities, there will be a focus on emerging and enabling technologies, such as information and communication technologies, and developing European e-infrastructure, which will support growing data capability.

Action: The EPSRC is developing a proposal for a national network of centres in big data analytics to be considered as part of the Research Councils' UK Strategic Framework for Capital Investment and dependent on their future delivery plan funding. The centres will develop world-leading capability and capacity in new, transformative tools and techniques to enable UK companies and the research community to be at the forefront of extracting knowledge and value from data.

Supporting collaboration

Just as important as research into new R&D tools and techniques is the potential for innovative use of data to transform R&D in general, maximising opportunities by working across disciplines.

Sir Andrew Witty's review 'Encouraging a British Invention Revolution' recognised the UK's strong record of collaboration between universities and business[44].

[42] Council for Science and Technology, The age of algorithms, 2013
[43] European Commission, Research & innovation, Horizon 2020
[44] Sir Andrew Witty Review, Encouraging a British invention revolution, 2013

The World Economic Forum has ranked the UK in the top five in the world for university-business collaboration in R&D for the past four years.

The government shares the view that UK business needs to make the most of the expertise contained in UK universities to support the rebalancing of our economy. Research Councils UK (RCUK) has developed a web-based portal that gives the public better access to information about research funded by the UK Research Councils. RCUK's Gateway to Research portal provides a mechanism for businesses and other interested parties to identify potential partners in universities to develop and commercialise knowledge, and maximise the impact of publicly funded research.

Working with the Technology Strategy Board, RCUK will organise a Big Data Festival in autumn 2014, which will connect different academic communities who are active in big data, as well as users in business and policy.

The government has created the Connected Digital Economy Catapult (CDEC), supported with over £50 million of funding from the Technology Strategy Board, which has been established with the vision to 'ignite digital innovation to power sustained economic growth in the UK'. CDEC will bring together innovators from industry – both large companies and SMEs – research and academia, promoting alignment and collaboration amongst the many players in the digital innovation community.

A key objective for CDEC will be to develop new tools, platforms and assets to enable UK innovators to take advantage of the opportunities offered by the huge growth in the volume and variety of data. Over the next five years, CDEC will create a 2,000m^2 Innovator Centre in London, where companies can collaboratively develop and showcase their ideas and products. In turn, this will accelerate understanding amongst businesses and the public sector of the transformational impact the digital economy can bring to organisations. Through this, and its network of partner centres, CDEC will promote some of the UK's most creative and innovative digital businesses. As well as CDEC, there are other key organisations that are fostering collaboration between different fields and across different sectors. For example:

Open Data Institute

The Open Data Institute (ODI), based in Shoreditch, is the world's first centre created to help businesses innovate, research and exploit the opportunities of open data.

Opened in December 2012, the ODI has exceeded its year one targets as it convenes world-class experts to collaborate, incubate, nurture and mentor new ideas, as well as promote innovation. The ODI now has 46 organisations as members and
117,000 people have been reached through the ODI programme.

The Institute is an independent, not for profit company, whose services include providing training and learning in open and linked data technologies, and incubation for start up companies working with open data.

midata Innovation Lab

The midata Innovation Lab (mIL) brings together industry, universities, and regulators to collaborate on new ideas for data services innovation and consumer protection.

mIL acts as an accelerator for businesses to access rich datasets to create new services for consumers, whilst also working closely with regulators and universities to create the world's best consumer protection framework in the emerging personal data market.

mIL is looking at how to interest consumers in this new market for personal data services and stimulate the creation of new services and applications. The mIL will publish a report in November which includes a number of prototypes of applications highlighting the benefits to consumers, government and industry.

Sharing and linking data securely and appropriately

Not all data can, or should, be shared openly and freely. For example, administrative data – that is, data collected by governments and other public sector organisations primarily for administrative (not research) purposes – often includes personal data which cannot be made open for confidentiality and security reasons.

However, in many cases there are significant opportunities and benefits, for both the research and commercial sectors, which can be achieved from sharing and linking datasets where the appropriate protections have been put in place, such as by removing personal identifiers or aggregating data.

This chapter looks at some of the different types of data – open, administrative, research and personal - and the ability to share and link this data, including addressing the issue of data standards. The final section covers privacy, data protection and security issues.

Open data and public sector information

Since 2010, the UK has released the largest amount of government data of any country in the world[45], through the data.gov.uk site. This is a searchable portal designed to release data held by the public sector, and currently contains over 10,000 datasets, including datasets held by the government and local authorities. To support this, the government has also supported the creation of the Open Data Institute, the first of its kind in the world, and the Open Data User Group.

Public sector data and information has particular significance given that the European Commission estimates that between 15% and 25% of total data used in e-commerce trading is based on public sector information[46]. Removing barriers to the use and re-use of public sector information can therefore help to deliver economic benefits as well as promote transparency and social engagement. The UK operates under the Re-use of Public Sector Information Regulations 2005[47]. An amended European Directive on the Re-use of Public Sector Information[48] was adopted in June 2013 and will be implemented by July 2015. The new Directive extends the scope of public sector information and provides greater transparency over charging.

The government also commissioned Stephan Shakespeare, then Chair of the Data Strategy Board, to undertake a review of the public sector information market, both current and future, in order to make recommendations to the government on how to widen access to public sector information and consider new and innovative opportunities for open data. In the government's response to Stephan Shakespeare's review, the government committed to setting out a National Information Infrastructure

[45] Cabinet Office, Open data white paper, 2012
[46] PIRA study, Commercial exploitation of Europe's public sector information report, 2000
[47] Re-use of public sector information regulations, 2005
[48] Directive 2003/89/EC on the re-use of public sector information, 2003

(NII) which will contain that data held by government which is likely to have the broadest and most significant economic and social impact if made available and accessible outside of government.

The government is developing the processes to support the maintenance of a dynamic NII and has launched a first iteration[49], which will be the basis for user feedback and for the identification of additional datasets. These processes broadly cover:

- Identifying and maintaining an inventory of data held by government
- Prioritising data to be included in the NII
- Supporting organisations to release data

The UK is widely acknowledged as leading the way on promoting the re-use of public sector information across Europe by introducing initiatives that promote and encourage re-use. Amongst the initiatives introduced by the UK are:

Licensing

The UK has introduced the Open Government Licence[50], which is an open and non-transactional licence widely praised for its clarity and simplicity and copied across the world. The UK Government Licensing Framework (UKGLF), which features the Open Government Licence, was developed in order to maximise the benefits of public sector information by providing a licensing framework which facilitates use and re-use as simply and transparently as possible.

Legislation.gov.uk

Launched in 2011, legislation.gov.uk provides easy access to UK legislation, both for legal practitioners and the general public. It enables users to see both the original (as enacted) legislation, and amended versions of legislation. The content can be used and re-used for commercial and non-commercial purposes under the Open Government Licence.

Information Fair Trader Scheme

The regulatory framework of the Information Fair Trader Scheme (IFTS)[51] establishes a set of standards and principles for public sector bodies in the field of information. IFTS requires the public sector bodies covered by the scheme (mainly trading funds and significant information traders) to maximise the re-use of the information they hold while ensuring they comply with standards of fairness and transparency.

[49] Cabinet Office, National Information Infrastructure, 2013
[50] The National Archives, Government licensing framework
[51] The National Archives, information fair trader scheme

As part of the UK's membership of the Open Government Partnership (OGP), the government has published our OGP UK National Action Plan 2013 to 2015[52] which details commitments to make the UK government more transparent. Specific commitments include developing the National Information Infrastructure and publishing legislation in draft format on www.gov.uk, whenever appropriate, to make the UK more accountable to citizens.

Administrative data

The Administrative Data Taskforce (ADT), chaired by Sir Alan Langlands, published a report in December 2012[53] which considered how access to and linkage between administrative data could best be achieved for research purposes. Responding to the recommendations of the ADT report, the government announced in October 2013 the creation of four new innovative administrative data research centres, and a data service, led by the Economic and Social Research Council (ESRC), which together will form the Administrative Data Research Network (ADRN).

The centres are led by the University of Southampton in England, the University of Edinburgh in Scotland, Swansea University in Wales, and Queens University Belfast in Northern Ireland, and will run for an initial period of five years. The Administrative Data Service will be led by the University of Essex. This network of investment is being enhanced by investment in safe setting environments s and in a UK wide researcher training and accreditation programme.

The newly-formed ADRN will enable research based on linked data from government departments, allowing the analysis of massive and mixed datasets, and making a significant contribution to ensuring the future sustainability of UK research competitiveness.

The development of the ADRN constitutes the first phase of the ESRC's Big Data Network. Phase 2 will focus primarily on business data and local government data, and phase 3 will focus on third sector data and social media data. This infrastructure will enable innovative and powerful new analyses of economic and societal challenge, drawing a rich and diverse range of data in secure and trusted settings. This will be important in securing and maintaining public confidence in the use of data and the ESRC and the Office for National Statistics are working in partnership to engage the public and better understand public views on administrative data[54].

The ADT also recommended that legislation should be enacted to facilitate research access to administrative data and to allow data linkage between departments to take place more efficiently. The government response to the ADT noted the complexity of the rules governing data sharing, and that there could be benefits across the whole policy spectrum from making better use of administrative data for research and statistics. The government has committed to establishing a centre of excellence to reduce the complexity of sharing data between public services, and will explore

[52] Cabinet Office, Open Government Partnership National Action Plan, 2013
[53] Administrative Data Taskforce, Improving access for research and policy, 2012
[54] ESRC, Public dialogues webpage

options for new legislation to simplify the legal landscape where issues are most pressing. This is in addition to the Law Commission's current broader review of the legal framework on data sharing, which will report in spring 2014.

Research data

There is significant activity underway to ensure access to publicly funded research data in particular. The Royal Society report 'Science as an Open Enterprise', published in 2012, recommended that publishing data in a reusable form to support findings must be mandatory for researchers. This should be supported with more experts in managing and supporting the use of digital data, as well as greater recognition of the value of data gathering, analysis and communication[55].

The government recognised the importance of this recommendation by establishing the Research Sector Transparency Board (RSTB). While much of the responsibility for delivering the report's recommendations rest with the research community itself, the RSTB has been set up to advise the government on the approach and to highlight areas where the government can take action.

The RSTB has now met twice and in its most recent meeting discussed the risks around a lack of coordination and standardisation across institutional repositories within a discipline, and the challenge of searching for data across disciplinary boundaries. To seek practical ways forward on this, and to address wider issues such as skills and career development for researchers and data specialists in the sector, the Royal Society will convene an Open Science Data Forum in early 2014. The aim of the forum will be to engage key stakeholders on the implementation of an open data regime in research, including representatives of universities, the Research Councils, and other experts.

Action: At the Open Science Data Forum in early 2014, the research community will work together to develop proposals to support access to, and use of, research data.

The government has also implemented the recommendations of the independent group, chaired by Dame Janet Finch, on expanding access to published research findings. This has helped put the UK in the vanguard of global open access developments, with the government making an initial £10 million of funding available, £10 million available to universities to establish Publication Funds, which will assist with swift implementation of the findings. This funding will be followed up by Research Councils UK's additional allocation of £37 million over the period 2013/14 to 14/15 to more than 100 universities. The Research Councils will review progress at the end of next year.

The government is currently introducing for research data a new qualified exemption from disclosure under the Freedom of Information (FOI) Act. The exemption will protect against the inappropriate or premature disclosure of pre-publication research,

[55] The Royal Society, Science as an open enterprise, 2012

and will provide clarity and reassurance about the ability of the FOI Act to protect genuinely sensitive information. This in turn will help strengthen the UK's leading position in international research, by ensuring that research is not undermined through fear of inappropriate disclosure under the FOI Act. The exemption will only be available where disclosure would be harmful and not in the public interest, and will not lead to a reduction in transparency. The government intends to bring this exemption into force next year.

There is still more to do to ensure that research data is made open safely and in a way that best supports innovation and the UK's thriving research sector, both domestically and working with global partners. There are several positive international moves on open research data, including the recent UK-led G8 Science Ministers' meeting which made a commitment to the principle of open scientific research data. The creation of the Research Data Alliance between the European Commission, the US government and the Australian government, the European Commission's recent Open Data Strategy, and the Horizon 2020 open data pilot are all helpful interventions, in which the UK is playing its part.

Personal data and midata

Businesses hold large amounts of data about consumers – for example, about their transactions and spending patterns preferences. The midata programme seeks to empower consumers by giving them greater access to, and control over, their data. The programme also aims to stimulate economic growth through promoting a new marketplace for services and products created to help consumers access and understand their data. The insights provided by better access to their personal data will allow customers to understand their patterns and find the deals that are the right fit for them.

There are three core parts to the midata programme:

- Working with companies who hold data about their customers to release it back to consumers in a reusable electronic format

- Developing new products and services that can use this data to provide greater insight and independent advice to customers

- Building trust and consumer confidence

The midata programme has already achieved initial successes, for example several energy companies now provide consumers with access to their consumption and tariff data in an electronic format which allow them to understand their energy use and make informed decisions about switching tariff or energy provider. The government wants to explore how we can take this further, in particular to get citizens back the individual data that is held on them by central government.

Action: The government will convene a working group on widening the midata programme, including stakeholders from the private sector, Connected Digital Economy Catapult (CDEC) and consumer organisations.

The midata Innovation Lab (mIL), established to act independently from the midata programme, is bringing together the relevant parties to create a handful of prototype services which highlight the innovative services possible using personal data. These prototypes enable better dialogue between organisations and consumers on the benefits of sharing data, as well as considering issues of consumer protection and trust and making sure UK digital innovators are at the leading edge of this part of the information economy. CDEC will convene a workshop, including stakeholders from the private sector and consumer organisations, to define the next phase of the midata Innovation Lab.

Text and data mining

Professor Hargreaves' 2011 review of Intellectual Property and Growth[56] recommended that the UK introduce an exception to copyright laws to permit the use of data mining for non-commercial research. In response, the government is changing the law to allow this.

The Intellectual Property Office has published draft legislation to amend the Copyright, Designs and Patents Act 1988 so that it is not an infringement of copyright for a person who already has a right to access a work (whether under a licence or otherwise) to copy the work as part of a technological process of analysis and synthesis of the content of the work, for the sole purpose of non-commercial research. This will enable key research without undermining publishers' control over IT systems or commercial exploitation.

This change will enable text and data mining, allowing the UK's world leading research community to deliver significant benefits to organisations and to citizens, through permitting researchers to handle vast quantities of potentially valuable data. It will increase research efficiency by saving researchers' time, and could lead to novel discoveries as large scale analysis uncovers unforeseen connections, and enhance the UK's skills base in analytics through the development of new techniques.

Action: Following the technical review of the published draft legislation, the government will bring into force secondary legislation to enable text and data mining for non-commercial purposes in 2014.

Data standards

The greatest opportunities for innovation come from working across disciplines, and from being able to combine and analyse information and data from a variety of sources and subjects.

Good data standards are essential components of ensuring that data from different sources can be shared and linked. Metadata, that is information about the data itself – for example, where and how it was collected – tells data users and reusers about the

[56] Ian Hargreaves review, Digital opportunity, 2011

provenance and quality of the data, which informs the conclusions they can draw from analysis. Ensuring data is collected and curated consistently with technical and metadata standards is an important task, and one which currently receives less focus and less recognition than it should.

Data standards should be developed by those working in the relevant disciplines, and accepted across the research community, not only within the UK but internationally as well. The Research Councils and the learned societies have roles to play in here, as do other bodies such as the Research Data Alliance, who are working internationally to promote standards for research data, and the Open Data Institute, whose Open Data Certificates ensure best practice in publishing open data, with over 520 datasets published so far.

The 'Science as an Open Enterprise' report recommended that common standards for sharing information are required to make open research data widely useable[57]. Some scientific disciplines have strong and widely accepted data standards and definitions, which allow researchers to compare and combine data for their mutual benefit. Others are less developed in this area, with competing standards creating both real and perceived barriers to data sharing and integration. The extent of data curation varies significantly, within disciplines and institutions, and between academia and industry. The Royal Society-led work on Open Science Data will look at the importance of data standards and metadata, and what practical actions can be taken to promote this.

Privacy and data protection

The use of data needs to be underpinned by a strong legal framework which protects the right to privacy of individuals, while enabling the analysis of data in order to provide innovative new services, products and insights.

Effective data protection is vital to our data capability. Individuals have a right to privacy and having their personal data protected, whether that data is being handled and processed by the government, a private company, or a research institution. As Stephan Shakespeare's review of public sector information recommended, the UK needs a clear and pragmatic policy on data privacy and confidentiality which ensures public trust in the confidentiality of their data, while increasing the availability of data to maximise its economic and social value. A report last year by Booz & Company[58] noted the importance of clear data protection legislation to investment in new start-up businesses.

A modern data protection framework needs to promote economic growth and prosperity by making sure that data protection rules reflect this, along with the reality and requirements of modern enterprise and innovation.

[57] The Royal Society, Science as an open enterprise, 2012
[58] Booz&Co, The impact of EU internet copyright regulations on early state investment: a quantitative study, 2012

The UK supports the need to bring data protection rules in line with the reality of the 21st century. The government does not believe that the European Commission's proposals for reform to the data protection framework strike the right balance between privacy and innovation. We should also be careful about overly prescriptive regulation that increases red tape and costs for businesses, the public sector, and for regulators themselves.

The government wants to see European data protection legislation that protects the civil liberties of individuals while supporting economic growth and innovation, and allowing for the necessary and proportionate use of data by law-enforcement authorities. These aims should be achieved concurrently, not at the expense of each other. The UK government is committed to playing a full part in negotiations with the European Commission to deliver a framework which achieves these objectives.

Stephan Shakespeare's review also recommended that the government should provide clear guidelines to all parties involved in data sharing. The government believes that good practice guidance is essential in enabling data sharing and encouraging open data. The Information Commissioner's Office produces a range of guidance on this subject, available through their website.

Action: Working with the Information Economy Council, the government will look at options to promote guidance and advice on the rights and responsibilities of data users.

In order to enable secure access to digital public services, the government is leading the way in using Identity Assurance (IDA) which enables users operating online to safely and securely assert their identity, and control their personal data. The government has committed to work closely with industry, privacy advocates and consumer groups to develop an IDA solution for government services. BCS, the chartered institute for IT, has also recently published its annual Aspects of Identity yearbook[59], which sets out practical measures for improving identity governance online.

Security of data

As well as ensuring that organisations are responsible users of data, people also want assurance that they are taking the necessary steps to ensure data is protected from theft or cyber attack. Security breaches may undermine the confidentiality, integrity and availability of data with implications for an organisation's competitiveness and reputation, as well as potential legal or contractual penalties.

The government is working with organisations, large and small, to ensure they are aware of how to protect their business and the data they hold from the risk of cyber attack. The government has launched the 'Ten Steps to Cyber Security'[60] guidance for large organisations, and 'What You Need To Know About Cyber Security'[61] for

[59] BCS, Aspects of identity yearbook 2012-13, 2013
[60] BIS, Ten steps to cyber security, 2013
[61] BIS, Cyber security: what small businesses need to know, 2013

small businesses and will shortly be identifying a preferred organisational standard for cyber security to help organisations understand the steps they need to take, for the benefit of their own business and the peace of mind of their customers. Industry is also able to share information on evolving cyber threats through the government-backed Cyber Security Information Sharing Partnership (CiSP).

The government is working in partnership with a number of sectors that are of economic importance to the UK and are targets of cyber attacks, in order to develop understanding of what cyber security means for them, and to encourage and support action. One such partnership is the higher education sector: The government is working with Universities UK to raise awareness of the challenge of cyber security in universities, where a high value is placed on the use and exchange of research data; and also supporting university participation in the CiSP. The government has also supported the establishment of 11 academic centres of excellence in cyber security and is exploring the possibility of supporting high quality cyber security degrees.

Annex A - Summary of actions

Actions
The government will work with employers, e-skills UK, Nesta, Universities UK and the Open Data Institute to explore the skills shortages in data analytics and set out clear areas for government and industry collaboration.
The government will hold a workshop in November 2013 bringing together representatives from universities, businesses and other relevant bodies to discuss computer science graduates and how to get the right skills to meet current and future needs.
Universities UK will review how data analytics skills are taught across different disciplines and assess whether more work is required to further embed these skills across disciplines.
The government will work with the Information Economy Council, e-skills UK and Intellect to develop a plan to bolster the image of the discipline by spring 2014. As part of this, the government will work with the Information Economy Council, Research Councils and other the relevant professional bodies, including BCS and IET, to collate career profiles of people working in data analytics and the different career pathways.
The government will work with the Data Centre Alliance, Intellect and UKTI on options to attract overseas investment and customers to the UK data storage market.
The E-infrastructure Leadership Council will monitor a programme of activity to drive awareness, support, and access to e-infrastructure for businesses across six key sectors, as well as a separate campaign specifically aimed at SMEs.
The EPSRC is developing a proposal for a national network of centres in big data analytics to be considered as part of the Research Councils' UK Strategic Framework for Capital Investment and dependent on their future delivery plan funding.
At the Open Science Data Forum in early 2014, the research community will work together to develop proposals to support the access to, and use of, research data.
The government will convene a working group on widening the midata programme, including stakeholders from the private sector, CDEC and consumer organisations.
Following the technical review of the published draft legislation on copyright exceptions, the government will bring into force secondary legislation to enable text and data mining for non-commercial purposes in 2014.
Working with the Information Economy Council, the government will look at options to promote guidance and advice on the rights and responsibilities of data users.

Annex B - Glossary of terms

Administrative data	Administrative data collection is the set of activities involved in the collection, processing, storage and dissemination of statistical data from one or more administrative sources. (Source: OECD Glossary of Statistical Terms)
Aggregated data	A form of anonymisation of data, involving combinations of records so that individual records are not disclosed or identified.
Anonymised data	Data containing only anonymised records, where direct identifiers have been removed. (Source: OECD Glossary of Statistical Terms)
App	An abbreviation for “application”, it generally refers to a piece of software which runs on a phone or mobile device, although apps can also run on computers or the internet.
Big data	Gartner describes big data as high volume, velocity and variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.
Cloud computing	On demand access to a shared pool of remote computing resources, such as software, data storage and analytics, accessed via the internet and managed by the user.
Data centre	A facility used to house computer systems and associated components such as telecommunications, storage, and cooling systems.
Data controller	A person who (either alone or jointly with other persons) determines the purposes for which and the manner in which any personal data are, or are to be, processed.
Data re-use	Use of information other than for the purpose for which it was originally produced – whether for commercial or non-commercial purposes. (Source: adapted from EU PSI Regulation)
Dynamic data	Data that are changed or updated at uncertain intervals as new information arises.
Identifier	A particular element or reference in a dataset that allows an individual or business’s identity to be known.

Internet of Things	A concept relating to devices and systems working together in ways that are helpful to the user and society. It is underpinned by machine to machine communications and other technologies such as RFID, radio, internet, mobile telecommunications.
Linked data	Data described by an identifier and addresses to permit linking with other relevant data which might not otherwise be connected, improving discoverability. It may contain embedded links to other data. (Source: Open Data White Paper)
Metadata	Data that describes or defines other data. This includes anything that a data user needs to know in order to make proper and correct use of the real data, in terms of reading, processing, interpreting, analysing and protecting the information. (Source: OECD Glossary of Statistical Terms)
Open data	Data that meets the following criteria: a) accessible (ideally via the internet) at no more than the cost of reproduction, without limitations based on user identity or intent; b) in a digital, machine readable format for interoperation with other data; and c) free of restriction on use or redistribution in its licensing conditions Open data is not limited to the public sector – it can also be provided by the private sector or by individuals. (Source: Open Data White Paper)
Personal data	As defined by the Data Protection Act 1998, personal data means: “Data which relates to a living individual who can be identified a) from those data; or b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller; c) and/or includes an expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual”

Pseudonymised data	Data relating to a specific individual where the identifiers have been replaced by artificial identifiers to prevent identification of the individual.
Public sector information	Data and information that the public sector collects, produces, reproduces, publishes and disseminates in many areas of activity while accomplishing their public task or other duties. (Source: Adapted from Deloitte's Market Assessment of Public Sector Information)

This publication available from www.gov.uk/bis
Any enquiries regarding this publication should be sent to:

Department for Business, Innovation and Skills
1 Victoria Street
London SW1H 0ET
Tel: 020 7215 5000

If you require this publication in an alternative format, email enquiries@bis.gsi.gov.uk, or call 020 7215 5000.

BIS/13/1250