ROYAL
ACADEMY OF
ENGINEERING

IET The Institution of
Engineering and Technology

Connecting data
driving productivity and innovation

November 2015

Connecting data
driving productivity and innovation

Contents

Foreword

The UK has many of the key assets, in terms of knowledge, technologies and experienced individuals, needed to play a major role as leading nations move to exploit an increasingly information-driven global economy. Harnessing the power of data analytics (more colloquially 'big data') and advanced connectivity, reliably linking key datasets, often in real time, has immense potential to drive innovation. This may come from converting the sale of existing products into a broader sale of capability, such as the already successful transformation of the aero engine business model. It may come through the creation of more integrated and sustainable urban development through 'smart city' initiatives. It may return power to consumers as they take more control of and profit from their personal data. Such techniques and technologies offer the potential to transform UK productivity, currently lagging in the G7.

This report highlights the experience to date across multiple sectors of the economy. It documents both best practice for further dissemination and pitfalls to avoid. The report identifies a series of policy actions for government, professional institutions and other stakeholders that would facilitate new areas of economic activity, such as markets in data, and also recommends steps necessary in areas from personal privacy to resilience and cybersecurity.

The UK is at the start of a dramatic transformation that will impact on the working and private lives of all our citizens. The Royal Academy of Engineering and the Institution of Engineering and Technology are determined that the way ahead should draw on the key principles of best engineering practice and look forward to continuing to facilitate these vital developments.

Professor Jim Norton FREng FIET
Chair of the Connecting Data Working Group

Executive summary

How can the UK create a 'data-enabled' economy through the use of data analytics (more colloquially 'big data'), supported by data science and advanced data connectivity? This was the question at the heart of a series of sector-specific stakeholder workshops run by the Royal Academy of Engineering and the Institution of Engineering and Technology (IET). Improved use of data analytics offers a critical tool for addressing the UK's nagging productivity deficit of 17% relative to the average of the G7 economies that was highlighted in HM Treasury's Command Paper[1] published in July 2015. The paper stressed the government's commitment to raising UK productivity by encouraging long-term investment and promoting a dynamic and innovative economy. The workshops sought to identify the key foundations of a data-enabled economy and where investment would be needed.

The sectors explored were advanced manufacturing, built environment, energy, transport, health, aerospace and defence, and insurance, chosen to cover a broad range of the sectors that make up the UK's economy and serve society. The workshops examined the future opportunities for organisations and sectors to improve products and processes and to innovate using data analytics, and explored barriers to success, both universal and sector-specific. The workshops also investigated the progress that had been made so far in each sector in the use of data analytics and advanced connectivity, and identified current examples of best practice. The scope extended to the networks of sensors and devices that will generate, communicate and respond to data, known as the Internet of Things, and the systems that will be created by integrating previously separate networks. The majority of workshop participants were engineers who were developing new applications of data analytics, or already had used data analytics or data science successfully.

The findings suggest that the UK is strongly placed to develop a leading data-enabled economy. There are early pockets of excellent practice in data analytics and some companies already have a strong capability in this area or the potential to realign existing capabilities in data to address new applications. However, the challenge will be to spread these examples of best practice through all sectors of the economy, particularly those that are only now becoming reliant on data. The area is still immature and there remains great potential for innovation and value generation in future years, to the benefit of business and society as a whole. Case studies presented in this report demonstrate benefits that are more broadly applicable, both within and between sectors.

THE AREA IS STILL IMMATURE AND THERE REMAINS GREAT POTENTIAL FOR INNOVATION AND VALUE GENERATION TO THE BENEFIT OF BUSINESS AND SOCIETY AS A WHOLE

Data analytics offers opportunities for individual organisations and sectors to transform the way that they provide products and services. Many of the applications that have emerged so far are single sector, but in the future, new opportunities will arise as a result of combining data across sectors in innovative ways. There are opportunities to drive innovation by combining open data and proprietary data shared between organisations, along with novel types of unstructured data such as social media and crowdsourced data.

Technical, organisational, regulatory and legal barriers common to all sectors are present, particularly where innovative products and services challenge existing ways of working. For example, there are technical challenges around developing high-quality datasets and complex data-driven systems, reliably linking data sets that have originated in different contexts, and ensuring that appropriate security and privacy mechanisms are established. The right conditions need to be in place to allow innovation to occur, and these include appropriate legislation around data protection and ownership of data, and adequate access to data, whether open or proprietary. Standards for achieving interoperability between datasets, devices and systems and for building trust in the use of data are also required. Individual organisations face challenges in creating new business models and developing the capability to capitalise on data analytics. The UK will need to build on its considerable existing capabilities in multidisciplinary innovation by addressing barriers that otherwise might reduce the UK's international competitiveness in this field, including the need to ensure that data sharing and the operation of data-driven systems can occur across international, as well as sectorial and organisational, boundaries.

The key messages and recommendations are:

Strengthening the foundations

The growing level of interconnection and interdependence of the supporting systems of a modern economy creates many opportunities for improved performance and innovation but, at the same time, increasing vulnerability to unanticipated emergent behaviour including cascades of failure. Moving to a data-enabled economy, while essential in terms of UK international competitiveness, will place further strain on these already fragile systems that are also subject to the threat of cybersecurity breaches. For example, relatively simple embedded systems[i] have proliferated throughout the UK's infrastructure and have been progressively linked to the internet, often without thought to the security implications. It is therefore vital that best engineering practices, and in particular systems engineering approaches, are deployed in software development. While this is a generic issue of relevance to all software development, it is of particular relevance to critical infrastructure and the Internet of Things because of the risks of creating vulnerable systems of systems.

The importance of deploying tools and methodologies for much higher quality in software development for critical infrastructure and the Internet of Things needs to be recognised. It is possible to build simple command and control systems[ii] using rigorous methods that build on mathematical foundations.

> **Recommendation 1**: Infrastructure regulators (such as Ofgem, Ofcom and Ofwat), working together with professional institutions and standards bodies, should work towards establishing an enabling structure that both promotes innovation and ensures safety and resilience. As part of this, the development of a coherent suite of technology-independent codes, guides and specifications could be considered that provide rigorous performance, compliance and resilience frameworks both for the implementation of new infrastructure and for the phased improvement in the resilience of legacy systems. Such a system, if well designed, should provide a solid foundation on which innovation could flourish.

Additional mechanisms to improve quality should be debated further by relevant stakeholders, such as the introduction of enhanced liability on producers and importers of software, hardware and systems in certain situations.

---

i   An embedded system is a computer system with a dedicated function within a larger mechanical or electrical system.

ii   Systems that have the ability to issue instructions to and monitor activities of multiple devices from a central location.

Enhancing the infrastructure

A prerequisite for a data-enabled economy is investment in broadband access, at least compliant with the EU Digital Agenda for Europe targets of universal fixed access at a minimum of 30 Mbits/sec download by 2020. Similarly, ubiquitous access to high-speed mobile broadband services is required. This is crucial to support a data-enabled economy, requiring the transfer of data in large volumes and at high speeds, or the resilient transfer of data in both rural and urban areas. It stands to benefit all sections of business and society. Sectors such as advanced manufacturing and agriculture, although making a significant contribution to UK GDP and rural economies, are particularly at a disadvantage. Furthermore, competitor economies such as Germany and the Irish Republic are already implementing EU targets.

> **Recommendation 2**: Government (DCMS) and Ofcom should continue to explore new avenues to extend the Universal Service Obligation (USO) structures to broadband access and should target 30 Mbits/sec download speeds rather than the 5 Mbits/sec initially suggested in the government's *Digital Communications Infrastructure Strategy*. This is compliant with the EU Digital Agenda for Europe targets already being implemented by competitor economies.

Creating markets in data

The open data initiatives of recent administrations have provided welcome first steps in the creation of a data-enabled economy. However, while government has led the way, much potentially valuable data remains locked away in corporate silos or within sectors, although some data is already traded within the supply chains of individual sectors. The next step, in areas that do not impinge on the privacy of personal data, should be the creation of platforms to enable proprietary datasets to be traded within a framework that promotes trust and practicality.

> **Recommendation 3**: Multidisciplinary working groups should be established and tasked with debating mechanisms to allow the controlled trading of engineering data up and down the supply chains within two trial sectors. They should explore new and sustainable business models for data trading, as data analytics facilitates an evolution from selling products to selling service or availability. These groups should work with all interested parties including the Open Data Institute and standards bodies, companies and professional bodies, to develop the necessary sector-specific and trans-sector standards for metadata, calibration, accuracy and timeliness to provide a firm and trusted foundation for data capture, trading and re-use.

Establishing methodologies for the formal valuation of data assets

Many information age companies have almost no physical assets reported in their statutory accounts and their key assets (such as customer datasets) are regarded as intangible and essentially impossible to value other than through a full corporate acquisition or disposal. By failing to measure this key and growing element of national value creation, UK assets are being underestimated and a big element of wealth creation omitted from the tax base. A consequence is a lack of focus and potentially poor decision-making on how best to develop, trade, protect and exploit such assets. This issue may also lie at the base of boom and bust cycles of stock market valuations of these new types of company. Despite the real challenges, the benefits of introducing valuation methodologies into corporate governance and accounting practice should be further explored.

> **Recommendation 4**: The Royal Academy of Engineering should investigate data valuation methodologies with relevant stakeholders. These could include representatives from the Financial Reporting Council (FRC), the accounting professional bodies, professional engineering institutions, the Open Data Institute and leading economic academics.

MUCH POTENTIALLY VALUABLE DATA REMAINS LOCKED AWAY IN CORPORATE SILOS OR WITHIN SECTORS

Sharing best practice

Within each of the sectors researched in this report, areas of excellent practice and innovation have been identified. If these practices were shared more broadly both within and between sectors, substantial improvements in performance could be realised. They could also be the basis of new education modules.

Recommendation 5: Sector Catapults should be tasked with developing and implementing mechanisms for sharing best practice in data analytics and connectivity with particular emphasis on demonstrably reaching SMEs within sector supply chains. This action could be linked to recommendations made on creating markets in data. The Catapults should seek the involvement of the professional networks operating within the professional engineering institutions and standards bodies for advice on the development of best practice codes and guides for SMEs applicable across business and industry sectors.

Developing multidisciplinary skills

A unanimous concern across all the sectors researched concerned the dearth of the multi-skilled workforce required to convert data analytics theory into genuine business practice and performance. The required combination of skills is challenging, drawing on engineering, computer science, mathematics and statistics as well as specific sector knowledge. Engineers with data science skills and specialist data scientists are in short supply. At present organisations painstakingly develop these skilled individuals in-house and then fight to retain them in a market of scarcity. This report makes recommendations as to how both undergraduate and postgraduate training might address these new challenges for a data-enabled economy.

Recommendation 6: The Royal Academy of Engineering and the IET should convene a group of professional bodies including the Institution of Mechanical Engineers, the Institution of Civil Engineers, the British Computer Society, the Institute of Mathematics and its Applications and the Royal Statistical Society to recommend changes in undergraduate and postgraduate education as well as continuous professional development to reflect the new demands for multi-skilled individuals and teams with data science skills.

A UNANIMOUS CONCERN WAS THE DEARTH OF THE MULTI-SKILLED WORKFORCE REQUIRED TO CONVERT DATA ANALYTICS THEORY INTO GENUINE BUSINESS PRACTICE

1. Introduction

The generation of 'big data'[iii] as a result of the increasing use of computers, sensors, and other digital devices, combined with increasingly networked systems and improved analytics, will help to improve, and possibly even transform, how all sectors of the economy operate. This will create economic value for the UK by increasing productivity, driving innovation and changing the pattern of employment. It will also achieve broader social and environmental benefits, by facilitating improvements in healthcare, transportation, energy use, the built environment and national security.

While there are best practice examples within engineering and data science where data analytics has been successfully applied, the area is still immature, so there remains great potential for innovation and value generation in future years, to the benefit of business and society as a whole. Some companies already have a strong capability in this area or the potential to realign existing capabilities in data to address new applications. Many of the emerging applications are from a single sector; new opportunities will arise as the result of combining data across sectors.

However, increasing reliance on data, devices and systems comes with major technical risks. It is critical that these are addressed to prevent the creation of vulnerable systems that fail in cascade, leading to broader practical and economic damage. Specific technical challenges need to be considered around the complexity, uncertainty and resilience of the systems of systems that are created by integrating previously separate networks of sensors and devices. Cybersecurity is another major concern. Regulatory and organisational issues will also need to be identified and addressed to enable the potential of data analytics to be fully realised, particularly where innovations in technologies, products and services challenge existing ways of working.

Policymakers, engineers, data scientists and others involved in big data and data analytics will benefit from collaboration in order to lay the foundations for a functioning data-enabled economy. As well as the technical challenges described in the previous paragraph, there are other important challenges: building a solid digital infrastructure, including expansion of access to fast broadband to all areas, to enable the effective collection and distribution of data; creating sustainable markets in data by finding ways to trade securely proprietary data that can be reliably linked with other data, to the benefit of supply chains both nationally and internationally; and the capability to value data so that companies focus on its protection as an asset from physical and cyber threats. The UK already possesses strong capabilities in multidisciplinary innovation but will need to address barriers that otherwise might reduce its ability to compete internationally, for example by ensuring that data sharing and the operation of data-driven systems can happen across international boundaries.

This report is based on findings from a series of seven sector-specific workshops convened by the Royal Academy of Engineering and the IET. The sectors explored were advanced manufacturing, built environment, energy, transport, health, aerospace and defence, and insurance. These sectors represent only part of the big data opportunities available, and were chosen to cover a broad range of sectors that make up the UK economy.

MANY OF THE EMERGING APPLICATIONS ARE FROM A SINGLE SECTOR; NEW OPPORTUNITIES WILL ARISE AS A RESULT OF COMBINING DATA ACROSS SECTORS

iii 'Big data' is the data that is produced in large volumes, at high velocity and with a diverse nature as a result of the digitisation of services, and the production of data by computers, sensors and other digital devices.

The workshops examined future opportunities to improve products and processes and to innovate using data analytics, and explored barriers to success, both universal and sector-specific. The workshops also investigated the progress that had been made so far in the sectors' use of data analytics and advanced connectivity and identified current examples of best practice. The scope of the workshops extended to consideration of the networks of sensors and devices that will generate, communicate and respond to data, known as the Internet of Things, and the systems that will be created by joining up these networks. The opportunities and barriers for individual sectors are described in Section 3, along with case studies illustrating current best practice. Sections 4 to 9 discuss the issues raised at workshops that are relevant across all sectors.

1.1 National policy context

Recognition by the government of the importance of big data to the economy was made explicit when big data was named as one of the Eight Great Technologies[2] in 2013. Since then, strategies for the information economy[3] and the UK's data capability[4] have been published. The latter observed that leading business sectors, such as aerospace, agri-tech, automotive, healthcare, life sciences and media and telecommunications, could be "leaders in understanding the context of data within their own domains", and recognised a need to place this within a strategic framework that took into account the UK's capability, infrastructure and data.

The importance of the Internet of Things and the algorithms that underpin data analytics techniques are recognised by the UK government's Chief Scientific Advisor[5] and by the Council for Science and Technology[6]. The Alan Turing Institute, a research institute specialising in data science, was created in response; scientific activities at the institute commenced in September 2015. This institute, along with other organisations such as the Science and Technology Facilities Council, are important as they enable explorative work around data, which is essential in understanding

future opportunities, as well as uncovering new insights and testing hypotheses.

Government made clear its commitment to open data and transparency in its white paper, published in June 2012[7], that coincided with the publication by individual government departments of open data strategies[8]. Government also signed up to a charter to promote the release of data by G8 governments[9].

The overall economic context for this report is highlighted in the HM Treasury Command Paper *Fixing the Foundations[10]: Creating a more prosperous nation,* published in July 2015, that highlighted the 17% productivity gap between the UK and the average of the G7 economies. That paper stressed the government's commitment to raising UK productivity by encouraging long-term investment and promoting a dynamic and innovative economy.

1.2 International policy context

Within Europe, Germany is focused on developing technologies and applications based on data and the Internet of Things[11][12][13]. In the US, too, the importance of big data and data analytics in addressing national challenges, alongside other digital technologies, has been recognised[14][15][16]. Big data continues to be a federal government funding priority at the time of writing[17].

The European Commission's *Digital Agenda for Europe* acknowledges that 'digital technologies have enormous potential to benefit our everyday lives and tackle social challenges'[18], and the Commission observes that there are significant opportunities for Europe in immature areas where global players have not started to dominate, despite the fact that the US is so far leading Europe on the establishment of major data-centric companies[19]. In May 2015, the European Commission set out its strategy for a digital single market[20]. It identified big data, along with cloud services and the Internet of Things, as being central to the EU's competitiveness. In October 2015, the European Commission published a *Single Market Strategy*[21] that builds on elements of its strategy for a digital single market.

BIG DATA WAS NAMED AS ONE OF THE EIGHT GREAT TECHNOLOGIES IN 2013

THE EU *DIGITAL AGENDA FOR EUROPE* AIMS TO ACHIEVE NATIONWIDE BROADBAND COVERAGE ABOVE 30 MBITS/SEC BY 2020

1.3 Context for connectivity

The UK is at a point of substantial change in the competitive markets underpinning the connectivity that is an essential prerequisite for an advanced data-enabled economy.

Regulatory and policy reviews have been announced including:

- Ofcom's strategic review of digital communications markets[22], announced on 12 March 2015 and

- DCMS-DCLG review of how the planning system in England can support the delivery of mobile connectivity[23], announced in July 2015.

The EU *Digital Agenda for Europe* includes an objective to increase the speed of internet access for all its citizens, that aims for nationwide broadband coverage above 30 Mbits/sec by 2020. The Commission is investing funds in broadband infrastructure to achieve this[24].

2. What level of economic and other benefits might be unlocked?
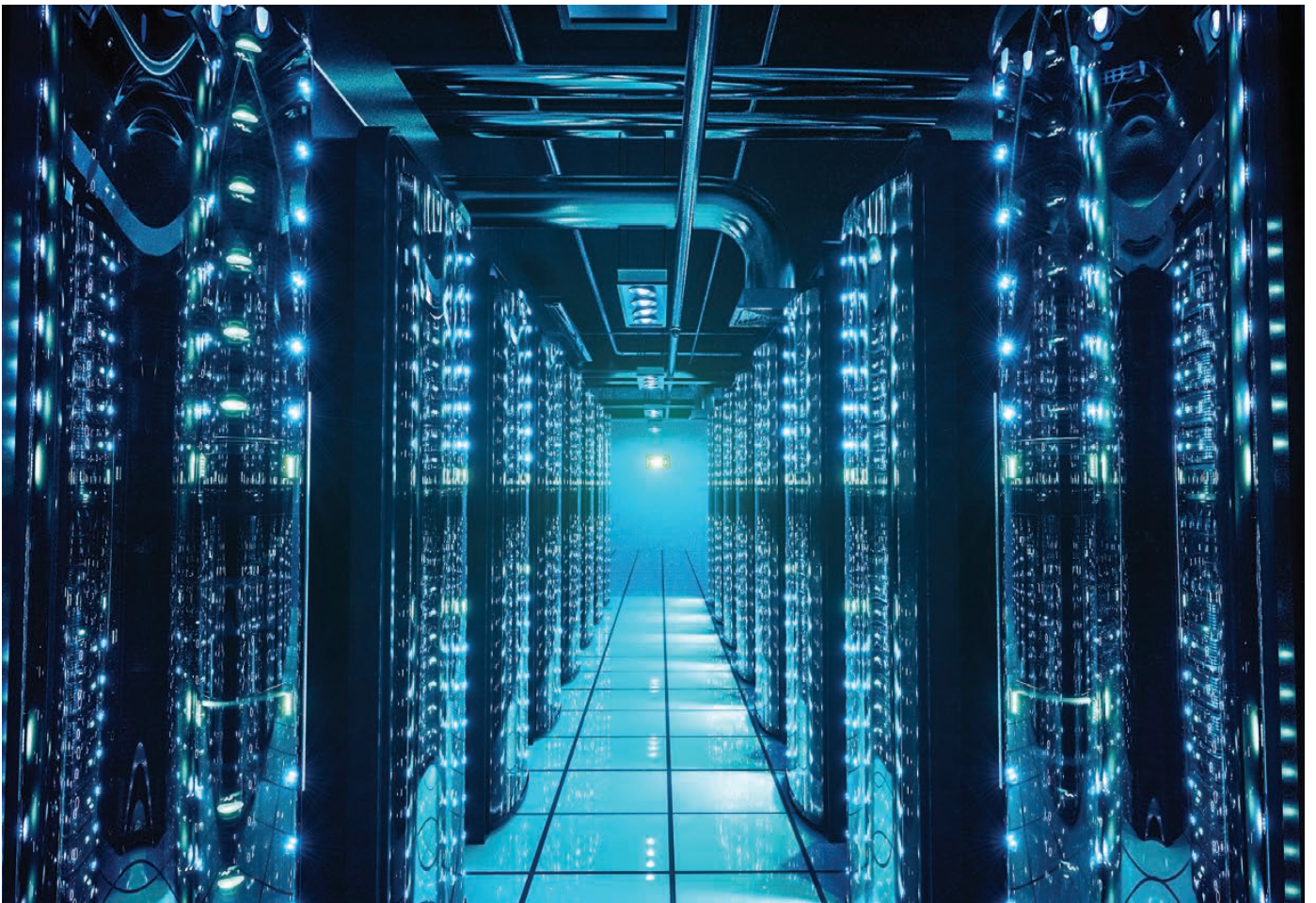
SOCIAL BENEFITS INCLUDE IMPROVED HEALTH AND WELLBEING, MORE CITIZEN-FOCUSED SERVICES AND INCREASED SAFETY AND SECURITY

2.1 Social, environmental and economic benefits

Data analytics has the potential to create social and environmental benefits, in addition to generating economic value. Environmental benefits include improvements in the use of resources such as energy and water. Social benefits include improved health[25] and wellbeing, greater citizen empowerment and more citizen-focused services, better-informed consumers, increased safety and security and reduction of traffic congestion. In 2011/12, Deloitte estimated the economic value of public sector information to consumers, businesses and the public sector to be approximately £1.8 billion (at 2011 prices)[26]. If social value were to be taken into account, an aggregate estimate of between £6.2 billion and £7.2 billion was generated.

CEBR[27] has estimated that, by 2017, the annual benefits to the UK economy of unlocking the economic value of big data through data analytics are expected to amount to £40.7 billion. Furthermore, the cumulative benefits of £216 billion over the years 2012–17 is equivalent to a 2.3% share of CEBR's forecast of total cumulative UK GDP over the same period. A study by McKinsey[28] indicates that open data has the potential to enable more than $3 trillion in additional value annually across seven domains of the global economy, including education, transportation, consumer products, electricity, oil and gas, healthcare, and consumer finance. The study does not include the societal benefits. It emphasises the value of combining open data (such as that generated by government) with proprietary data shared between organisations. In the UK, the Open Data Institute has been formed to capitalise on the potential value to be created from open data[29].

Big data also has the potential to impact on the creation of jobs. A report by e-skills (now The Tech Partnership) predicts that the employment of big data specialists will increase by 243% between 2012 and 2017 (from 20,000 to 69,000), and the employment of big data users will increase by 177% between 2012 and 2017 (from 233,000 to 644,000). However, the benefit from new jobs will only be realised if jobs can be filled with the appropriately skilled people.

2.2 Data opportunities and threats

Boston Consulting Group (BCG) describes three disruptive waves of change unleashed by the global internet[30]. It observes that in the most recent wave "half of the world's data can now be put together, at near-zero cost, to reveal patterns previously invisible. Half of the world's data is already, technically, a single, universally accessible document". It discusses a new business architecture that is emerging for businesses whatever their size, as a result of their ability to access data. However, there are major challenges around linking data together, due to a lack of consistency in the content and assumptions behind each dataset.

New types of company have emerged, particularly in the US, whose business models are based on the aggregation of data and provision of cloud services, such as Amazon, Google (now part of Alphabet), Facebook, Microsoft and Apple. The value of some of these companies is unprecedented, and the value that they generate is of a similar magnitude to that described earlier in CEBR's report[31]. These companies hold huge quantities of data and are in a position to compete with traditional engineering

SOCIAL MEDIA GENERATES DATA THAT ALLOWS SOCIAL PROCESSES TO BE STUDIED AS AN ALTERNATIVE TO SURVEYS

sectors for a share of the market in such areas as autonomous cars and smart cities. They are also competing for graduates skilled in data science, and are able to pay large salaries. Considerable value is also being generated in smaller data companies, such as US-based Planet Labs, which collects and analyses earth observation data with myriad applications[32]. The UK does not have a good defence against such competition, although the creation of the Alan Turing Institute will help counter this, as will the increase in data science courses provided by UK universities.

Novel kinds of data are emerging from social media and more broadly from the digital footprints that individuals leave, whether intentionally or otherwise. Social media generates data that allows social processes to be studied as an alternative to surveys. Three major applications of social media data include elections and polling, security, and commercial applications[33]; such data is therefore useful to government, national intelligence agencies and commercial organisations. The challenges of extracting information from new types of data, and of dealing with a combination of 'soft' data from social media and 'hard' data from sensors, is discussed further in Section 4.1.

Crowdsourcing is also generating new types of data. This data is provided by members of the public specifically to help solve a common problem; for example, Waze[34], a navigation app, gathers map data and traffic information from its users to provide routing and real-time traffic updates. Crowdsourcing also has the potential to aid disaster response. For example, following the Haiti earthquake in 2010, crisis-mapping software[35] provided a way to capture and share critical information coming directly from Haitians, by means of social media and text messages. The scientific community is also benefiting from data contributed by the public, such as weather observations and observations about wildlife, and platforms are emerging that facilitate this[36].

Often simply joining up existing internal data can help develop better products, deliver better services and combat fraud, and offers huge benefits. In the public sector, the use of internal data is helping to ensure legitimate and automated access to services. However, combining such data introduces uncertainty that needs to be managed. For example, it is necessary to

determine the level of confidence needed to ensure that records in two systems are about the same person.

Using data to build customer or citizen-centric views from both internal and external sources allows commercial and public sector organisations to differentiate and personalise their services. For example, retailers are able to provide relevant and timely offers. Customers expect information that they provide to be used by organisations to deliver useful services, and the ability of the organisation to sustain the trust of individual customers and to protect organisational reputation is a central concern in the ethical use of customers' data.

There are considerable business opportunities for the digital communications industry in supporting activities around data. The GSM Association presents the opportunities for mobile network operators around intelligent transport systems and smart cities[37]. Mobile phone data can be used in numerous other ways. One example, involving data sharing across sectors, is the use of mobile phone data in fraud management to alert banks when a customer is overseas[38].

Opportunities also exist in developing technologies and applications for robotics and autonomous systems (RAS) that have been described as "the arms and legs of big data"[39]. Applications extend beyond manufacturing into transport, energy, healthcare and smart cities, and are often based around data gathering and remote monitoring[40].

Many companies are already gathering high volumes of data and are users of sophisticated data analytics[41]. There are also examples of UK-based companies that have capitalised on new opportunities, and are able to provide services around big data across a range of sectors[42].

2.3 The importance of valuing data

Many information age companies have almost no physical assets reported in their statutory accounts and their key assets (such as customer datasets) are regarded as intangible and essentially impossible to value other than through a full corporate acquisition or disposal. At present, the value

THE KEY TO
DETERMINING THE
VALUE OF DATA IS
TO UNDERSTAND ITS
PROVENANCE, THE
PROCESSES THAT
HAVE CREATED IT AND
THE ASSUMPTIONS
MADE WHILE IT IS
BEING PROCESSED

of data assets (along with other intangibles) on corporate balance sheets is only estimated in the context of the 'goodwill' imputed by the difference in price paid in an acquisition or disposal and the value of physical assets transferred.

Without recognition of the value of data assets, the contribution of data to the UK economy cannot be assessed and its growth over time cannot be measured. In addition, a large component of wealth creation is missing from the tax base. Furthermore, decision-making on how best to develop, trade, protect and exploit such assets is potentially compromised. The lack of recognition may also lie at the base of boom and bust cycles of stock market valuations of these new classes of company. The benefits and challenges of valuing data are set out by CEBR[43]. The Information Society Alliance, EURIM[44] and others[45 46 47], also discuss the importance of considering information as an asset.

Data typically has a specific target use, and value is determined according to that use. A dataset does not have a single value, but rather it has different values for different purposes. The key to determining value is to understand its provenance, the processes

that have created it and the assumptions made while it is being processed. The ability to reliably link datasets also creates new value. Datasets may be particularly valuable if they enable correlation and validation (triangulation) of data. For example, a dataset that ties together passport numbers, driving license numbers, national insurance numbers and name and address data would be incredibly valuable as well as being highly sensitive. One that contains a bus timetable is useful but not so valuable unless it has reference points that can be tied to geospatial data.

As for other intangible assets, it is hard to put a value on data[48], particularly if there are few existing examples of how the data is useful. Notwithstanding, the benefits of valuing data are of such potential importance that the development of valuation techniques should be addressed by the accounting industry and by government, as data assets. It is vital that the value of data is formally recognised on company balance sheets. Valuation methodologies should be developed and gradually introduced into corporate governance and accounting.

3. What are the opportunities within and between sectors?

Table 3.1 shows the breakdown of predicted economic value of big data by sector: these predictions may underestimate value as they were calculated at a time when weak growth was anticipated. It is possible too that opportunities to create value from sharing data between sectors are not accounted for.

Table 3.1: Economic value broken down by sector

Industry	2011	2012-17
Manufacturing	5,965	45,252
Retail	3,406	32,478
Other activities	3,446	27,929
Professional services	3,039	27,649
Central government	2,517	20,405
Healthcare	1,450	14,384
Telecoms	1,465	13,740
Transport and logistics	1,360	12,417
Retail banking	708	6,408
Energy and utilities	660	5,430
Investment banking	554	5,275
Insurance	517	4,595
UK economy	**25,087**	**215,964**

Source: CEBR[49], units £million

Arguably, the greatest potential lies in sectors such as advanced manufacturing and healthcare where the value is high. Other sectors such as energy and transport already have systems that are subject to data-driven optimisation; for these sectors, the potential to extract additional value is more limited.

The following sections describe the opportunities for each of the individual sectors that were the subject of the sector-specific stakeholder workshops, as well as opportunities across sectors, and highlight some of the main barriers.

3.1 Advanced manufacturing

The 2013 Foresight report on manufacturing[50] identifies big data, in combination with other enabling technologies such as the Internet of Things, mobile internet and cloud computing, as important components supporting developments in manufacturing, including the digitisation of manufacturing value chains connecting customers, suppliers and manufacturers; higher performance and more flexible manufacturing systems; and better customisation of products and services, embracing bespoke production. These have the potential to make manufacturing processes faster and more flexible and resource-efficient, and to help the UK manufacturing sector become more competitive.

"Most supply chain companies, as they become smaller, are less well equipped to deal with data and understand the value of dealing with data... whether looking at connectivity, security or skills, at each level, as you move down towards the SME size of the supply chain, there is progressively less competence to handle that."

Andrew Churchill, Managing Director, JJ Churchill Ltd

"The real challenge is not showing either people on our sites or the regulator what the shiny analytics are, but the traceability of the action taken from that information, and how that has changed how we operate."

Nicola Walker, Informatics Lead, Advanced Manufacturing Technologies, GSK

The vision of "decentralised, autonomous real-time production" is also embraced by the German 'Industrie 4.0' concept. Embedded systems and global networks, of which data and the internet are key components, are identified as the technologies driving Industrie 4.0[51]. Germany has taken a strategic approach to the development of standards that will underpin the Industrie 4.0 concept[52], a necessary component to ensure the success of the initiative.

The integration of internet-connected sensors into products supports the evolution of existing business models or the emergence of new 'servitisation'[iv] models for manufacturers (see Box 1). New business models will emerge around the ownership of data, where data is shared between consumer and manufacturer. Business standards used in the procurement of goods and services between companies will be required. The challenge of enabling these technologies across the supply chain is considerable, particularly for smaller companies. Supply chain companies that are able to embrace new technologies will be able to compete on a global basis.

With management and analysis of data in a digital form, regulators can more easily investigate alleged breaches by manufacturers and businesses generally can more easily demonstrate compliance with industry standards to their customers. For example, auditors checking compliance with standards will be able to scrutinise performance quality and products within primes[v], contract manufacturers and supply chains to a much greater depth if information around certification and batch documentation is available digitally.

The sector will need to address the barriers to taking on big data and associated technologies, particularly within smaller companies in the supply chain. These include a leadership culture in some organisations that is closed to new techniques and technologies; the difficulty of attracting new recruits with the right skills to the sector; the reluctance of the supply chain to share data because of perceptions that this will damage their competitiveness and the lack of sufficient connectivity in rural areas, where factories may be based.

---

iv 'Servitisation' describes the process of enhancing the value of products by adding services, or even selling services in place of products.

v 'Primes' are the companies at the top of supply chains that assemble or distribute finished products

BOX 1: SERVITISATION MODELS

New business models are emerging across a number of sectors, including built environment, transport, defence and aerospace, where data is used to provide a service that goes beyond the sale or handover of the product.

A service model is possible where a product is remotely monitored so that the need for maintenance or spare parts is identified. For example, in construction, 'product lifecycle management' technologies can enable a more intelligent approach to asset management, by means of a condition-based monitoring regime through the lifecycle, rather than a regime of planned preventative maintenance[53].

At a more sophisticated level, a function is sold: assets are leased, and data flowing from the asset is used to underpin the leasing contract. For example, the MoD now leases helicopters for the Royal Navy fleet, and the contract contains predefined hours of operation. For its drivetrains, Siemens sells the production of torque on a consistent and measureable basis. This has the added benefit of ensuring that customer loyalty is built into the development and supply of torque.

A service model for high-value assets

Rolls-Royce works across many sectors, including civil and defence aerospace, marine and nuclear sectors. Data services provide support to the after-market business, both to Rolls-Royce directly and Rolls-Royce customers which include more than 300 airlines globally. Services include ensuring reliability of the equipment, predicting and planning maintenance, and also long-term business forecasting. In particular, monitoring and predicting equipment failure is possible by means of small sets of data from the field, so that disruptive situations can be avoided.

The business model incentivises Rolls-Royce to have more reliable engines. In the last decade, reliability across the fleet has increased by 73%. Data from servicing and operations can be used to design better engines. Alongside this, computational fluid dynamics (CFD) data has allowed greater precision in design, although processing huge quantities of CFD data remains a challenge.

BAE Systems provides military aircraft. Data analytics is being developed to support both diagnostic and prognostic activities. Providing data to the supply chain to inform maintenance and support is key, as is making sure that the various platforms are in the appropriate state of readiness, such as 'combat ready'. Rather than carrying out scheduled repairs, it is possible to do targeted repairs on aircraft, combat systems and ship management systems. This is of particular value if replacement of a component or subsystem has a long lead time.

A business model around the system, not just the individual asset

Caterpillar has a long history of monitoring construction equipment, which includes such parameters as GPS position, engine temperature and revolutions. Information about the performance of the individual asset allows the company to improve the design of its products. The company has a servitisation model in which, for example, spare parts can be ordered to anticipate demand.

Caterpillar has developed a business model around a whole system, rather than an individual asset: so, for example, the company is able to advise quarry operators on optimising the productivity of a quarry. From information about GPS position and weight of trucks in use at the quarry, it is possible to identify when a truck is full but not moving, and therefore when there is a loss of production time. If Caterpillar does not own all the equipment, data needs to be shared across the different providers of equipment.

3.2 Built environment

Big data has the potential to yield knowledge about the built environment and its users that can be used to improve the design and performance of buildings. This will lead to greater sustainability, better energy efficiency and more precisely engineered systems, more efficient use of space, a higher quality environment and increased satisfaction of occupants[54].

New business opportunities and models are emerging in areas such as:

- linking energy performance during life to the original construction contract

- building management systems that offer occupants options for the reduction of energy usage

- greater use of predictive asset maintenance.

For example, Demand Logic[55] has developed a web-based tool that uses data generated by building management systems to find energy savings and performance improvements in commercial buildings.

Data analytics can help to improve the cost-effectiveness and risk management of construction projects. For example, during the construction of Crossrail, existing buildings and structures were instrumented during tunnelling works to monitor risk of damage (see example in *Case studies: where are the success stories?*).

The ability to share data between the different built environment stakeholders, such as building operators, owners, tenants, designers and contractors already exists within Level 2 Building Information Modelling (BIM). The UK will be the first to link design and construction standards with through-life operation, maintenance and facility management standards with Level 3 BIM. With the ubiquity of sensors being embedded in the built environment, the quantity of data available will surpass

that contained within current BIM models. Innovative ways of extracting knowledge from data, and making use of it at all stages of the lifecycle to understand and monitor design, will be needed.

At city scale, big data generated from both embedded sensors and social media, combined with enabling technologies such as the internet and wireless networks, supports the development of so-called smart cities. These will help to facilitate the participation of people in community life, as well as enhancing the performance of services and reducing costs and resource consumption. Intersectorial sharing of data is an important component of smart cities, extending across the built environment, energy, transport, digital communications and healthcare sectors. In combination with demographic data and weather data, for example, data from these sectors can support decision-making for policymakers[56] and this will be enhanced with extended BIM systems[57]. Big data will inform the procurement of built environment assets: data feeds from the city will drive what performance is required from the asset. Autonomous technologies will be used to enhance the capabilities of smart cities, for example, in inspection and maintenance systems, integrated personal transport and assistive living[58].

The export of construction services worldwide, including engineering consultancy and architecture, relies on the UK maintaining its position as a construction leader in key growth areas[59]; and this must include big data and associated technologies. For example, the UK has the world's leading set of BIM standards, driven by government with the support of industry. These are now being promoted as the international standards for the construction sector. Successfully addressing the barriers will help the growth areas to flourish. Examples of barriers include cybersecurity risks that necessitate the development of mechanisms to collect, analyse, share and store safely and securely[60]; standards for ensuring interoperability between data, devices and systems; and the difficulty of designing building-integrated systems that have shorter lifetimes than the fabric of the built environment.

"Data allows us to solve difficult problems and gain new insights in ways that we could not do before."

Duncan Price, Director of Sustainability, Buro Happold

3.3 Energy

The central or local generators, pipes, wires, and energy centres will, over time, be viewed as an energy system that delivers services to its end-users, rather than considered as separate electricity, gas, or heat networks. Data will therefore contribute to achieving convergence across different energy networks, allowing greater interoperability and flexibility in the future. Data will also be needed to understand and manage the operation of new types of energy generation, including community energy systems and heat networks. New types of data will be needed for local strategic energy system planning; for example, data on building typologies and construction types, spatial layouts, tenure type, and potential energy resources, as well as on the existing energy infrastructure. A whole-systems approach, including the ability to exchange data, is key to achieving interoperability in an increasingly dynamic and complex sector[61].

While existing data sources have supported the traditional incremental development of infrastructure, there is likely to be a much greater dependence on the integration of external data, such as market growth, to support the design processes needed for such a significant system transition. There are additional ways in which data sharing between the energy sector and other sectors can bring benefits, such as with local authorities to provide indicators on fuel poverty, with manufacturers of electrical appliances to improve energy efficiencies and with city planners for the purpose of designing new energy infrastructure. Data is also being used by third parties, benefiting consumers in new ways. For example, Open Utility is currently developing an online market for renewable energy[62].

The successful operation of the smart grid in the future will be underpinned by the availability of real-time data, as the grid moves from one-way to two-way flow, and needs to be capable of more integrated and complex transmission, distribution and trading. More data will be needed to optimise decision-making for operation and investment planning. In particular, systems will be required that integrate processes to enable long-term planning, day/hour ahead forecasting and immediate operational management. Data will also be used to set tariffs and manage system carbon intensity so that there are incentives for peak demands to be reduced, since providing marginal generating capacity for the highest load remains very expensive. Currently, the electricity network has high availability because of asset redundancies, but in the future, with greater inclusion of renewable energy, providing sufficient asset redundancy will be untenable or costly. The system will need to use assets and control demand more intelligently while ensuring that the safety and resilience of the infrastructure is maintained. Probabilistic techniques are likely to become increasingly important and these will be underpinned by sound data.

As well as centralised cloud services, distributed intelligence will play a role. For example, a micro-energy management system serving a community of devices could work hand-in-hand with the higher analytics engines – filtering and feeding data on current conditions, and using external data to assist its own decisions.

The UK government has set a deadline for gas and energy suppliers to install smart meters in every home and business by 2020, a first step in achieving a smart grid. Smart meters will play an important role in generating data, to inform the planning and operation of the grid, but also in understanding and empowering consumers to reduce energy use. Consumer behaviour in this area is poorly understood and may exhibit counterintuitive features[63]. There is considerable activity being undertaken at both national and international level around the development and roll out of standards for smart meters[64] [65].

"The difficulty lies in the very asymmetric nature of the transaction, in that the user of the data usually has a much better idea of what its potential value might be than the consumer has".

Erwin Frank-Schultz, Industry Technical Leader, IBM UK Ltd

There is a perception that the public's privacy will be compromised by the installation of smart meters, as a result of the possibility of interpreting data on energy usage to obtain insights into personal aspects of a household. The energy sector will need to address how it demonstrates the benefits of smart meters both for individuals and to society, and to ensure that appropriate technical solutions are in place to address privacy issues. The question of public trust is critical and is being addressed in a number of ways[66].

The oil industry already makes extensive use of data and analytical techniques to improve the productivity of assets. For example, detailed subsurface seismic information helps to achieve enhanced oil recovery; sophisticated prediction techniques are used to deal with the challenge of uncertainty and drive predictive maintenance scheduling. Data analytics is used to improve safety and equipment reliability, improve the process of refining and manufacturing oil-based products and understand the performance of fuels.

© efergy.com

3.4 Transport

Greater use of data analytics has the potential to support more efficient and smarter transport of people and goods through an integrated system that can optimise capacity, make better use of resources and provide a more customer-focused experience. Data can support both the planning and operation of the system, and make it more responsive. Data can be used to create better-integrated transport models that give insight into how the system is working, and can also be used as predictive tools to predict what might happen in different scenarios or extreme events, and to plan accordingly.

A transport system with consumers' needs as the central consideration will require an integrated approach. Spreading peak demand has major benefits for both users and providers. Transport providers can gain deeper insight into passenger experience, alongside more conventional performance measures, and improve their services accordingly. Passengers are already empowered to make transport choices by being provided with the necessary information about timetables and updates on delays through text messages and apps. For example, Citymapper[67] uses real-time data on a range of transport modes to help users plan trips. INRIX[68] provides information including parking and weather conditions to drivers, as well as providing real-time and predictive information on traffic which can help with planning new roads and transport systems.

'Connected cars' will contribute to improved road safety and more effective vehicle maintenance, and allow drivers to plan journeys better. They also have the potential to provide better transport opportunities for an ageing population or for individuals with disabilities. These would bring together the transport, built environment, insurance and digital communications sectors, and rely on effective sharing of data across sectors.

New business models are possible, in which, for example, vehicle pools accessed on an as-required basis will gradually displace private car ownership, principally in urban areas. Vehicle tracking services also allow businesses to optimise fuel consumption and reduce running costs for their fleets[69]. Business models designed around predictive maintenance are also emerging across all modes of transport, requiring integration across the supply chain to ensure spare parts are located and maintenance services are delivered promptly.

Rail freight also stands to improve planning and operational efficiencies. Information about the system is already connected with the goods supply chain, allowing the exact location of goods at any point in time in the system to be established. Better data about assets, including how they are being used, will contribute to the ability to predict their performance and improve efficiencies.

To optimise the potential contribution of big data and data analytics to the creation of an integrated transport system, the different transport modes will need to work together to find ways of sharing data using standards that are applicable to all modes. However, the different operators do not have a vested interest in providing efficient interchanges and ultimately in achieving interoperability, and will require incentives to do so. As with other sectors, the integration of different systems will result in a complex system that is vulnerable to cascade failure and therefore requires an additional degree of assurance.

"If your predictive maintenance systems on your train are telling you that the traction package is failing, it would be nice if your remote condition monitoring system could contact your suppliers' enterprise resource plan system and book maintenance under the service agreement."

Bryan Donnelly, National Fleet Performance Manager, Association of Train Operating Companies

3.5 Health

The use of big data and data analytics in the healthcare sector is driven by the need to improve the quality of care and reduce the cost of delivering care, supporting an ageing population within constrained funding. The planning and operation of healthcare delivery stands to benefit, as does improving wellbeing and preventing illness through public health interventions. The conversion of NHS patient records to electronic form yields data that, subject to patient informed consent, can be used in combination with other datasets to support prevention and treatment, as well as aiding research[70] to develop new medicines and treatments. The combination of genomic data with electronic health record data is also a possibility, subject again to appropriate personal privacy safeguards[71]. The National Information Board has published a framework[72], and more recently a series of roadmaps, on using data and technology to improve health and care[73].

Telehealth can make major savings possible in the treatment of common illnesses that often require hospitalisation. New technologies that use data to aid assisted living are emerging, although the NHS has not put in place the systems to deploy them. Patients are able to play a bigger part in managing their health; for example, patient-generated sources of data can be used to help manage long-term health conditions, by allowing patients to track and monitor their health and choose to share their data with their doctors[74][75] or to achieve a remote medical diagnosis[76]. The increasing prevalence of sensors embedded in wearable technologies such as Fitbits[77] and associated apps encourage wellbeing through healthy lifestyles, as well as providing monitoring data to caregivers. Individuals can buy analyses of their own genetic information that allow them to understand their risk factors and make lifestyle choices accordingly[78].

More effective treatment can be achieved through 'personalised medicine', which brings together data on genetics with environmental factors to tailor treatments to individuals, and alongside monitoring, allows care providers to assess how patients respond to treatment. New sources of data generated by smartphones, fitness monitors and wearables, and patient-controlled medical devices help researchers identify behavioural and broader social and economic factors that have the greatest influence on health outcomes[79], and thus help shape future health policies. The effectiveness of drugs can be assessed by remote monitoring of recipients of the drugs, providing an additional approach to testing for safety and efficacy, and informing value-based pricing of drugs.

The importance of demonstrating to the public the benefits of sharing personal data, and of putting in place a robust system of informed consent, is particularly critical for the health sector. Where effective anonymisation is not possible, the government and the NHS need to take a patient-centric approach to explain the society-wide benefits openly and honestly alongside the risks to individual privacy. One possible option for reducing the risk of loss of privacy is for consent to expire after a period of time has elapsed, at which point data will be withdrawn, or for individuals to be able to withdraw their data at will. Mechanisms for compensation could be made available in case data is leaked or misused. Standards also help to build trust around the use of data in health, such as the

"We have a phenomenal data machine learning environment in the UK; we have a phenomenal digital environment in the UK and we have a highly respected health service in the UK. Wouldn't it be great if we could bring that all that together to take innovation forwards?"

Julie Bretland, Founder, Our Mobile Health

recently published standard for developers of health and wellness apps[80], and other work in development[81]. Additional ways of balancing social and individual needs may be possible.

In order to realise the potential benefits of big data by creating interoperable systems, there is a need to ensure joined-up working between different stakeholders such as hospitals, GPs and patients. The existence of proprietary software in hospitals often precludes the possibility of their systems being connected with others. Another consideration is the usability in electronic health records of data generated by health apps, as well as the quality of the apps themselves. Early work around the quality of apps is going on, but more investment is needed. Notwithstanding, the existence of the National Health Service provides a platform in which innovations can be validated and subsequently applied in other situations.

TO CREATE INTEROPERABLE SYSTEMS, JOINED UP WORKING IS NEEDED BETWEEN DIFFERENT STAKEHOLDERS SUCH AS HOSPITALS, GP'S AND PATIENTS

3.6 Aerospace and defence

The aerospace and defence sectors represent a specialised subset of advanced manufacturing, characterised by high-value assets. It is therefore of particular value to the sector to create data services that support the after-market business. Indeed, innovations in servitisation models were initially developed within this sector. Data can be used to ensure the reliability of the equipment, help in predicting and planning maintenance, and in long-term business forecasting. Data from servicing and operations can also be used to design better components and subsystems, such as engines. As well as individual components, a servitisation model can extend to a whole capability such as the complete aircraft, and in a defence application, can monitor whether the capability is 'combat ready'. The defence supply chain is particularly sensitive to the challenges of data and system security.

Other applications of data analytics within aerospace and defence are well developed. For example, during flight testing, data from an aircraft is downloaded via telemetry during manoeuvres and analysed in real time.

In defence and national security, data analytics provides essential intelligence. As the range of types of data – both open and closed – increases, new styles of conflict are emerging, from state-on-state to identity-on-identity, for which data analysis and data exploitation will be part of the armoury. New techniques and technologies around data will be needed, such as methods of reducing data volume, deep machine learning and better human-computer interaction. Data and associated technologies will be integral to defence equipment too. For example, the Ministry of Defence recently launched its 'Future Soldier Vision'[82], which enables enhanced situational awareness through the integration of data connectivity and sensors into soldiers' clothes and equipment.

While the appropriate oversight mechanisms are still subject to debate, the value of connectivity and data analytics to the intelligence community is clear as detailed in the report *A Democratic Licence to Operate* by the Royal United Services Institute (RUSI).[83]

"We have retrospectively made modifications to the existing fleet based on service information and those then influence future designs."

Peter Chapman, Director, Rolls-Royce Controls and Data Services Limited

3.7 Insurance

Data is already used by the insurance sector to develop risk models and the increasing availability of data has the potential to improve estimates of uncertainty in risk ratings. In the consumer market, new sources of data are emerging, and companies are mashing together different types of data to get a more detailed profile of an individual. In future, the sector faces competition from companies, such as supermarkets and social media networks, that have access to large customer databases[84] and can provide insurance to consumers, although these will be constrained by data protection laws and the requirement for informed consent.

There is a data conundrum in this sector: insurance markets currently work on the basis that there is uncertainty in quantifying risk. If perfect data were available, low risk clients might realise that it is unnecessary to purchase insurance and the market would be left with only high risk clients whose premiums could become unaffordable. In an extreme case, the industry could decide not to provide cover in part of the market, rendering access to insurance unobtainable for some. This happened for cover for terrorism activity in London and was only resolved with the creation of Pool Re as a reinsurance vehicle[85].

In the case of motor insurance, telematics systems for road vehicles can be used to monitor driving performance, and are already being used by insurance companies to offer cover to drivers on a different basis to traditional motor cover. For example, detailed information on the circumstances around an accident can be obtained by combining telematics data with virtual mapping data[86]. Telematics can be used

to incentivise people to improve their behaviour, for example, by giving the driver feedback on their driving in real time. This will make insurance more affordable for drivers who would otherwise have to pay high premiums but could incentivise jamming and hacking by drivers whose telematics systems would reveal risky or illegal behaviour.

Advances in sensors used to monitor equipment or people in real time will benefit the insurance industry, for example, to forecast the availability and reliability of engineering equipment. This is of interest in regard to loss of production or other types of incident.

The insurance industry is faced with the challenge of pricing cyber risk for data-driven systems such as autonomous vehicles, as well as for individual companies. At present, it is hard to quantify the risks owing to the lack of historical data and rapid developments in the type of attacks. There are also systemic risks due to the interconnected nature of IT systems that are hard to quantify[87], although most incidents of cybercrime to date have been targeted at individual businesses. However, the demand for insurance in this area is growing, and products are beginning to appear on the market[88].

If better software engineering methods can be shown to reduce the risks of successful cyberattacks, insurance may become more affordable for companies that adopt such methods and, in a virtuous spiral, this may create a financial incentive for the take-up of improved software engineering; that is to say, insurance has the potential to be a valuable tool for enhancing the management of, and resilience to, cyber risk[89].

TELEMATICS SYSTEMS FOR ROAD VEHICLES CAN BE USED TO MONITOR DRIVING PERFORMANCE AND ARE ALREADY BEING USED BY INSURANCE COMPANIES

3.8 Case studies: where are the success stories?

The following provide examples where the use of big data and data analytics has brought benefits. These benefits are more broadly applicable across all sectors. The examples are based on detailed case studies that will be published separately from this report[90].

Transport – Modelling the UK's rail network

ORBIS is Network Rail's £330 million seven-year programme to create a detailed digital model of the UK's rail network in order to improve the efficiency, cost-effectiveness and safety of the organisation's asset management capability. Digital solutions are being developed to collect, join and exploit accurate asset data on the rail infrastructure. Geographical data allows an interoperable spatial model of the railway to be created containing information about how the assets are used, and their capability and performance. Successful implementation resulted partly from the involvement of on-the-ground maintenance staff in developing interfaces for collecting data using tablets and smartphones.

Transport – Operation of buses in Helsinki

A Helsinki-based bus company (Helsingin Bussiliikenne Oy) is using a data warehouse to combine data from its fleet of 400 buses, in order to improve cost-effectiveness, reduce carbon emissions and improve the service for users. The sensors monitor and analyse fuel usage and other data for each driver, route and vehicle, allowing mechanical problems to be identified, helping to improve individual drivers' performance and increasing understanding of why accidents have occurred. Other cities are interested in applying this to their own bus services.

Built environment – Construction of Crossrail

Data analytics were used to improve the performance of infrastructure delivery during the construction of Crossrail. Existing buildings were at risk of damage from settlement and ground movements during tunnelling works, so instrumentation was put in place to monitor this. The application of novel data analytics techniques to data obtained from monitoring equipment, based on a collaboration between engineers and specialist data scientists, helped project managers to manage risk and reduce costs. Part of the success lay in effective visualisation of the data in a dashboard, allowing project managers to interpret and act quickly on settlement data.

Healthcare – Improving end-of-life care

Vitas Innovative Healthcare, a US-based healthcare provider, is using mobile technology to improve end-of-life care for patients by increasing the effectiveness of caregivers and other staff. The technology facilitates better communication of information and allows caregivers to be routed more efficiently and quickly to patients based in their own homes or residential care facilities.

Healthcare – An app for treating burns injuries

The Mersey Burns Application (app) for use on smartphones was designed to improve the assessment and resuscitation of patients following a burn injury, before the patient reaches a specialist unit. The app allows clinicians to shade the burn pattern onto the screen and generate detailed fluids protocols as well as send an email to a receiving burns unit. The greatest challenge was to develop a robust process that satisfied the regulator, MHRA. The application is the first UK healthcare app carrying a CE mark from MHRA.

Energy – Optimising performance in the energy industry

A tool has been created by Wood Group Intetech that helps oil and gas operators optimise performance. A global database of well failure data has been assembled, overcoming operators' reluctance to share data by keeping sources confidential. The data allows operators to benchmark their performance with other operators, and identify where changes in operations or maintenance regimes are beneficial. This example demonstrates how sharing commercially sensitive data, albeit in an anonymous form, can bring benefits to the companies that participate.

Manufacturing – Monitoring the condition of industrial equipment

Finning is an international company that sells and rents Caterpillar equipment to a diverse range of industries and also provides customer support. It has developed a new service that allows customers to optimise their asset management through the proactive, holistic use of condition-based monitoring data, thus minimising customers' owning and operating costs. Factors such as operator performance, site conditions and equipment history are reviewed in addition to component health, and possible future problems flagged up. The development of a dashboard for presenting information aids its communication, along with a bespoke customer portal, MyFinning, where regular reports at asset, site and fleet level can be accessed.

Learning from other sectors – Performance development for Williams F1 racing cars

Big data collected from F1 cars is vital for maximising performance. Effective use of data goes beyond aerodynamics to design, manufacturing and race engineering. Performance development teams rely on real-time data from multiple data sources including sensors, video-feed, on-car telemetry and simulations. Stringent F1 regulations limit the time and resource that teams are allowed to use to develop the car, both on track and in the factory. This puts a premium on extracting the maximum amount of useful information from the minimum amount of testing, requiring carefully designed experiments and efficient processes. A consistent method is needed to evaluate whether a new part or change increases performance according to the agreed development direction.

Learning from other sectors – Weather forecasting by the Met Office

Data used in numerical prediction models is obtained from myriad sources, including observation stations, radiosondes, aircraft and satellites. The challenge is to bring data into prediction models that vary considerably in timeliness, granularity and quality, and to process and store huge volumes of data. Weather forecasting benefits greatly from the agreement between national meteorological services worldwide to exchange data freely between them. The usefulness of meteorological data is broadened by using it in combination with other data sources, for example, combining wind data with information about leaf drop to predict when 'leaves on the line' are going to be a problem for the rail operator.

Learning from other sectors – Improving productivity and safety in the mining industry

Big data is being used to increase safety and productivity in the mining industry. Rio Tinto collects data from sensors attached to fixed and mobile equipment, and applies advanced analytics techniques. The analysed data allows engineers to identify safety issues or predict and prevent operational events, such as engine breakdowns, before they occur. The capability, launched in March 2015, is staffed by 12 mineral experts who continuously examine processing data from five of the company's coal sites in Australia, and operations in Mongolia and the US. This information includes laboratory data, operator log sheets, control systems, process surveillance cameras, maintenance systems and data histories. On the basis of the data analytics, the company has already made control system and instrumentation improvements and enhancements to operator decision-support tools.

4. Engineering challenges around data, analytics and systems

This section outlines engineering challenges, many of which can be successfully addressed by applying best engineering practice in order to ensure that accepted methods are used, and that solutions are appropriate, cost-effective and well-documented, user-focused and compliant with legal and regulatory requirements. The central role that areas such as software development play in creating data-driven systems necessitates the extension of best engineering practice to these areas as well.

Good data governance is also essential in any data analytics project. It is key to ensuring that the data is of adequate quality, and it is possible to understand and integrate data. This is necessary to build trust in the data. It requires an information engineering approach.

"You have to do better engineering on the data."

Dr Martyn Thomas CBE FREng FIET, Director, Martyn Thomas Associates Ltd

"We simply do not have the metadata to allow us to understand what these measurements represent. They are just numbers."

Hugh Boyes FIET, Principal Fellow, Cyber Security Centre, WMG, University of Warwick

4.1 Engineering challenges around data and analytics

Ensuring good quality data and analytics

The level of data quality that is required depends on the intended use for the data. For example, a higher tolerance is acceptable in the quality of personal data used in market prospecting in comparison to the quality of patient data used in healthcare. The challenge of acquiring and preparing data for analysis is considerable, and typically accounts for two thirds of the effort in any analytics project.

Data quality is influenced by a number of factors that include bias, timeliness, granularity, the quality of metadata and the possibility of calibration error. There is also a threat of 'spoofing', whereby data is falsified intentionally without the knowledge of the data recipient. Any uncertainty about the data remains when datasets are combined for analysis. It is important to document how data has been collected to reduce this uncertainty.

In turn, data quality influences the quality of the analytics and the confidence to make decisions based on the outputs. The quality of the analytics is also influenced by the availability of data, the appropriateness of the data for the problem, the correctness of the algorithm and the placement of the execution of the model in the running systems, a software development issue. The quality of analytics development is maximised when developers work closely with those who will be using the software, are able to experiment with the real data

and evaluate the analytics throughout the process using a feedback loop of input and outcome data.

It is also important to document assumptions about how data has been analysed. For example, metadata should include information on how data is processed and interpreted: communicating the assumptions behind aggregation, abstraction, knowledge and hypothesis makes it possible to revisit data to discover things that were not noticed originally.

Effective visualisation can help data users extract useful information from data. For this, an understanding of the user and how they will make decisions based on the information is needed. The KPMG Data Observatory, launched at Imperial College London in November 2015, is an example of a sophisticated visualisation facility that will allow businesses to gain insight from complex datasets.

Dealing with volumes of data

One of the challenges of big data is how to process the large volumes of data being collected. Often, data is collected simply because it can be but without any experimental design or hypothesis that justifies the need to collect this specific data. Data has a timeframe within which its analysis has the highest value, and there is also a timeframe within which acting on the results of the analysis has the highest value. Outside of these windows, the value of the data can be negligible. If too much data has been collected, there is a risk that analysis cannot be done sufficiently rapidly (for instance, because there is insufficient bandwidth available for transmitting the data for remote processing). The engineering skill lies in judging the rules for selecting data according to its purpose and the timeliness requirement, in order to extract the maximum value. Data analysis techniques, such as machine learning, facilitate real-time or near-real-time analysis of large data volumes.

There is a trade-off between throwing away data, which risks losing valuable information, and keeping unmanageable quantities of data. If there is uncertainty about which data will be valuable in the future, it will be necessary to carry out the analytics earlier on to identify what is valuable, or to keep the data for a certain period of time so it is possible to revisit it. An example of this might be video analytics of CCTV footage for security purposes or the retention of mobile phone location and other data by phone companies. These problems remain even as the cost of data storage falls, for the reasons discussed in the section on *Preserving data for the future*.

Dealing with heterogeneous data

By combining both structured and unstructured datasets, it is possible to enrich the information that can be extracted from data. Both soft data from social media and hard data from physical systems will be increasingly available. Finding ways to combine such different types of data in meaningful ways is a challenge. Datasets may have varying amounts of noise and

"The volume of interesting, useful data is not increasing at the rate that the total volume is rising."

Dr David Watson FREng FIET, Director, IBM Research, UK

"Sooner or later we need to think whether we need to collect that data, how often we need to collect and what volume."

Dr Payam Barnaghi, Assistant Professor, University of Surrey

"The exciting science comes in because we need new techniques that can fuse genetic data with highly unstructured, noisy sensor data. This is a unique contribution that engineers can provide."

David Clifton FIET, Group Leader, Computational Health Informatics Laboratory, University of Oxford

levels of certainty. Different types of data will allow different questions to be asked. Understanding the nature of the questions that can be asked of different types of data is a first step; it will then be necessary to develop the tools required to answer them.

Preserving data for the future

By addressing data legacy, organisations will preserve the ability to analyse and interpret data. Although data stays static, the problems to which data is applied may change over time. Therefore appropriate metadata is needed that lasts over time so that information about the context and content of the data is preserved; or data migration is carried out to preserve data usability over time. The sensitivity of data also changes over time as it is linked and aggregated. The challenge is in judging what data is valuable and worth keeping. A robust security, classification and governance model is needed to optimise the effective retention of data and to ensure the appropriateness of its use at a future time.

It will be of benefit to demonstrate to individuals within organisations the value of data owned by the organisation and the benefits of preserving it. For example, Schlumberger is starting with small pilot groups to preserve research data for decades, with the intent that these groups will communicate the benefits of preserving

data to other groups. The scientists who generate the data are involved in classifying it to ensure it maintains its usefulness.

Future accessibility of data by computer applications also needs to be considered. The data should be migrated periodically to ensure that current applications can access it. A robust process is needed for identifying and destroying data that does not need to be held indefinitely.

4.2 Engineering challenges around systems

Considering system legacy

It becomes harder to maintain or upgrade a system the longer the system remains in service. If components are documented digitally, they can be made as and when needed rather than holding stockpiles. It will be important to ensure that these digital files can be accessed and processed in future decades, of course, just as this has to be true for all the important data that an organisation retains and continues to need. Migration to updated systems is preferable, alongside annual audits of databanks to ensure this happens, although this decision is driven also by cost.

"When you use complex communicating sets of machines, you don't know what that world looks like."

Denton Clutterbuck FIET, Director of Sales and Marketing, Altran UK

"It is no good coming along in 40 years' time and finding that you can't read the files."

Professor Ian Poll OBE FREng, Chief Executive, Poll Aerosciences Ltd
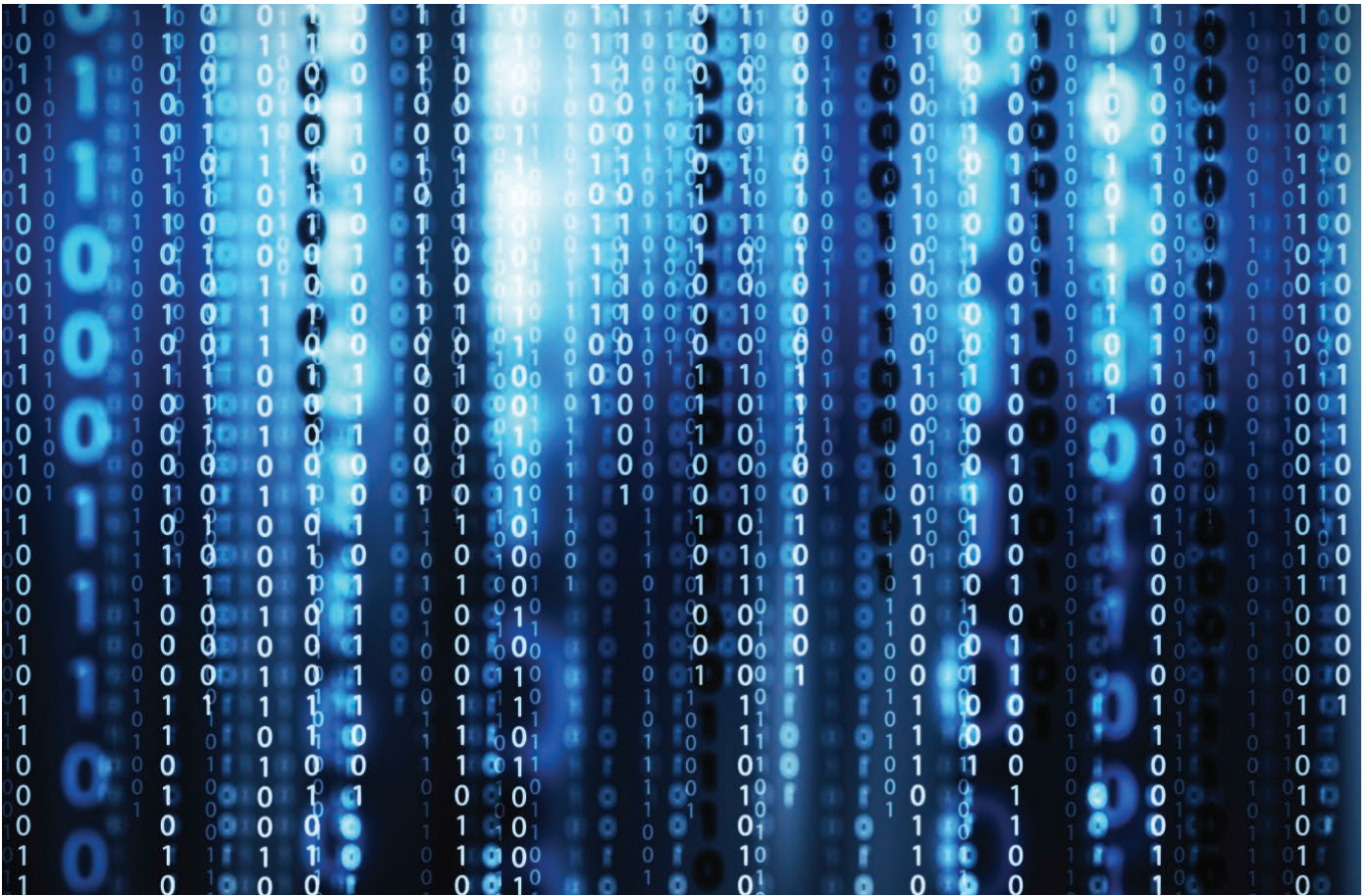
"Open machine learning systems will be put in place and things will arise from that which are totally unpredictable."

Professor David Delpy CBE FREng, Chair, Defence Scientific Advisory Council

"The failure of one system has much more potential to cascade through others in that very tightly integrated environment".

Professor Jim Norton FREng FIET, Past President BCS

Managing the uncertainty of complex systems

An important area emerging from discussions at the workshops is the integrity and assurance of data, devices and systems, and the implications of connecting these to the internet. Devices will be connected together in such a way that creates increasingly complex, integrated systems, either by design or accident. Emerging properties that are either desirable or undesirable may result. This is of particular concern for safety-critical applications: as society grows more dependent on new systems, it is an urgent priority to ensure that they are dependable.

It will be important to understand how to manage risks associated with uncertain data and the impact they might have on decision-making. Ways will need to be found to assure more complex systems, and methods to certify systems that are less predictable in their behaviour, such as neural nets[vi] that must decide from among a wide variety of responses. More integrated and complex systems will need to be analysed and tested to assess the security vulnerabilities and the probability of failure occurring. Interconnected systems should be designed to degrade gracefully when one or more related systems fail.

[vi] A neural net is a type of machine learning method and can be applied to a wide range of tasks that are hard to solve using ordinary rule-based programming.

5. Privacy and security: overcoming the key barriers

5.1 Privacy

The basis of relationships between individuals and organisations will need to become more defined as more information is produced and different types of organisation hold the information. Engineers will need to be familiar with these issues so that engineered systems provide appropriate solutions.

There are trade-offs to be made in letting others have access to personal data. The question needs to be asked, 'What you are prepared to give up in terms of privacy in order to gain a service?'. There is evidence that the public are willing to share personal data to get a better service in relation to transport[91], for example. It is necessary for organisations to demonstrate the benefits of opening up data to win public trust and it may be necessary to offer indemnities against damage or loss, however inadvertently caused. It is important for organisations to have clear definitions of 'opting in' to a particular system, and to communicate the implications to individuals. A change in context, such as a change of governance or a change in their personal circumstances that makes their earlier decision unsafe, may lead to an individual wishing to opt out, even if they trusted the system earlier. Consent is currently permanent and irrevocable, although new EU data protection law will give individuals the 'right to be forgotten' (see Section 5.3). Given the potential benefits to consumers of opening up data, regulators could be encouraged to participate in this debate more proactively.

IBM's paper on *Ethics for big data and analytics*[92] presents an ethical awareness framework, that sets the ethical position within the questions 'What can be done legally?', 'What can be done technically?' and 'What would an organisation like to do?'. Balancing the benefits between different stakeholders is a prerequisite of an ethical solution. There are potential ethical issues associated with combining different datasets, so that 'new' information about an individual is created. There are also questions around who owns this information and whether it is legal to process personal data that has been derived in this way.

Both large and small organisations have said that they are deterred from moving forward because of these ethical and legal uncertainties. More compliance toolkits may be needed to help organisations deal with such issues.

"Open data goes with some loss of privacy, even if it is anonymised'.

Professor Will Stewart FREng FIET, Chair of the Communications Policy Panel, IET

"When you opt in, are you signing that data away for life or are you saying 'You can keep it and use it for a year' or 'For these purposes, but at the end of that time you have to delete it?' Those questions ought to be asked explicitly".

Dr Martyn Thomas CBE FREng FIET, Director, Martyn Thomas Associates Ltd

5.2 Security

An engineering axiom for security professionals is that you cannot bolt 'security' onto a system design as an afterthought. It has to be designed in from the beginning and with mathematical rigour.

The word 'security' has broad connotations: it is both about guarding against possible malicious actions, and about assuring safety-critical systems. There is a risk that the individual is compromised by the failure of interacting systems. Lessons can be learned from public safety networks and from the military community. The security of wireless connected systems is an important area.

Security in data-driven systems is paramount, both to meet the law and also to protect critical infrastructure; however there are numerous products that are unfit-for-purpose. Often devices and systems have been designed with no forethought that they might one day be connected to the internet. An important consideration is

how data can be taken out of devices and systems safely and securely. As well as the systems themselves, datasets too can be hacked into and used for hostile purposes. For example, a drop in household power consumption over a period of time indicates that the householder is away from home, information that is of potential interest to criminals. Malicious access to engineering design data is also a risk. Organisations need data management and stewardship skills to maintain the privacy of their customers' data and ensure that systems are secure.

Both open source and proprietary software can play a role in developing high-quality software products. The paper *Security in Open versus Closed Systems*[93] makes clear that there are balancing benefits and that the argument for security supremacy of open versus proprietary systems is largely an irrelevance. A move though to building the base software for simple infrastructure command and control systems using efficient analytical methods and then making these available free of charge through the Open Source community was recommended in the IPPR report *Shared Responsibilities: A national security strategy for the UK*[94] published in 2009. This recommendation remains pertinent today.

As technology continues to improve, encryption that is adequate at the current time is very unlikely to be adequate for the whole lifetime of a system. It is important to understand the security risks of a product over its design lifetime, especially for major assets that must remain in service for many years to repay their investment.

Box 2 provides a series of questions that relate to the integrity and assurance of data and data systems and their connectivity, and can provide a helpful framework when designing a system comprising connected data and devices.

"We need to be addressing what is the mechanism which allows us to take the data off those sensors, use it for the purpose it was collected and also share it, where appropriate, in a safe and secure fashion."

Hugh Boyes FIET, Principal Fellow, Cyber Security Centre, WMG, University of Warwick

"As technology continues to improve, what is adequate encryption today is very unlikely to be adequate encryption in the lifetime of the satellite or whatever you have."

Nic Holt FREng, Distinguished Engineer and Systems Architect, Fujitsu

BOX 2: FRAMEWORK FOR THE DESIGN OF SYSTEMS COMPRISING CONNECTED DATA AND DEVICES

1. Who could be at what risk if these data is leaked or compromised?

2. Should the system be connected to a network or other systems at all?

3. Should any connection be one-way only (so that this system cannot be controlled across the network)?

4. Should the device be protected? If so, what are the threats against which it should be protected?

5. Should the data be protected? At rest? In transit across the network?

6. How important is data availability?

7. How important is data integrity?

8. How important is data confidentiality?

9. How confident do I need to be that the data is protected against these risks?

10. How will I achieve that degree of confidence?

11. How will I demonstrate the confidence level that has been achieved?

12. What systems are we creating by connecting two or more devices?

13. Are these systems that have been designed as systems, or are they accidental?

14. What integrity properties do I need these systems to have, and how will I assure myself that they have them?

15. Who carries the risk if something goes wrong?

16. Who carries the liability?

5.3 Legislation around data and software

Technologies and processes around big data, data analytics and advanced connectivity are developing quickly, and the legal and regulatory environment has not always kept up. It is necessary to question whether traditional legal and regulatory principles still work, and how they need to change.

Data

Stricter EU data protection laws[95][96] are currently being developed that are expected to come into force in the next few years. These laws support individuals' fundamental right to data protection. They were initiated in response to the increasingly large quantities of personal data generated and the globalised nature of data flows as a result of cloud computing. Cloud computing allows data to be stored in a remote location that is in a different jurisdiction from the individual associated with that data. In response, a single set of rules will be in force across the EU, and clarity on when EU law applies to data controllers outside the EU. The benefits to data protection of restricting cloud services to national or regional boundaries is also a matter of debate[97]. The 'safe harbour' rules that allow US companies to use a single standard for consumer privacy and data storage in both the US and Europe were deemed invalid by the European Court of Justice in October 2015, complicating the regulatory environment.

Another principle in the developing EU legislation requires that data protection safeguards are built into products and services from the earliest stage of development. The requirements for informed consent will be also stricter. Tighter legislation is being prepared on private companies giving data to third parties and tougher punishments put in place if data protection breaches are discovered. The new legislation also

establishes the right for individuals to be forgotten, which means that they can request that their data is deleted. The UK's data protection agency, the Information Commissioner's Office (ICO), already provides guidance[98] to companies and other organisations, including codes of practice that will help them comply with the new legislation.

Security in software

Another issue that concerns both companies and consumers is the threat of breaches of cybersecurity. Many classes of goods already have strict safety requirements. In the modern world, security vulnerabilities may often contribute to safety hazards, because malware and hacking are now so prevalent that everyone is potentially at risk even if there is no reason for them to be targeted individually. It is important to ensure that security is adequate in these systems since the protection of critical infrastructure (transportation, energy, healthcare) is crucial to society. It is therefore necessary to treat cyber threats as part of the normal environment for the purposes of assessing acceptable quality. Since this a problem of the modern information society, innovative legal solutions have to be sought.

The reasoning behind the wish to explore new approaches is that existing law, both general principles as well as statutory

regimes, does not adequately incentivise software and system providers to ensure that their products are always fit-for-purpose, especially in terms of security functionality. However, UK legislation is based on the principle that those who create a hazard have a duty to manage the risks[99]. This shows that accountability is a fundamental principle in the UK legislation which has to be translated into the connected world. The Consumer Rights Act 2015, for example, has already extended consumer protection to digital data[100]. With regard to cybersecurity, one avenue to tackle this problem might be to put laws into force that establish accountability for software vendors. Enhancing liability for faults should lead to an improvement in quality, but it may have an adverse impact on innovation and the speed of development cycles.

The need for legislative changes to improve software quality requires further debate. If changes are pursued, during the 'preparation period' of the legislation there should be a public consultation on the classes of software that should be covered urgently and on which, if any, classes of software should be exempt. Furthermore a state-of-the-art defence could be included, allowing a trader to avoid liability if the state of the art was such that the particular defect could not have been detected at the time the product was put on the market.

SECURITY VULNERABILITIES MAY OFTEN CONTRIBUTE TO SAFETY HAZARDS, BECAUSE MALWARE AND HACKING ARE NOW SO PREVALENT THAT EVERYONE IS POTENTIALLY AT RISK

6. Connectivity

Broadband access and ubiquitous access to high-speed broadband services are necessary to ensure that connectivity benefits all sections of business and society, including public and private organisations, government and consumers. This is of particular importance in a data-enabled economy, where the transfer of data in large volumes and at high speeds, or the need for resilient transfer of data in urban and rural areas, are vital components of many data-driven systems. Furthermore, consumers expect a level of connectivity that meets their needs.

Sectors such as advanced manufacturing and precision agriculture, despite making a significant contribution to the UK's GDP, are at a significant disadvantage as a result of their rural locations. Nesta recently highlighted the increasingly damaging impact of insufficient broadband access to precision agriculture[101].

6.1 What level of connectivity is needed?

The requirements for connectivity will vary according to the particular application, and organisations will need to identify where systems are critical for business purposes or for safety. For example, telecare systems used to monitor patients in their own homes rely on robust connectivity to ensure that any adverse event needing an immediate response is rapidly communicated. The nature of data generated and the analytical techniques being applied to the data will also influence connectivity requirements. For example, in the case of distributed systems that comprise a number of components carrying out computations between which messages are passed, 'latency'[vi] is a problem for some kinds of data analytics, including graph analytics.

6.2 The decision to process locally or centrally

Connectivity affects the decision about whether to process data locally, or transfer the data to a central location for processing. This decision is also influenced by how essential or timely the data analysis needs to be. If transferring the data to a central location is too slow, the information can become outdated. The decision to centralise or decentralise processing of data is a trade-off between more efficient processing in the centre versus requiring more bandwidth. For example, to keep peak electricity load at a minimum, it is more efficient if energy-using devices within a building can communicate with each other and adjust their usage accordingly, rather

"Lots of connectivity is being put in homes, but not in the industrial environment".

Irene Lopez de Vallejo, Head of EU Partnerships, Connected Digital Economy Catapult

"Make sure all smart cities, all smart homes have connectivity and data designed in."

Dr Mike Short CBE FREng FIET, Vice President R&D, Telefonica

---

vi 'Latency' is the time delay in a system

than sending information backwards and forwards between the building and a central location. Security concerns also influence the decision about whether to store and process data locally rather than centrally. This is a consideration in deciding whether health data is stored locally with a GP or in a centralised storage location. 'Distributed' analytics enables local analysis of data, alongside the exchange of information from a centralised system. This can help in dealing with reducing the data volumes converging at a central location, and will be used for applications such as connected cars and local energy systems.

"If broadband is not resilient in that area, then your healthcare system also fails".

Cameron Steel FIET, Director, BK Design Associates UK Ltd

"Connectivity is the big issue, in terms of the bandwidth, the timeliness and the resilience, because you need it most when things are going wrong".

Adrian Grilli, Managing Director, JRC Ltd

6.3 Resilience of infrastructure

The resilience of broadband infrastructure in the UK to support the proliferation of internet-connected devices is a concern. Broadband systems should be able to provide sufficient data rates at all times to support the increasing number of users and devices that are connected to the internet, whatever the overall level of usage. Routers and other communications hubs are critical components in many networks as their failure or disruption can compromise the availability or integrity of the entire network which will prevent the organisation from operating efficiently or at all. Too many failures have already occurred because communications hubs were a single point of failure and vulnerable to floods, fires, disruption during building works or cyberattacks.

7. Creating the right conditions for innovation

7.1 Open data as a driver of innovation

In an ideal world, data should be widely available and accessible, with the business value created downstream, through innovation and unlocking of markets. A switchover to open data is needed, where every sector is considering what data it holds and in what circumstances it can be released. Regulatory issues, commercial constraints and cultural attitudes, for example, affect willingness and ability to share data. It is not always appropriate for data to be free and publicly available; there may be a need for access controls appropriate to its sensitivity and use. Leadership is needed to address the evolving intellectual property and legal constraints, which vary across different countries.

The Open Data Institute suggests that a 'data infrastructure' is needed to create the right conditions for data accessibility[102], that is informed by considerations such as what data can reasonably be traded, and what should be a public good and openly available. The Open Data Institute classifies three types of data – open, shared and closed – according to the nature of the permissions for use and publication.

Opening up data is not without its risks: any potential threat to assets and individuals needs to be addressed, alongside the benefits. And even anonymised data can, in certain circumstances, be manipulated and lead to loss of privacy for individuals. It is also important to recognise that opening up datasets does not itself create value, but that the ability to link datasets reliability generates value.

"Open data is a fundamental driver to the business models".

Ruth Boumphrey, Head of Research Grants, Lloyd's Register Foundation

"Unless there is a switchover programme towards open data I do not think we can pass go as a digital economy."

Dr Mike Short CBE FREng FIET, Vice President R&D, Telefonica

7.2 Demonstrating the benefits of sharing data

Much engineering performance data in sectors such as transport and energy is held behind firewalls and is proprietary and commercially valuable. However, in some situations it is beneficial to the parties involved to share data, and these benefits need to be publicised. Sharing data requires organisations to work collaboratively. It is also only useful if the shared datasets can be linked together. Investment will be needed and the means found to overcome the perceived risks of sharing data, but returns could be considerable. For example, a scheme that enables oil and gas companies to share data with a neutral party but keep data confidential allows the companies to benchmark their performance with other operators and identify where investment in de-bottlenecking or

improving operational performance can improve relative performance (Section 3.8).

In some circumstances, there are legal and contractual barriers, as well as technical barriers, to sharing data. Systems may have been set up specifically to prevent data from being shared. Westminster City Council, which is involved in smart city initiatives, has now changed the terms of its contracts to permit data sharing, and other local authorities are considering this.

For some applications, the value of sharing data is clear, as long as the problems of data anonymity and data security can addressed by filtering data to preserve anonymity, or to share only some parts of the data.

7.3 Creating markets in data

While the opening up of data by government has contributed greatly to the development of a data-enabled economy, much potentially valuable proprietary data held by corporate bodies is still inaccessible. Access to this data will not necessarily be without cost. The ability to trade data in 'data markets' under rigorously specified conditions is therefore a necessary condition for gaining access. UK organisations such as Mydex[103] are already pioneering the role of third party trusted data management in areas such as personal data, which are more challenging than, for example, trading engineering data.

A vital component of creating markets in data will be the formation of trading hubs and exchanges where controlled access can be bought and sold and datasets appropriately curated. Good

"How do we encourage investment and innovation when the data is the key currency?"

Dr Mike Short CBE FREng FIET, Vice President R&D, Telefonica

"Open data is often misinterpreted as free data and publically available data. You can have open data, but with the right access controls, with trust and privacy."

Dr Payam Barnaghi, Assistant Professor, University of Surrey

examples already exist within the world of e-commerce such as in the civil aircraft industry (Aeroxchange[104]) where there are intermediaries that track the location and availability of controlled or high value spare parts or facilitate the exchange of information up and down supply chains (Covisint[105]). Methods for valuing data, discussed in Section 2.3, will be needed.

There is potential too for individual consumers to trade their data and benefit directly. For example, a technical and market platform, the Hub-of-All-Things (HAT), is being developed for personal data owned by individuals[106].

7.4 Standards: achieving interoperability within and between sectors

Standards can help to drive innovation; a lack of standards can conversely inhibit it. They will also help to create a common language to aid multidisciplinary teams. A national strategy for the use of standards in data analytics, similar to that being developed by Germany for Industrie 4.0[107], would help to accelerate innovation. Standards that are applicable across different sectors are needed in many areas including data concept models[108], data taxonomies, data integration, data privacy, security and accessibility. CRISP-DM[109] is an example of a standard process model for data mining. ICT standards addressing technological interoperability are also needed. Government are implementing open standards principles to provide a level playing field for open source and proprietary software providers competing for government IT contracts[110].

Standards development must be market-led and vendors will want to engage in the process to ensure their features and requirements are included in the standard that appears. The rate of evolution of any industry standard should depend on market feedback, and standards can be developed to be updated as necessary to meet user requirements. Many organisations will need to develop a strategy for managing the rate of change, participating in and tracking emerging standards. Standards are key to achieving interoperability. In the absence of interoperability, there will be a proliferation

ENGINEERING DATA
SHOULD INCLUDE
DETAILED METADATA
DESCRIBING ITS
CONTEXT. THIS WILL
ALLOW DATASETS
TO BE USED WELL
BEYOND THE
APPLICATIONS FOR
WHICH THEY WERE
COLLECTED

of bespoke systems that will allow vendors to lock in customers and act as a barrier to new entrants. As the Internet of Things takes off, new networks and standards are appearing, so it is currently difficult to judge what connectivity standard should be selected for Internet of Things devices. Expensive upgrades may be needed in the future, if the wrong decision about which standard to use is made now.

Standards for data will allow its meaning and context to be preserved. This is particularly an issue when data is traded between sectors, where there may not be a shared understanding of the particular context in which the data has originated. It will help to provide an understanding of the different types of data, their levels of uncertainty and how they have been calibrated. For example, BSI[111] is currently developing Publicly Available Specification (PAS) standards for smart cities that will help to address this[112 113]. In addition to technical and interoperability standards, non-technical standards for the use of data by city authorities have been published and are being developed for citizen engagement, vital if trust is to be developed in the use of the data[114].

In emerging fields of strategic importance to the UK, it is vital to secure first mover advantage. This is best achieved through a coordinated effort to define and establish the relevant international standards committees, as Germany is in the process of doing for the new industry standards that will underpin Industrie 4.0. The UK must maintain a leadership role in developing international standards so that UK companies do not lose out against international competitors. As the UK member of ISO and IEC, BSI is promoting smart city standards into the international domain.

Standards are usually created at industry or national level; one difficulty is that SMEs working individually are too small to drive the standard for a new product and sometimes do not have the resources to participate in international standardisation

workshops. It is important to raise awareness among research, development and innovation funding bodies and private investors of the value of this activity for accelerating routes to market. It would be helpful if university researchers, who often have relevant expertise, are provided with the necessary funding or career incentives to participate in international standardisation activities.

A key conclusion across all of the stakeholder workshops was the need to ensure that engineering data should include detailed metadata specifying limits of uncertainty, calibration applied, and other aspects of describing its context. This will allow datasets to be curated and used potentially well beyond the applications and sectors for which they were originally collected. Defined standards remain a crucial facilitator of this work.

7.5 Ownership of data and intellectual property rights

An important area of legislation relates to ownership of data. Ownership of rights in data allows businesses to prevent others copying it, or permits them to charge revenue for its access[115]. It is often difficult to identify where the ownership of data that has been generated in new ways lies, as data is often collected and analysed as a collaborative effort. For example, it may be unclear whether owners of connected cars, or the manufacturers, are the owners of data generated by the vehicles. Similarly, there is a question over who owns energy data used to monitor the performance of buildings[116]. For any sector, the ownership of data and the freedom to use the data are critical components in protecting it as an asset and realising its value[117].

Organisations will also need to ensure that intellectual property issues are considered in relation to new tools and techniques that they develop.
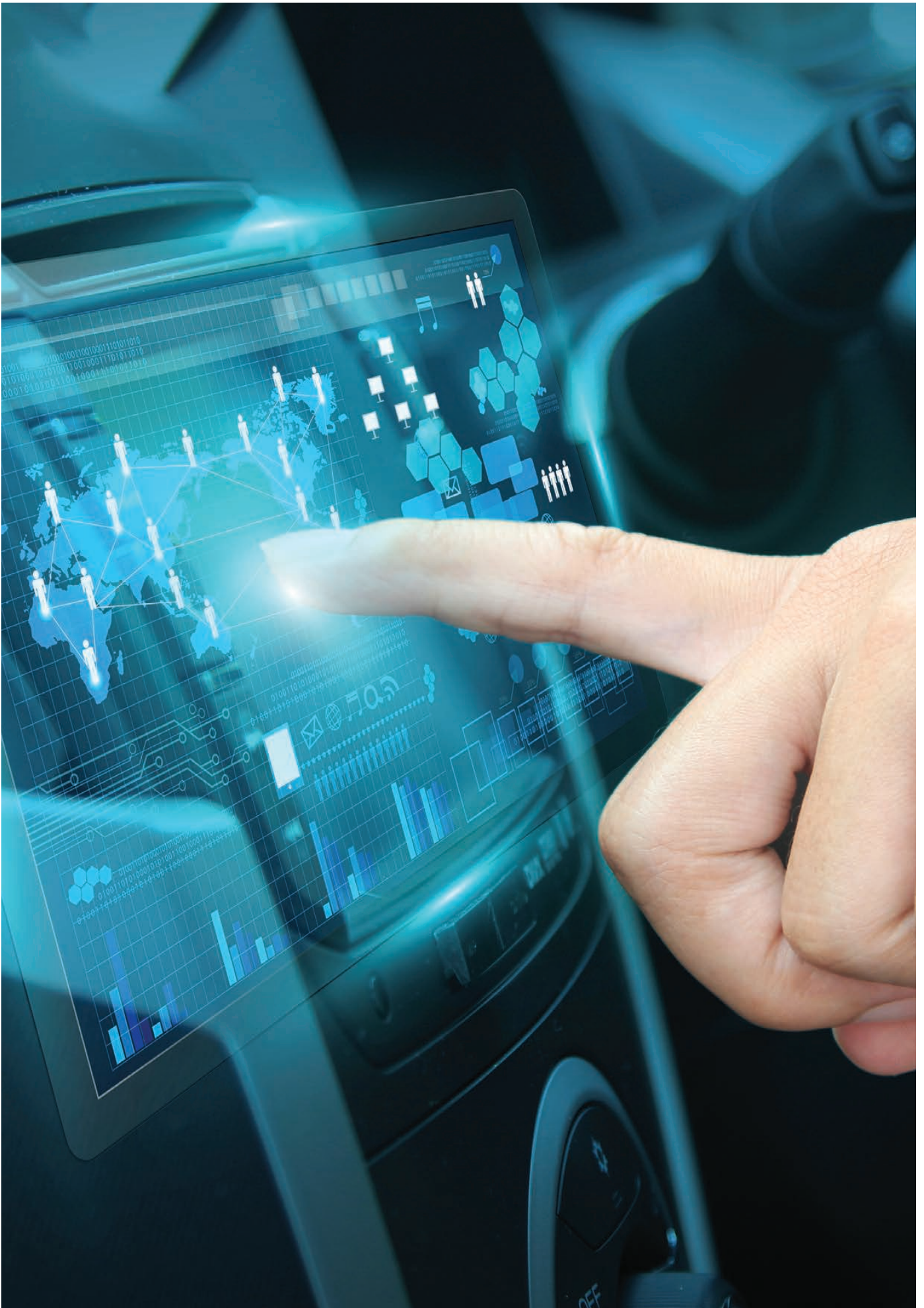
8. The challenges for organisations

8.1 Creating new business models

Existing organisations and new businesses face a major challenge in creating business models around data. There is potential to learn from other sectors that have already done this, such as aerospace or retail. Business models must reflect the need for dynamic data to be regularly updated and maintained. Servitisation models are one example. These were initially developed in the aerospace industry and now have increasingly broad application as a result of better data analytics and improved connectivity; they are set out in more detail in Box 1.

"The business case may not be clear from the beginning, but if people start to play around with new technology, they will find lots of ways in which they can improve their part of the business."

Steve Pryce, Head of Marketing and Wagon Management, DB Schenker Rail UK

"We should challenge some of the business processes that have been put in place before the era of big data just to see whether they are still the right processes given big data".

Dr Martyn Thomas, CBE FREng FIET, Director, Martyn Thomas Associates Ltd

New business models could also emerge for products and services as a result of collaboration between sectors traditionally separated within vertical silos[118]. For example, the satellite industry can provide earth observation data to new markets within the agriculture and natural resource sectors[119].

Business models are being developed that are centred on creating benefits for the consumer, although benefits could also accrue to commercial organisations concurrently. For example, the Hub-of-All-Things (HAT) creates value by trading personal data by means of a platform that also collects, stores and manages data to optimise its value[120]. The data might originate from devices such as Fitbits or smart meters, or services such as social media, telecoms or banks and be combined with third party data to create new services.

There are barriers to creating new business models around disruptive technologies if new technologies are significantly cheaper and potentially bring an industry less revenue. For example, instead of having conventional behind-the-ear hearing aids, it is now possible to wear a connected chip that allows the user to download different algorithms, depending on the level of background noise to which they are exposed. This has the potential to reduce the cost of hearing aids from thousands to tens of pounds.

8.2 Embracing new technologies

The technology that is deployed in a company needs to be part of the corporate vision, so that technology 'push' is considered alongside market 'pull' and end value. Two main components of success are having the right skills at all levels of the organisation and building a mature digital platform. A data analysis system needs to embrace the rapid innovations in data platforms in an ongoing manner since they enable cost-effective analysis of data for specific purposes. It is important not to lock a solution down to one data platform, and to recognise that a perfect data platform for all types of data analysis problems is never going to materialize because each needs to make compromises between flexibility and efficiency.

Senior managers are loath to try and gain value from something that they do not understand: the challenge is in overcoming this perception. Demonstration projects driven by end users, rather than by technology providers, will be helpful.

"The most powerful examples I have seen of the use and benefits of big data and data analytics are when you get the right group of teams together to address specific problems. You have to be able to articulate and frame what the problem or the opportunity of the challenge is, however big or small."

Dr Angela Strank, Chief Scientist, BP

"Most boards lack engineers in leadership positions. There is a lack of curiosity about technical issues."

John Patsavellas, Manufacturing and Technical Director, Altro

8.3 Recruiting and retaining staff

Industry needs many more engineers with the required skills, or data scientists, to convert data analytics theory into business practice. Manufacturing is an example of a sector that has struggled to recruit, in part at least owing to outdated perceptions about manufacturing and an expectation that the job will not be sufficiently interesting. However, particularly in high value manufacturing, the nature of jobs is changing, and the prospect of new technologies and techniques such as the Internet of Things, informatics and data analytics will help to attract potential recruits.

A corporate culture is needed that shows that a company is providing an exciting environment that can nurture and grow its employees, so that people can be recruited and recruits remain energised and want to stay with the company. Senior leaders need to be open to the new ideas of their junior staff.

8.4 Building the right teams

The right type of engineers should be used for the right job. For example, in some cases it may be more appropriate to use specialist software engineers and IT systems engineers, rather than engineers from another area with some software skills.

Not all the skills need necessarily sit within an organisation. As long as commercial restrictions can be overcome, there is an argument for letting others from outside the industry solve problems. For example, Quantum Black, a data science consultancy, worked with Atkins and Arup on developing tools to monitor the construction of Crossrail (see Section 3.8). External organisations may also be able to bring in sector-specific expertise, or expertise in a specialist area such as cybersecurity. New kinds of partnerships are likely to emerge. In some cases, it is possible to realign existing capabilities to tackle big data and data analytics problems.

9. Developing the UK's data capability

9.1 Learning from what is already out there

Big data and data analytics are immature areas across all sectors, and much still needs to be done to develop them into robust, practical tools. There are common issues that span all sectors, as demonstrated by the series of seven sector-specific workshops that formed part of this project, each of which to a greater or lesser extent raised similar issues. Knowledge transfer between companies and countries, as well as sectors, will facilitate more rapid implementation, and help ensure that the benefits are reaped. The case studies in this report provide lessons for companies and sectors outside their immediate domain. Further sharing of excellent practice and innovation will help to further improve the performance of companies, organisations and sectors.

 "The oil and gas sectors have invested heavily in performance measurement of equipment and developing analytic tools for processing high volumes of data to look for outliers that may give early warning signs of failure or improved processes. Sharing best practice would be beneficial."

Dr Dougal Goodman OBE FREng, Chief Executive, The Foundation for Science and Technology

"We can try and optimise the miles per gallon on a car to the fifth decimal, when actually the answer is to take the train. I think quite often we miss those opportunities, where a completely different solution is necessary and it requires humility on all sides".

Erwin Frank-Schultz, Industry Technical Leader, IBM UK Ltd

9.2 Is a new type of engineer needed?

In the sector workshops, there was consensus that broad skills are needed for big data and data analytics in individuals and across teams, which might include some combination of the following: mechanical and electronic engineering, statistics, machine learning, mathematical modelling, complex system modelling, informatics, internet skills, data science and computing. Systems engineering is a foundational skill, and is useful both as a technical tool and as a basis for broader business thinking across all sectors. The teaching of systems thinking in universities to students studying computer science and software engineering is also needed, along with the introduction of information architecture and the design of information supply chains. The ability to visualise data is a necessary part of extracting useful information, as well as the ability to articulate information to decision-makers.

An understanding is required of the engineering issues, combined with an understanding of the sector, so that individuals or teams are able to identify whether or not data analytics is a suitable tool, and how they might be applied in a particular context. It is risky to deploy abstract data analysers who do not understand how the data is generated or how the data analysis should be interpreted.

Engineers can develop data science skills alongside their own specialism, but may also find themselves working alongside data scientists who bring specialist skills that cross the disciplines of statistics, mathematics, computer science and engineering, as well as an understanding of how data analytics may be applied to business in practice.

As digital data elements become ubiquitous, it is no longer possible to compartmentalise IT, and science and technology. Similarly to other areas of engineering, the solution to the problem may lie in the hardware, or the software, or both. People with open minds, who are able to work in teams and across disciplines, will be needed. Leaders will need to embrace these new ways of working.

9.3 How can education address this?

Education needs to start in schools, to nurture a curiosity about data and to develop an understanding of data, risk and statistics, and of the difference between correlation and causation. Education needs to foster a combination of engineering and IT thinking. Schoolchildren also would benefit from the career opportunities in big data and data analytics.

"Systems engineering is how we work in business. It is how we design, develop and manage complex systems through writing a business case, requirements capture, design, delivery, it is analysing data and it is validating and verifying the system."

Dr Graham Herries FIET, Director of Systems Integration, Engineering Excellence Group, Laing O'Rourke

"How do you prepare people for bigger picture leadership, because the picture is getting bigger all the time thanks to the availability of more and more data?"

Professor Ian Poll OBE FREng, Chief Executive, Poll Aerosciences Ltd

Training in the stewardship and use of the new data resources and new sorts of analytics will need to be part of existing professional engineering disciplines, as well as providing more emphasis on applied systems engineering. Certification of courses by professional institutions will need to be reviewed, including joint accreditation courses by different institutions to address how individuals obtain skills that go across traditional boundaries.

The training of engineers in response to big data will be supported by the formation of courses in data science and other related areas such as big data, cloud computing, cybersecurity, machine learning, intelligent systems and the Internet of Things that are springing up in UK universities at both undergraduate and postgraduate level. Responding to the need for interdisciplinary skills, the courses are often run jointly between different university departments, such as statistics and computer science, or engineering and computer science, although it remains to be seen how some of these courses will be accredited[121].

9.4 Further education

Further education is also of considerable importance and more further education is needed to create the required numbers of specialist data scientists and engineers with expertise in data science. New funding models are also required, as current funding is insufficient to meet this challenge. Conventional engineering apprentice courses do not include data skills, systems integration or process engineering. Ideally, engineers from graduate and apprentice training routes should be taught the same skills. The Manufacturing Technology Centre is opening an advanced manufacturing training centre which will equip people, from apprentices to those undertaking lifelong learning, with the appropriate skills.

10. Conclusions

THE OPPORTUNITIES
ARE WIDE-RANGING
AND DIVERSE, AND
THE UK IS RELATIVELY
WELL POSITIONED
TO EXPLOIT THE
NEW TOOLS OF
DATA ANALYTICS
AND ADVANCED
CONNECTIVITY

This report is based on seven sector-specific workshops convened by the Royal Academy of Engineering and the IET. It includes the views and experience of those already using big data and data analytics, and those who are grappling with new ways of working and with the development of new applications.

The report demonstrates that the opportunities are wide-ranging and diverse, and concludes that the UK is relatively well positioned to exploit the new tools of data analytics and advanced connectivity to enable improvements to existing products and services, and innovation in products and services. However, a number of key legal, regulatory, standards development and organisational conditions need to be in place to ensure that innovation in products and services is facilitated, and steps need to be taken to minimise technical risks associated with data, devices and data-driven systems.

The report identifies the key issues that need to be addressed in order to create a data-enabled economy. These include:

- deployment of best engineering practice to new areas such as software development that is critical to the development of the UK's data analytics capability

- investment in fixed broadband services and in ubiquitous high-speed mobile broadband services equivalent in performance to the targets set by the EU digital agenda

- the investigation of methodologies that allow the value of data assets to be formally recognised

- the development of industry standards, where appropriate, that accelerate innovation and market-making and establish a leading position internationally for the UK

- the exploration of mechanisms for creating markets in data that permit access to proprietary but potentially valuable data.

To facilitate this, the UK's data capability needs to be enhanced by ensuring that best practice is shared, and that people with the appropriate multidisciplinary skills are available to companies wishing to exploit opportunities.

Specific messages and recommendations (see Executive summary) aimed at industry, professional institutions, academia, regulators, standards bodies and government show how these issues can be taken forward. Their successful implementation will contribute to the formation of a thriving data-enabled economy that will help to address the UK's nagging productivity deficit and maintain the UK's competitiveness in an economically important area.

Appendix 1: Acknowledgements

The project was overseen by a working group:

Professor Jim Norton FREng FIET, Chair of the working group

Professor Nick Jennings FREng FIET, Regius Professor of Computer Science, University of Southampton; Chief Scientific Advisor for National Security, HM Government

Professor Cliff Jones FREng FIET, Professor of Computing Science, University of Newcastle

Dr Mike Short CBE FREng FIET, Vice President Research and Development, Telefonica

Dr Martyn Thomas CBE FREng FIET, Director, Martyn Thomas Associates Ltd

Dr David Watson FREng FIET, Director, IBM Research, UK

with support from the following staff:

Ahmed Kotb, Head of Sector – Information and Communications, IET

Dr Philippa Westbury, Policy Advisor, Royal Academy of Engineering

The Academy and the IET would like to thank the following peer reviewers:

Professor Jon Crowcroft FREng FRS FIET, Marconi Professor of Communications Systems, University of Cambridge

Liam Maxwell, Chief Technology Officer, HM Government

Dr Scott Steedman CBE FREng, Director of Standards, BSI

Members of the **Royal Academy of Engineering Digital Systems Community of Practice**

Members of the **IET Sector Panels**

We would also like to thank **Dr Dougal Goodman OBE FREng**, Chief Executive, The Foundation for Science and Technology, for his contribution to the study.

Appendix 2: Workshop participants

The Academy and the IET would like to acknowledge the contribution of participants at the seven workshops.

Advanced manufacturing

Professor Steve Garwood FREng (Chair)	Imperial College London
Dr Tony Heenan (Speaker)	Williams Formula One
Paul Calver	UK Trade and Investment
Ben Carter	Vodafone
Andrew Churchill	JJ Churchill Ltd
Professor Richard Clegg	Lloyd's Register Foundation
Sahar Danesh	IET
Harald Egner	Manufacturing Technology Centre
Dr Dougal Goodman OBE FREng	The Foundation for Science and Technology
Professor Robert Harrison	Warwick Manufacturing Group
Paul Hingley	Siemens plc
Lina Huertas	Manufacturing Technology Centre
Dr Irene Lopez de Vallejo	Connected Digital Economy Catapult
Professor Duncan McFarlane	University of Cambridge
Shaun Mundy	Oxford Instruments plc
Professor Andy Neely FREng	Cambridge Service Alliance
John Patsavellas	Altro
Madeleine Scott	EEF – the manufacturers' organisation
Professor Ashutosh Tiwari	Cranfield University
Nicola Walker	GSK

Built environment

Dr David Watson FREng FIET (Chair)	IBM UK
Michael Devriendt (Speaker)	Arup
Dr Payam Barnaghi	University of Surrey
Alan Berry	IET
Hugh Boyes FIET	IET
Tim Clark	IET
Professor Richard Clegg	Lloyds Register Foundation
Professor David Fisk FREng	Imperial College London
Dr Dougal Goodman OBE FREng	The Foundation for Science and Technology
Dr Graham Herries FIET	Laing O'Rourke
Dr Julia de Juan Verger	BSI
Professor Jim Norton FREng FIET (Chair of the Connecting Data Working Group)	
Duncan Price	Buro Happold
Dr Nishanth Sastry	King's College London
Philippa Shelton	Royal Academy of Engineering
Dr Mike Short CBE FREng FIET	Telefonica
Cameron Steel FIET	BK Design Associates
Professor Will Stewart FREng FIET	IET

Energy

Professor Jim Norton FREng FIET (Chair of the Connecting Data Working Group)	
Doug Johnson (Speaker)	Met Office
Jim Anning	British Gas Connected Homes
Adrien Bouyssi	Baringa
Professor Richard Clegg	Lloyds Register Foundation
Helen Farr	IET
Erwin Frank-Schultz	IBM UK Ltd
Dr Dougal Goodman OBE FREng	The Foundation for Science and Technology
Gordon Graham	IET
Professor Tim Green FIET	Imperial College London
Adrian Grilli	JRC
Professor Cliff Jones FREng FIET	University of Newcastle
Dr James Martin FREng	Mott MacDonald

Colin Myles	Schlumberger Ltd
John Okas	Real Wireless
Ashley Pocock	EDF Energy
Roisin Quinn	National Grid plc
Oliver Rix	Baringa
Dr Mike Short CBE FREng FIET	Telefonica
Dr Liane Smith FREng	WG Intetech
Dr Mahesh Sooriyabandara	Toshiba Research Europe Ltd
Dr Angela Strank	BP
Jennifer Woodruff	Western Power
Ian Welch	Ian Welsh Power Consulting Ltd and Energy Systems Catapult
Professor Xiao-Ping Zhang FIET	University of Birmingham

Transport

Professor Jim Norton FREng FIET (Chair of the Connecting Data Working Group)	
Davin Crowley-Sweet (Speaker)	Network Rail
Simon Ball	Transport Research Laboratory
James Cemmell	Inmarsat
Sahar Danesh	IET
Professor Mischa Dohler	King's College London
Bryan Donnelly	Association of Train Operating Companies
John Easton	University of Birmingham
Dr Yolande Herbath	Transport Systems Catapult
Sharon Kindleysides	ITS UK
Dr James Martin FREng	Mott MacDonald
Eman Martin-Vignerte	Bosch
Alan Newbold	Arup
Professor John Polak	Imperial College London
Steve Pryce	DB Schenker Rail UK
Theo Quick	CGI
Professor Simon Saunders FIET	Real Wireless

Health

Dr Mike Short CBE FREng FIET (Chair)	Telefonica
Maria Hernandez (Speaker)	CISCO
Julie Bretland	Our Mobile Health
Dr David Clifton FIET	University of Oxford
Ed Conley	UK Health Informatics Research Network
Professor David Delpy FREng FRS FMedSci	Defence Scientific Advisory Council
Dr Patrick Finlay	Institution of Mechanical Engineers
David Gough	NetScientific
Dr Sina Habibi	Diamentech Limited
Chris Lewis	Lewisinsight
Professor Ray Meddis	University of Essex
Cameron Steel FIET	BK Design Associates
Professor Mehdi Tavakoli	HealthTech and Medicines KTN
Dr Martyn Thomas CBE FREng FIET	Martyn Thomas Associates
Dr James Wilson	University College London
Dr Alex Wyke	Patientview

Aerospace and defence

Professor Jim Norton FREng FIET (Chair of the Connecting Data Working Group)	
Professor Tom McCutcheon (Speaker)	DSTL
Martin Agnew	Airbus Defence and Space Ltd
Peter Chapman	Rolls-Royce Controls and Data Services Ltd
Denton Clutterbuck FIET	Altran UK
Gary Elliott	Aerospace Technology Institute
Dr Paul Galwas	Connected Digital Economy Catapult
Nic Holt FREng	Fujitsu
Professor Nick Jennings FREng FIET	University of Southampton; Chief Scientific Advisor for National Security, HM Government
Professor Peter Johnson	University of Bath
Professor Roy Kalawasky FIET	Loughborough University
Dr Andy Mills	University of Sheffield
Professor Tim Norman	University of Aberdeen
Professor Ian Poll FREng	Poll Aerosciences Ltd

References and endnotes

1   HM Treasury (2015), *Fixing the Foundations: creating a more prosperous nation*, Cm 9098
     www.gov.uk/government/publications/fixing-the-foundations-creating-a-more-prosperous-nation

2   David Willetts (2013), *Eight great technologies* (Chapter 2 *The big data revolution and energy-efficient
     computing*, Policy Exchange. *www.policyexchange.org.uk/publications/category/item/eight-great-
     technologies* (accessed 14/10/15).

3   Department for Business, Innovation and Skills (2013), *Information economy strategy. www.gov.uk/
     government/uploads/system/uploads/attachment_data/file/206944/13-901-information-economy-
     strategy.pdf* (accessed 14/10/15).

4   Department for Business, Innovation and Skills (2013), *UK data capability strategy: seizing the data
     opportunity. www.gov.uk/government/publications/uk-data-capability-strategy* (accessed 14/10/15).

5   Government Office for Science (2014), *The Internet of Things: making the most of the Second Digital
     Revolution*, a report by the UK Government Chief Scientific Advisor. *www.gov.uk/government/uploads/system/
     uploads/attachment_data/file/409774/14-1230-internet-of-things-review.pdf* (accessed 14/10/15).

6   Letter to the Prime Minister from the Council for Science and Technology (2013), *The age of algorithms:
     algorithms, analytics, modelling and data for growth and public sector efficiencies. www.gov.uk/government/
     publications/the-age-of-algorithms* (accessed 14/10/15).

7   HM Government (2012), *Open Data White Paper: Unleashing the Potential. www.gov.uk/government/uploads/
     system/uploads/attachment_data/file/78946/CM8353_acc.pdf* (accessed 14/10/15).

8   Data.gov.uk (2012), *Open data strategies. data.gov.uk/open-data-strategies* (accessed 14/10/15).

9   Cabinet Office (2013), *G8 Open Data Charter. www.gov.uk/government/publications/open-data-charter*
     (accessed 14/10/15).

10   HM Treasury (2015), *Fixing the Foundations: creating a more prosperous nation*, Cm 9098. *www.gov.uk/
     government/publications/fixing-the-foundations-creating-a-more-prosperous-nation* (accessed 14/10/15).

11   Acatech position paper (2011), *Driving force for innovation in mobility, health, energy and production.
     www.acatech.de/fileadmin/user_upload/Baumstruktur_nach_Website/Acatech/root/de/Publikationen/
     Stellungnahmen/acatech_POSITION_CPS_Englisch_WEB.pdf* (accessed 14/10/15).

12   German National Academy of Science and Engineering (2011), *Agenda CPS (Cyber-physical systems).
     www.gtai.de/GTAI/Navigation/EN/Invest/Industries/Smarter-business/Smart-technologies/cyber-physical-
     systems,did=626022.html* (accessed 14/10/15).

13   Industry-Science Research Alliance Prospect Study (2013), *Prosperity through research – What tasks lie ahead
     for Germany? www.forschungsunion.de/pdf/isra_prospect_study_2013.pdf* (accessed 14/10/15).

14   Executive Office of the President, President's Council of Advisors on Science and Technology (2010), *Designing
     a digital future: federally funded research and development in networking and information technology*,
     Report to the President and Congress. www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-nitrd-
     report-2010.pdf (accessed 14/10/15).

15 Executive Office of the President (2012), *Obama Administration unveils 'Big data' initiatives: announces $200 million in new R & D investments*, press release. *www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf* (accessed 14/10/15).

16 Executive Office of the President (2012), *Big data Across the Federal Government. www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final_1.pdf* (accessed 14/10/15).

17 Executive office of the President (2015), *Multi-Agency Science and Technology Priorities for the FY 2017 Budget*, Memorandum for the heads of executive departments and agencies. *www.whitehouse.gov/sites/default/files/microsites/ostp/m-15-16.pdf* (accessed 14/10/15).

18 European Commission, Digital Agenda for Europe – Europe 2020 strategy, *Pillar VII: ICT-enabled benefits for EU society. ec.europa.eu/digital-agenda/en/our-goals/pillar-vii-ict-enabled-benefits-eu-society* (accessed 14/10/15).

19 European Commission (2014), *Towards a thriving data-driven economy*, communication to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions, Jul 2. COM(2014) 442 final, Brussels.

20 European Commission (2015), *A Digital Single Market Strategy for Europe*, communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, COM(2015) 192. *ec.europa.eu/priorities/digital-single-market/docs/dsm-communication_en.pdf* (accessed 14/10/15).

21 European Commission, 28 October 2015, press release, *A deeper and fairer Single Market: Commission boosts opportunities for citizens and business. europa.eu/rapid/press-release_IP-15-5909_en.htm* (accessed 29/10/15).

22 Ofcom (2015), *Strategic Review of Digital Communications*, discussion document. *stakeholders.ofcom.org.uk/binaries/consultations/dcr_discussion/summary/digital-comms-review.pdf* (accessed 14/10/15).

23 DCLG and DCMS (2015), *Review of How the Planning System in England Can Support the Delivery of Mobile Connectivity*, call for evidence. *www.gov.uk/government/uploads/system/uploads/attachment_data/file/443864/150709_Mobile_Infrastructure_Planning_Review_final.pdf* (accessed 14/10/15).

24 European Commission, Digital agenda for Europe – Europe 2020 strategy, *Pillar IV: Fast and ultra-fast internet access. ec.europa.eu/digital-agenda/en/our-goals/pillar-iv-fast-and-ultra-fast-internet-access* (accessed 14/10/15).

25 National Information Board and the Department of Health (2014), *Personalised health and care 2020: a framework for action. www.gov.uk/government/publications/personalised-health-and-care-2020/using-data-and-technology-to-transform-outcomes-for-patients-and-citizens* (accessed 14/10/15).

26 Department for Innovation, Business and Skills (2013), *Market assessment of public sector information*, written by Deloitte. *www.gov.uk/government/uploads/system/uploads/attachment_data/file/198905/bis-13-743-market-assessment-of-public-sector-information.pdf* (accessed 14/10/15).

27 CEBR (2012), *Data equity: Unlocking the value of big data*, report for SAS. *www.cebr.com/wp-content/uploads/2013/03/1733_Cebr_Value-of-Data-Equity_report.pdf* (accessed 14/10/15).

28 McKinsey and Company (2013), *Open data: unlocking innovation and performance with liquid information. www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information* (accessed 14/10/15).

29 Shadbolt, N. and Chui, M. (2014), *Big data, big new businesses. www.project-syndicate.org/commentary/nigel-shadbolt-and-michael-chui-estimate-that-the-annual-value-of-open-data-could-reach--3-trillion* (accessed 14/10/15)

The Open Data Institute, *The economic impact of open data: what do we already know? https://medium.com/@ODIHQ/the-economic-impact-of-open-data-what-do-we-already-know-1a119c1958a0* (accessed 03/11/15)

Nesta and the Open Data Institute (2015), *Open Data Challenge Series Final Report. opendatachallenges.org/wp-content/uploads/2015/10/Nesta-Final-report-26.10.15.pdf* (accessed 03/11/15).

30   Evans, P. and Forth, P. (2015), *Borges' Map: Navigating a world of digital disruption*, Boston Consulting Group. *www.bcgperspectives.com/content/articles/borges-map-navigating-world-digital-disruption/* (accessed 14/10/15).

31   Google Inc. (2015), *Consolidated Balance Sheets. investor.google.com/pdf/2015Q1_google_earnings_data.pdf* (accessed 14/10/15).

32   Planet Labs website, *In space for life on earth. www.planet.com/impact/* (accessed 14/10/15).

33   Parliamentary Office of Science and Technology (2014), *Social media and big data*, POSTnote number 460. *researchbriefings.files.parliament.uk/documents/POST-PN-460/POST-PN-460.pdf* (accessed 14/10/15).

34   Waze website, *Get the best route, every day, with real-time help from other drivers. www.waze.com* (accessed 14/10/15).

35   Ushahidi website, *Changing the way information flows. www.ushahidi.com/about* (accessed 14/10/15).

36   Zooniverse website, *People-powered research. https://www.zooniverse.org/* (accessed 14/10/15).

37   GSMA (2015), *Intelligent Transportation Systems Report for Mobile*, GSMA Connected Living Programme. *www.gsma.com/connectedliving/wp-content/uploads/2015/06/ITS-report.pdf* (accessed 14/10/15).

38   TeleSign website, TeleSign and Telefónica announce mobile identity and fraud prevention agreement, press release, London, 27 August 2015. *https://www.telesign.com/resources/press-releases/* (accessed 14/10/15).

39   Robotics and Autonomous Systems Special Interest Group (2014), *RAS 2020: A national strategy to capture value in a cross-sector UK RAS innovation pipeline through co-ordinated development of assets, challenges, clusters and skills. https://connect.innovateuk.org/documents/2903012/16074728/RAS%20UK%20 Strategy?version=1.0* (accessed 14/10/15).

40   Innovate UK and the Robotics and Autonomous Systems Special Interest Group (2015), *The UK Landscape for Robotics and Autonomous Systems. connect.innovateuk.org/documents/2903012/19163277/The%20 UK%20Landscape%20for%20Robotics%20%26%20Autonomous%20Systems* (accessed 14/10/15).

41   For example, Solomon Associates (*solomononline.com*) who provide monitoring and performance management services to the energy industry, and Dunn Humby (*www.dunnhumby.com*) who provide analytics services to the retail sector (accessed 14/10/15).

42   These include Kainos (*www.kainos.com*, accessed 14/10/15), providing data analytics and cloud services, as well as an Electronic Medical Records platform; Quantum Black (*www.quantumblack.com*, accessed 14/10/15), a data science consultancy using analytics and visualisation that has worked with the built environment, aerospace, finance and healthcare sectors, and in Formula One racing; and Mastodon C (*www.mastodonc.com*, accessed 14/10/15), a big data start-up who use an open source technology platform and are working with the Greater London Authority (GLA) on city applications.

43   CEBR (2013), *Data on the balance sheet*, Report for SAS. *www.cebr.com/reports/data-on-the-balance-sheet* (accessed 14/10/15).

44   EURIM The Information Society Alliance (2009), *Unlocking the value of information*, status report. *www.eurim.org.uk/activities/ig/0909-Unlocking_the_Value_of_Information.pdf* (accessed 14/10/15).

45   Wilco Engelsman (2007), *Information Assets and their value. http://referaat.cs.utwente.nl/conference/6/ paper/6807/information-assets-and-their-value.pdf* (accessed 14/10/15).

46   Higson, C. and Waltho, D. (2010), *Valuing information as an asset. faculty.london.edu/chigson/research/ InformationAsset.pdf* (accessed 14/10/15).

47   EURIM Valuation of Information Subgroup, *Valuation of information assets* [list of reports containing valuation methods]. *www.eurim.org.uk/activities/ig/voi/valuation.php* (accessed 14/10/15).

48   CEBR (2013) *Data on the balance sheet*, report for SAS. *www.cebr.com/reports/data-on-the-balance-sheet* (accessed 14/10/15).

49   CEBR (2012), Data equity: *Unlocking the value of big data*, report for SAS. *http://www.cebr.com/wp-content/ uploads/2013/03/1733_Cebr_Value-of-Data-Equity_report.pdf* (accessed 14/10/15).

50   Government Office for Science and the Department for Business, Innovation and Skills (2013), *Future of manufacturing: a new era of opportunity and challenge for the UK. www.gov.uk/government/publications/ future-of-manufacturing* (accessed 14/10/15).

51   Germany Trade and Invest (2014), *Industrie 4.0 – Smart manufacturing for the future. www.gtai.de/GTAI/ Navigation/EN/Invest/Service/Publications/business-information,t=industrie-40--smart-manufacturing-for- the-future,did=917080.html* (accessed 14/10/15).

52   VDE Association for Electrical, Electronic and Information Technologies (2014), *The German Standardization Roadmap Industrie 4.0,* Version 1.0. *http://www.vde.com/en/dke/DKEWork/NewsfromtheCommittees/2014/ Documents/German%20Standardization%20Roadmap%20Industrie%204.0.pdf* (accessed 14/10/15).

53   IET, *Digital engineering and project controls in the construction industry*, Case study provided by Laing O'Rourke. *www.theiet.org/sectors/built-environment/files/Laing-digital-casestudy.cfm* (accessed 14/10/15).

54   For example, data on user behaviour generated in novel ways will help achieve designs that reduce resource use, and increase health and wellbeing, and performance and productivity: Royal Academy of Engineering, Arup and ESRC (2015), *Built for living: understanding behaviour and the built environment through engineering and design.*

55   Demand Logic website. *www.demandlogic.com* (accessed 14/10/15).

56   London Datastore website [case studies illustrating how data is being used to drive policy in London]. *data.london.gov.uk/case-studies/* (accessed 14/10/15).

57   HM Government (2015) *Digital Built Britain: Level 3 Building Information Modelling – Strategic Plan. www.gov.uk/government/publications/uk-construction-industry-digital-technology* (accessed 14/10/15).

58   Robotics and Autonomous Systems Special Interest Group (2014), *RAS 2020 – A national strategy to capture value in a cross-sector UK RAS innovation pipeline through co-ordinated development of assets, challenges, clusters and skills. https://connect.innovateuk.org/documents/2903012/16074728/RAS%20UK%20 Strategy?version=1.0* (accessed 14/10/15).

59   CBI (2013), *Building Britain's future: An industrial strategy for construction. www.cbi.org.uk/media/2118883/building_britains_future.pdf* (accessed 14/10/15).

60   The suite of national standards for Level 2 BIM includes a cybersecurity standard for BIM software systems: BSI (2015), *PAS 1192-5 – Specification for security-minded building information modelling, digital built environments and smart asset management.*

61   Work by the IET (*Power Network Joint Vision, www.theiet.org/pnjv*, accessed 14/10/15) has raised awareness of the need for whole-systems thinking and has drawn attention to a potential gap in the sector's technical governance, as there is no party with whole-system oversight or accountability. This matter is being addressed by further work sponsored by DECC, through the Energy Systems Catapult, and undertaken by The IET.

62   Open Utility website. *https://www.openutility.com/* (accessed 14/10/15).

63   Phys.org (2015), *A seemingly obvious way to make the electricity market better may actually make it worse. phys.org/news/2015-07-seemingly-obvious-electricity-worse.html* (accessed 14/10/15).

64   Standards in existence include for example: PD CEN/TR 16061:2010, *Gas meters – Smart gas meters.*

65   Standards in development include for example: BS EN 62056, *Electricity metering – data exchange for meter reading and communication profiles for local data transmission* and BS EN WI 16836, *Remote reading of gas meters – wireless mesh networking for meter data exchange.*

66   The work of the Digital Catapult and its Trust Framework initiative includes the development of standards that may lead to a 'Kite mark' type of assurance scheme to build public confidence. Similar schemes already exist, for example, *BSI Kitemark for secure digital transactions. www.bsigroup.com/en-GB/our-services/product- certification/industry-sector-schemes/bsi-kitemark-for-secure-digital-transactions/* (accessed 14/10/15).

67   Citymapper website. *www.citymapper.com* (accessed 14/10/15).

68   INRIX website. *www.inrix.com* (accessed 14/10/15).

69 For example, Masternaut: Mobile resource management website, *Transport and logistics sector.* *www.masternaut.co.uk/vehicle-tracking-sectors/transport-logistics/* (accessed 14/10/15).

70 Parliamentary Office of Science & Technology (2014), *Big data and Public Health,* POSTnote number 474. *researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-474/#fullreport* (accessed 14/10/15).

71 See for example the work of the Farr Institute (*www.farrinstitute.org*, accessed 14/10/15) and the Computational Health Informatics laboratory at Oxford University (*www.robots.ox.ac.uk/~davidc/*, accessed 14/10/15).

72 National Information Board and the Department of Health (2014), *Personalised health and care 2020: a framework for action. http://researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-474/#fullreport* (accessed 14/10/15).

73 National Information Board (2015), *Plans to improve digital services for the health and care sector.* *www.gov.uk/government/publications/plans-to-improve-digital-services-for-the-health-and-care-sector* (accessed 14/10/15).

74 uMotif website, *Digital Health. umotif.com* (accessed 14/10/15).

75 ADI Health website, *Digital Health Assets. www.adi-uk.com/adi-digital-health-assets.html* (accessed 14/10/15).

76 Cupris Health website. *www.cupris.com* (accessed 14/10/15).

77 Fitbit website. *www.fitbit.com/uk* (accessed 14/10/15).

78 23andme website. *www.23andme.com* (accessed 14/10/15).

79 Recent research highlights the significant contribution of a small number of determinants, mostly behavioral in nature, to health outcomes: Health Policy Brief: *The Relative Contribution of Multiple Determinants to Health Outcomes,* Health Affairs, August 21, 2014. *http://www.healthaffairs.org/healthpolicybriefs/brief.php?brief_id=123* (accessed 14/10/15).

80 Innovation UK and the British Standards Institution (2015), *PAS 277: Health and wellness apps – Quality criteria across the life cycle – Code of practice. http://shop.bsigroup.com/upload/271432/PAS%20277%20(2015) bookmarked.pdf* (accessed 14/10/15).

81 For example, the national committee IST/35 on *Health informatics* is developing standards relating to records, data and communications in the health sector, facilitated by the Health and Social Care Information Centre (HSCIC).

82 Ministry of Defence, Defence Science and Technology Laboratory and The Rt Hon Michael Fallon MP, 16 September 2015, *MOD unveils futuristic uniform design,* press release. *www.gov.uk/government/news/mod-unveils-futuristic-uniform-design* (accessed 14/10/15).

83 RUSI (2015), *A Democratic Licence to Operate,* report of the Independent Surveillance Review. *https://www.rusi.org/sites/default/files/20150714_whr_2-15_a_democratic_licence_to_operate.pdf* (accessed 14/10/15).

84 Financial Times, 30 June 2015, *Insurers at risk from Uber-style upstarts.*

85 Pool Re website, *Reinsuring terrorism risk. www.poolre.co.uk* (accessed 14/10/15).

86 Financial Times, 29 June 2015, *Admiral seals connected car agreement with Redtail Telematics. www.ft.com/cms/s/0/c7c8cc46-1a92-11e5-8201-cbdb03d71480.html#axzz3n2obGMTa* (accessed 14/10/15).

87 Lloyd's and the University of Cambridge Centre for Risk Studies Business Blackout (2015), *The insurance implications of a cyber attack on the US power grid. http://www.lloyds.com/~/media/files/news%20 and%20insight/risk%20insight/2015/business%20blackout/business%20blackout20150708. pdf#search='cyber%20cover'* (accessed 14/10/15).

88 Lloyd's market news, February 2013, *International demand for cyber cover increases. www.lloyds.com/news-and-insight/news-and-features/market-news/industry-news-2013/international-demand-for-cyber-cover-increases* (accessed 14/10/15).

89 Lloyd's and the University of Cambridge Centre for Risk Studies Business Blackout (2015), *The insurance implications of a cyber attack on the US power grid.*

90 *Connecting data* case studies, *www.raeng.org.uk/connectingdata/casestudies*

91 A recent study of travellers' attitudes to intelligent mobility found that 57% of respondents would not mind sharing their personal data in order to get a better service: Transport Systems Catapult (2015), *Traveller needs and UK capability study: supporting the realisation of intelligent mobility in the UK.*

92 Chessell, M. (2014), *Ethics for big data and analytics*, IBM. *http://www.ibmbigdatahub.com/sites/default/files/whitepapers_reports_file/TCG%20Study%20Report%20-%20Ethics%20for%20BD%26A.pdf* (accessed 14/10/15).

93 Anderson, R. (2002), *Security in Open versus Closed Systems – The Dance of Boltzmann, Coase & Moore. www.cl.cam.ac.uk/~rja14/Papers/toulouse.pdf* (accessed 14/10/15).

94 Institute for Public Policy Research Commission on National Security in the 21st Century (2009), *Shared Responsibilities: A national security strategy for the UK. www.ippr.org/publications/shared-responsibilitiesa-national-security-strategy-for-the-uk*

95 European Commission, *Protection of personal data. ec.europa.eu/justice/data-protection* (accessed 14/10/15).

96 European Commission (2015), *Data Protection Day 2015: Concluding the EU Data Protection Reform essential for the Digital Single Market,* fact sheet. *europa.eu/rapid/press-release_MEMO-15-3802_en.htm* (accessed 14/10/15).

97 TAP blog, June 2015, *Christopher Millard Asks: Is Privacy the Real Driver of Forced Localization of Cloud Services? www.techpolicy.com/Millard-IsPrivacyDriver-ForcedLocalizationOfCloudSrvcs.aspx* (accessed 14/10/15).

98 Information Commissioner's Office, *Guide to data protection. ico.org.uk/for-organisations/guide-to-data-protection/* (accessed 14/10/15).

99 For example in the common law of Tort, this is the principle that underpins the classic case of Rylands v Fletcher. In addition to that it is also the central principle of the 1974 Health and Safety at Work Act.

100 See Business Companion website, *Trading standards explained. http://www.businesscompanion.info/en/quick-guides/digital/digital-content* (accessed 14/10/15).

101 Nesta blog, 14 October 2015, *Precision Agriculture: separating the wheat from the chaff. http://www.nesta.org.uk/blog/precision-agriculture-separating-wheat-chaff?utm_source=Nesta+Weekly+Newsletter&utm_campaign=a6d127350d-Nesta_newsletter_14_10_1510_13_2015&utm_medium=email&utm_term=0_d17364114d-a6d127350d-180631397* (accessed 14/10/15).

102 Open Data Institute (2015), *Who owns our data infrastructure? www.scribd.com/doc/266889068/Who-Owns-Our-Data-Infastructure#download* (accessed 14/10/15).

103 Mydex provides the individual with a hyper-secure storage area to enable them to manage their personal data, including text, numbers, images, video, certificates and sound. No-one but the individual can access or see the data. *mydex.org* (accessed 14/10/15).

104 Aeroexchange website, *Aviation supply chain solutions. www.aeroxchange.com* (accessed 14/10/15).

105 Covisint website. *www.covisint.com* (accessed 14/10/15).

106 RCUK Digital Economy HAT Project (2014), *Briefing Paper 1 – Engineering a Market for Personal Data: The Hub-of-All-Things (HAT). hubofallthings.com/downloads/hat-briefing-paper-engineering-market-personal-data/* (accessed 14/10/15).

107 VDE Association for Electrical, Electronic and Information Technologies (2014), *The German Standardization Roadmap Industrie, 4.0,* Version 1.0.

108 British Standards Institute (2014), *PAS 182 – Smart city concept model. www.bsigroup.com/en-GB/smart-cities/Smart-Cities-Standards-and-Publication/PAS-182-smart-cities-data-concept-model/* (accessed 14/10/15).

109 CRISP-DM website. *Cross industry standard process for data mining. crisp-dm.eu/home/about-crisp-dm* (accessed 14/10/15).

110 *Open standards and licencing: using open standards to build your service, https://www.gov.uk/service-manual/making-software/open-standards-and-licensing.html*

111 BSI in its role as the national standards body maintains the infrastructure for over 11,000 UK experts to participate in national, European and international standards development activity, where the UK is recognised as one of the leading players. *www.bsigroup.com*

112 BSI Group, *Smart cities standards and publications, www.bsigroup.com/en-GB/smart-cities/Smart-Cities-Standards-and-Publication/* (accessed 14/10/15).

113 BSI Group (2014), *PAS 182:2014 – Smart city concept model: Guide to establishing a model for data interoperability. shop.bsigroup.com/upload/Smart_cities/BSI-PAS-182-executive-summary-UK-EN.pdf* (accessed 14/10/15).

114 British Standards Institution (2014), *PAS 181 Smart City Framework. www.bsigroup.com/en-GB/smart-cities/Smart-Cities-Standards-and-Publication/PAS-181-smart-cities-framework/* (accessed 14/10/15).

115 Aerospace journal, August 2015, *Big data in the aerospace sector*, Royal Aeronautical Society

116 Klaschka, R. (2015), *Our buildings, our data... maybe not*, CIBSE Journal, July 2015.

117 Aerospace journal, August 2015, *Big data in the aerospace sector*, Royal Aeronautical Society

118 Ng, Irene C.L. (2014), *New business models in the connected digital economy*, Journal of Revenue and Pricing Management, 13(2).

119 Surrey Satellite Technology (2014), *Applications of earth observation. www.sstl.co.uk/Downloads/Brochures/SSTL-Applications-Brochure-Web* (accessed 14/10/15).

120 RCUK Digital Economy HAT Project (2014), *Briefing Paper 2 – Engineering a Market for Personal Data: The Hub-of-All-Things (HAT): Economic Model of the Multi-Sided Market Platform and Ecosystem. hubofallthings.com/downloads/economic-model-multi-sided-market-platform-ecosystem-briefing-paper-2/* (accessed 14/10/15).

121 Royal Statistical Society and Nesta (2014), *Model workers: how leading companies are recruiting and retaining their data talent.*

Notes

Royal Academy of Engineering

As the UK's national academy for engineering, we bring together the most successful and talented engineers for a shared purpose: to advance and promote excellence in engineering.

We have four strategic challenges:

Make the UK the leading nation for engineering innovation

Supporting the development of successful engineering innovation and businesses in the UK in order to create wealth, employment and benefit for the nation.

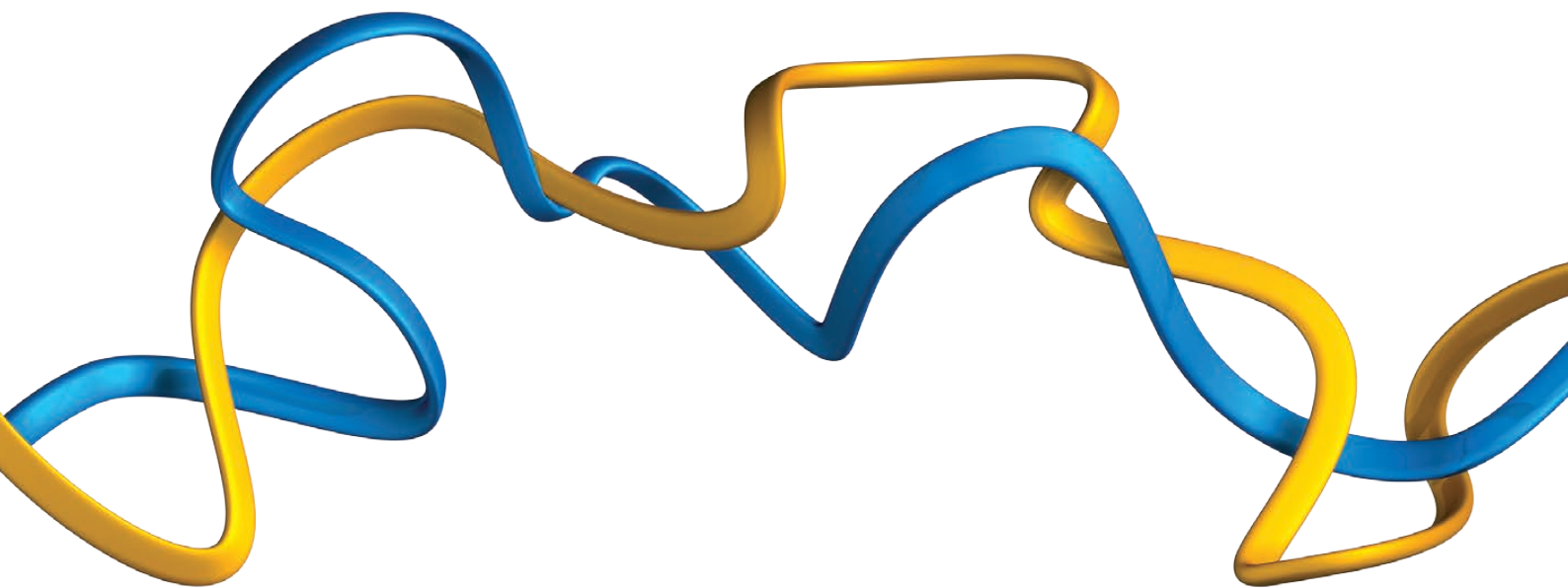**Address the engineering skills crisis**

Meeting the UK's needs by inspiring a generation of young people from all backgrounds and equipping them with the high quality skills they need for a rewarding career in engineering.

Position engineering at the heart of society

Improving public awareness and recognition of the crucial role of engineers everywhere.

Lead the profession

Harnessing the expertise, energy and capacity of the profession to provide strategic direction for engineering and collaborate on solutions to engineering grand challenges.

The cover is treated with a recyclable laminate