The IET
Michael Faraday House
Six Hills Way, Stevenage
Hertfordshire, SG1 2AY
United Kingdom
T +44 (0) 1438 313311
F +44 (0) 1438 765526
www.theiet.org

Ref: S1018/ab

House of Commons
Science and Technology Committee
E: scitechcom@parliament.uk

03 September 2015

Science and Technology Committee (Commons)

Re: The Big Data Dilemma inquiry

This response has been compiled by the IET and the Royal Academy of Engineering. It is based on the views of Members and Fellows from both organisations with expertise in digital systems as well as related systems such as transport, Energy, Innovation and manufacturing.

Evidence was also taken from seven sector-specific stakeholder workshops on big data that formed part of *Connecting data*, a joint project between the Royal Academy of Engineering and the IET.
The IET and the Royal Academy of Engineering are happy to discuss these points with the Ministers or Officials.

Yours sincerely,

Paul Davies
Head of Policy
Tel: 01438 765687 Email: pdavies@theiet.org

Enc.

‘Big data’ – Science and Technology Committee inquiry

This response has been compiled by the IET and the Royal Academy of Engineering. It is based on the views of Members and Fellows from both organisations with expertise in digital systems as well as related systems such as transport and manufacturing. Evidence has also been included gathered from seven sector-specific stakeholder workshops on big data that formed part of *Connecting data*, a joint project between the Royal Academy of Engineering and the IET[1].

Introduction

Big data, while poorly understood and defined, is all around us. Citizens, by default, add to big data sets by their actions, for example through purchasing, making a phone call, using energy, using social media, banking and finance and using public or private transport. The privacy of their data needs to be protected by ensuring appropriate data protection legislation keeps pace with developments in big data technologies.

Big data has come into existence as a result of the rapidly declining cost of storage together with the emergence and rapidly declining cost of the communications technologies underlying the internet and web that have together enabled the storage and harvesting of data as users physically transact or access remote network services by both fixed and, increasingly, mobile devices.

Other technologies enabled by these trends such as cloud computing and utilisation of open source software have been a major enabling force both in reducing the cost of storage and improving resilience while enabling new and innovative services.

It is important to realise that there are different types of big data:

- dynamic data sets that are rapidly changing, or nearer real time data sets such as mobile phone connection and tracking, weather data and air traffic control

- new and expanding data sets that can be dynamically generated from any network-connected device that interacts with the world around us. These devices, generically known as the Internet of Things (IoT), along with the standards surrounding connection and data exchange can be used in conjunction with other data to exploit a whole range of more detailed information

- slower moving data sets such as credit card data, energy billing, loyalty cards, revenue and tax collection and mapping services

- big data that is still disconnected and potentially subject to privacy concerns such as health data, where much data is held in many silos that are difficult to share across organisations. This can result in duplication errors, business inefficiency and high cost of ownership which further limits the usefulness and reliability of the data. There are many other instances of such data sets. Sometimes when such data sets are linked, major benefits can be unlocked.

[1] *Connecting data* investigates how advanced various sectors and organisations are currently in their use of big data and advanced connectivity, and how this might develop in the future. The study focuses on both engineering and other sectors, including transport, healthcare, built environment, energy, advanced manufacturing, aerospace and defence and insurance, and is due to be published in Autumn 2015.

Responses to inquiry questions

1. The opportunities for big data, and the risks

General opportunities

- Big data, combined with increasingly networked systems and improved analytics, could help to improve, and possibly even transform, how all sectors operate. This would create economic value for the UK by increasing productivity, driving innovation and transforming employment. It could also achieve broader social and environmental benefits by facilitating improvements in healthcare, transportation, energy use, the built environment and national security. Failure to invest in big data technologies is likely to place the UK at a considerable disadvantage.
- Big data is already used in a range of applications including managing infrastructure, telecoms and air traffic control; facilitating banking and ecommerce; helping with medical diagnosis, research, genetic engineering and the development of new drugs; and for fundamental research (for example particle physics at CERN).
- The IET/Royal Academy of Engineering *Connecting data* study, due to be published in the autumn of 2015, has shown that, while there are best practice examples where big data and data analytics have been successfully applied, the area is still largely immature, so there remains great potential for innovation and value generation in future years, to the benefit of business and society as a whole.
- For engineering sectors, big data has the potential to drive innovations in engineering products and services, and to improve engineering systems throughout their lifecycle, including design, manufacturing, maintenance and decommissioning[2]. Benefits include better efficiency, cost-effectiveness and safety, and more sophisticated ways of evaluating and managing risk.
- Business models based on the use of big data are emerging that both represent incremental improvements in business performance, through the optimisation of existing processes and also the creation of new products and services[3]. For example, 'servitisation', which describes the process of enhancing the value of products by adding services, or even selling services in place of products. The aerospace sector helped to pioneer this with Rolls-Royce's 'Power by the hour' [4] business model, but new big data technologies allow it to be applied much more widely. Increasingly sophisticated predictive analytics techniques enhance the potential of these business models further: for example, through the ability to predict system failures and identify performance improvements. This could offer enhanced customer value as well as being of value to the product or service provider.
- Opportunities to solve broader societal challenges by making links across sectors are numerous: for example, 'smart cities' will require the effective sharing of data between sectors including energy, transport and built environment, in combination with novel sources of data such as that generated by social media or crowd-sourcing. Insurance policies are being written for drivers based on evidence obtained from monitoring driving behaviour. New

[2] Lloyd's Register Foundation (2014), Foresight review of big data: towards data-centric engineering http://www.lrfoundation.org.uk/publications/bigdata.aspx

[3] Hartmann, P.M., Zaki, M., Feldmann, N. and Neely, A. (2014) Big Data for Big Business? A Taxonomy of Data-driven Business Models used by Start-up Firms (working paper), Cambridge Service Alliance

[4] Rolls-Royce press release (2012) Rolls-Royce celebrates 50th anniversary of Power-by-the-Hour http://www.rolls-royce.com/news/press-releases/yr-2012/121030-the-hour.aspx

wearable technologies and mobile applications are changing approaches to wellbeing and healthcare. Innovative services such as these will challenge the existing regulatory environment.

- Autonomous systems rely on the rapid transmission of large volumes of data: they have been called the 'arms and legs [and eyes] of big data'[5].
- Other applications of big data to autonomous systems lie in advanced manufacturing, providing the potential to improve efficiencies, create a quicker 'prototype' to market" turn-around, a greater degree of mass-customisation and allowing the manufacturing supply-chain to become leaner and more sustainable.
- In health, as in other sectors, the growing ability to monitor individuals or systems in real-time, which requires rapid analysis of large volumes of data, holds much potential.
- The UK has considerable strengths in multidisciplinary innovation and is therefore well-placed to compete internationally, as long as international standards and consistent approaches to data protection and cybersecurity are in place. Many internet companies have a global view and use Europe as a single market, and the UK needs to continue to look attractive as a place to innovate in big data.
- A more detailed picture of the opportunities presented by individual sectors including transport, manufacturing and energy is given in the following section. A discussion of the opportunities in a range of sectors will also be given in the *Connecting data* report (see Footnote 1 for the list of sectors), along with case studies illustrating successful applications of big data.

Specific opportunities by sector

A. Transport

There is great potential for the utilisation of big data in transport. Use of such data will allow the transport system to be more responsive and agile in overcoming congestion and reducing delays. Data analytics can assist in monitoring the patterns in a transport network and so help predict performance and throughput while improving efficiency. These parameters can then be used to help take a proactive approach to planning and management of infrastructure and support services. An example of this is Intelligent Transport Systems (ITS). These make use of data, sensor technologies and wireless communications to provide innovative services that may be coordinated between different modes of transport.

Big data has led to a major change in the options available for system-to-system communications in transport. This has created new opportunities for integration and data sharing between various systems, even when they are in different authorities or geographical locations.

Integrated travel data can deliver bespoke assistance to the traveller by helping re-plan journeys and understanding the individual traveller's profile and preferences; it can also exert a 'nudging' influence by trying to persuade travellers to use the more environmentally friendly options for travel when practical. To deliver integrated information, there is a need for a city and its transport providers to deliver open data.

[5] The Robotics & Autonomous Systems Special Interest Group (2014) RAS 2020 https://connect.innovateuk.org/documents/2903012/16074728/RAS%20UK%20Strategy?version=1.0

Big data matters from a transport perspective because it can:

- inform users of alternative options, thereby influencing modal choice. This could be used to encourage a modal shift to more sustainable forms of transport (including individuals sharing transport)
- filter data such that only pertinent information reaches a driver, hence significantly reduce driver distraction and stress which are known to be common causes of incidents.
- enable real time personalised travel planning
- facilitate the reduction in street furniture (information signs, lines on roads, etc.) as information could instead be provided in-vehicle.
- provide service providers with better intelligence that allows them to make more strategic operational interventions in a time of disruption or during major events
- provide service providers with meaningful information that can assist planning and investment decisions. This can be small scale, such as altering the frequency or route of a public transport service or much wider, such as informing planners tasked with regional development of the wider transport needs
- enable increasing autonomous operation of vehicles and other transport platforms. For example, connected cars[6] will be the first mobile big data platforms. They will need to be highly connected to other road users (cars, lorries, cyclists, pedestrians), and the road infrastructure (the road, roundabouts, signalling, buildings). They may also be connected to other parties including insurers, regulators, owners and service providers.

B. Energy

Within the energy sector, sophisticated data mining is in its infancy but growing. There is the potential to improve asset utilisation and identify anomalies in asset performance and consumer behaviour in real time. Given the extended lifetimes of many of the assets in the energy sector, particularly in generation and grid systems, this could significantly improve the efficiency and cost-effectiveness of the system.

Data will be particularly important as the electrical system evolves to become much more dynamic and two-way than it currently is. Increasing amounts of embedded generation such as solar PV are already putting strain on some local distribution systems. Data to understand and manage the flows more precisely will be important for system operators in the future.

Smart meters will be the main source of data once the government roll-out programme is completed and will be the first step towards a more complete 'smart grid' that will connect individual demand loads with the network and sources of supply. It is hoped that this will drive a reduction in overall demand as well as more sophisticated demand management that could ultimately reduce both the generating capacity required and the energy consumed.

Greater access to data should also increase the level of consumer engagement and empowerment in the energy system if presented in clear and user-friendly way. However, the risk of public resistance owing to privacy issues could adversely impact progress. To prevent this, there needs to be a debate on balancing privacy and individual/public benefits and convenience.

C. Innovation and manufacturing:

[6] The Transport Systems Catapult is currently running a trial of connected cars in Milton Keynes https://ts.catapult.org.uk/en/web/transport-systems/pods

Big data is now playing a major role in modern manufacturing. Data on consumer and market use of products can be utilised to personalise a product and develop products based on the consumer's needs. Currently, there is competition between online and high-street retailers on gaining consumer data in the fast-moving consumer goods sector. Manufacturers need to be prepared for the debate on the ownership of data and how product and consumer data is being used. Appropriate use of loyalty card information could contribute to sustainable manufacture and distribution by reducing wastage and allowing better resource planning.

Big data is also an issue in high value manufacturing, as know-how and intellectual property (IP) become entwined with product data that will be shared across a supply chain. The potential for leakage of IP in this manner is high and protection needs to be appropriate. This effect will be felt more strongly by the supply chain and the smaller manufacturers who could be obliged to effectively distribute parts of their IP –- so ensuring that their product innovation remains their own will be vital.

D. The end-user

A key opportunity in big data is delivering 'smart cities' and their outreach into rural communities. Smart cities could enable the participation of people in community life, as well as enhancing the performance of services and reducing costs and resource consumption, leading to a better quality of life, strengthened economies and a reduced impact on the environment. They could contribute to delivering strategic objectives that are laid out in local plans, including:

- improving safety
- improving access to education, healthcare and places of employment
- improving the environment (both aesthetically and reducing pollution and the carbon footprint)
- creating an environment for business sustainability and growth.

Examples of how big data in smart cities can help deliver these objectives include:

- informing public transport providers of the need for new or revised services to make it easier to travel to education or healthcare establishments, workplaces, etc. and/or providing information that might enable these establishments to alter the way in which they deliver their services
- defining the requirement for new energy infrastructure such as heat networks in conjunction with city planning
- supporting decision-making for policy-makers concerned with the governance and management of urban areas
- providing information to people on local services, empowering them to make more informed decisions

Public users will obtain significant benefit from inputting personal data into the system but there is the downside, the potential use of the data to target the end users, whether this is to advertise products or services or undesirable activities, such as social engineering to enable fraud or knowing when their homes are left unoccupied and hence more susceptible to burglary.

The IET's recent proposal to create a next generation of '999' services, noted that accurate data made available to an operator could save lives. Data can be utilised to give emergency services and healthcare professional more information at a quicker pace. Optimum usage here may well require a combination of public and private initiatives.

The risks

Big data is not without risk and this is an area that is not well understood. Examples of big data related risks include:

- Increasing reliance on data and data-driven 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 widespread practical and economic damage. Where data is used to drive physical systems, specific technical challenges need to be considered around the complexity, uncertainty and resilience of systems of systems that are created by integrating previously separate networks of sensors and devices..
- Cybersecurity is another significant concern. Guidance is available on how organisations can protect themselves[7], but wider implementation is still needed. Poor cybersecurity puts at risk privacy as well as national, physical and personal security and intellectual property. In the case of the latter, this could be through revealing trade secrets, proprietary information and enabling reverse engineering. Hacking also provides the potential for data corruption.
- Cybersecurity is often thought of as primarily addressing confidentiality issues, such as who should be able to access the data. However, for big data to retain its value there is a need to address other technical security considerations such as utility, integrity and authenticity of the data. Such considerations should be addressed in a holistic fashion taking into account the people, process and physical aspects of cybersecurity [8].
- A further risk associated with big data is the identification of sensitive information either through data aggregation (the analysis of large volumes of data) or data association (analysis of multiple sources of data).
- There are risks around maintaining public trust in data, and ensuring the right balance of data protection for individuals versus the need to support business in following opportunities. The risks can be mitigated by having open and honest conversations about the benefits to society as a whole of making data available versus the risks to individuals.
- Care must be taken when using big data to represent the world and preventing over-simplification. It is important to ensure that the right questions are being asked of data, and that its interpretation incorporates an understanding of the context. Furthermore, it should not be assumed beforehand that the right solution to a problem involves using data, otherwise we are at risk of a 'monoculture' where all problems require data to solve them.
- There is a need to create sustainable markets in data and standards that ensure that data is fit-for-purpose and useable both within and across sectors. Without these, there is a risk of insufficient accessibility to high quality data within and across sectors,
- There is also a risk to the success of businesses and the UK economy as a whole if senior management fails to understand and develop a formal strategy to take advantage of the big data revolution.

[7] Cabinet Office, Centre for the Protection of National Infrastructure, CESG, Department for Business, Innovation & Skills (first published 2012, updated January 2015) Cyber security guidance for business, including the 10 steps to cyber security https://www.gov.uk/government/publications/cyber-risk-management-a-board-level-responsibility

[8] See British Standards Institute PAS 1192-5:2015 "Specification for security-minded building information modeling, digital built environments and smart asset management". Preparation of the PAS was sponsored by the Centre for the Protection of National Infrastructure

2. **Whether the government has set out an appropriate and up-to-date path for the continued evolution of big data and the technologies required to support it**

The IET and the Royal Academy of Engineering applaud the progress that government has made to open up data. However, government could go further in advocating the opening up of valuable data that remains locked away in some corporate silos, particularly those where there has been significant public investment in the original development of the data. The UK needs to create sustainable markets in data to ensure that both access to and protection of the data is sufficient to support the data-driven economy.

In addition, the government should also consider the following actions:

- establishment of methodologies for the formal valuation of data assets. Without this, it is not possible to measure a key and growing element of national value creation. The consequence is potentially poor decision-making on how best to develop, trade, protect and exploit data assets
- expansion of access to fast broadband to all areas, that is at least compliant with the EU Digital Agenda for Europe targets for universal fixed access. Ubiquitous access to high-speed mobile broadband services is also required
- support for the application of best engineering practice into less mature areas such as software development that is used to generate, manage and analyse data. In particular, regulators will need to ensure that the necessary performance, compliance and resilience frameworks are in place. This will help to make systems less vulnerable to failure
- consider whether the acquisition, processing and storage of some big data represents a strategic asset for the UK, which should require the hosting, processing and storage of such data to be onshore so that adequate safeguards can be put in place and maintained

Legislative and policy framework

Big data technology and devices generating data such as the increasing use of the IoT will place a progressively increasing strain on legislation since currently legislation is relatively slow moving when set against a rapidly evolving technology where development or control is outside of the UK.

The IET and the Royal Academy of Engineering believe that the current legislation and regulation is fragmented and not fit for purpose in the digital age. There is a need for the government to:

-
- ensure that the Data Protection Act (1998) fit for purpose in the big data era
- introduce additional legislative measures to protect citizens' lifestyle data
- ensure that open data policy recognises the risks to UK security at national and individual levels, and develop the necessary legislation, regulation and guidance
- develop legislation, regulation and guidance to ensure that open data and big data do not expose sensitive commercial information and/or intellectual property
- review existing statutes and regulations such as the Environmental Information Regulations (2004), the Freedom of Information Act (2000), the Public Records Acts (1958 1nd 1967) and the Re-use of Public Sector information Regulations (2005) in light of the era of big data
- review whether the future evolution of big data can be anticipated in order to prepare for any future changes in legislation that are needed

- consider if the Information Commissioner needs new powers to regulate the big data market and deter the misuse of big data

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

Big data skills are required urgently to address the gaps in knowledge and expertise. The relevant areas in which these skills are needed include:

- data engineering
- data analysis
- programming and algorithm development
- data governance
- optimum collection standards
- data in manufacturing and production to enable industry to capitalise on big data opportunities

A number of postgraduate courses (as well as some undergraduate courses) have sprung up in data science and other areas related to big data including data science, big data, cybersecurity, cloud computing, machine learning and intelligent systems. Responding to the need for interdisciplinary skills, the courses are often run jointly between different departments, such as statistics and computer science, or mathematical sciences and computer science, although it remains to be seen whether some of these courses will be accredited by multiple professional institutions.

We welcome the new National Occupational Standards created around data analytics by Tech Partnership (formerly e-skills), the Sector Skills Council for Business and Information Technology, include the following:

- design and implement data analysis studies
- manage data analysis services
- lead the data management capability, strategy and framework
- manage the data science capability for data driven business insight.

The *Connecting Data* study identifies a large and diverse range of skills that are needed by engineers for big data and data-driven systems beyond the more specific data skills noted above. These include internet skills, computing, informatics, statistics, mathematics, machine learning, systems engineering, cybersecurity, complex systems modelling as well as mechanical and electronic engineering.

Representatives of the different sectors involved in the study are unanimous in their concern about the lack of multi-skilled recruits, and importance of individuals understanding the engineering or other context in which big data and data analytics are being applied.

Universities, professional institutions and industry need to work together to identify how existing engineering courses should be modified to enable engineers to tackle big data, or new courses created across disciplines. It is possible that some areas of expertise required are already taught – such as systems engineering and statistics - but without the necessary depth or emphasis, or in too theoretical a way.

Other skills that will be required by engineers include the ability to visualise data, and to communicate the outcomes of data analytics with decision-makers who may not have in-depth technical knowledge. In addition, other less tangible but equally important skills include the ability to work in interdisciplinary teams and to address problems from a systems perspective.

Professional institutions also need to consider ways to accredit new or adapted courses. As the expertise resides across disciplines, professional institutions representing the following areas will need to collaborate: information and communications technology (IET), electronic engineering (IET), computer science (British Computer Society), Mathematics (Institute of Mathematics and its applications), statistics (Royal Statistical Society), civil engineering (ICE) and mechanical engineering (IMechE).

Continuous professional development and vocational training are two further important areas that need to be addressed by the professional engineering institutions working together with industry.

The immaturity of this area means that organisations are still exploring different ways to assemble the appropriate expertise to solve big data problems. A challenge will be in assembling the right teams, whether in-house or by working with external data science consultancies, for example, and by recruiting and training individual team members in-house. Lessons learned and career pathways will need to be disseminated by industry and professional organisations working together.

Professional standards for engineering developed by UK-SPEC already include reference to competences that would benefit from an understanding and ability to utilise data[9]. For example, engineers are required to:

- identify, organise and use resources effectively to complete tasks, with consideration for cost, quality, safety, security and environmental impact (EngTech)
- use market intelligence and knowledge of technological developments to promote and improve the effectiveness of engineering products, systems and services (IEng)
- prepare, present and agree design recommendations, with appropriate analysis of risk, and taking account of cost, quality, safety, reliability, appearance, fitness for purpose, security, intellectual property (IP) constraints and opportunities, and environmental impact (CEng).

It is then up to the individual professional engineering institution to interpret the standards for its particular discipline. It is highly likely that specific competences in handling big data will be a serious consideration in future reviews of UK-SPEC.

Gaps need to be filled in the short term by making working in more traditional public sector industries, such as the transport sector, offer comparable remuneration to those in the private sector for similar types of work, such as banking and retail intelligence. In the longer term, gaps need to be filled by promoting and encouraging a better awareness of this type of work as part of the education of engineers.

4. How public understanding of the opportunities, implications and the skills required can be improved, and 'informed consent' secured

The public would benefit from a balanced presentation of the new opportunities that big data could provide both personally and to society as a whole. There is a need to engage the public with the likely impacts of big data on society, on employment and ways of working. Any presentation would need to take into account the difficulty of communicating complex technical issues around big data.

Whilst there may be moderate public interest in big data, in general, there is likely to be little appetite or interest in the subject unless the relevance to them is demonstrated.

[9] Engineering Council (2014) UK-SPEC UK Standard for Professional Engineering Competence http://www.engc.org.uk/engcdocuments/internet/Website/UK-SPEC%20third%20edition%20(1).pdf

It will be important for stakeholders with an interest in data protection at national level as well as individual organisations to develop processes for building trust around informed consent that makes clear the benefits to individuals and society as a whole as well as the risks. At the moment, there is too much focus on the benefits for providers, such as in the case of smart meters. Where relevant, some form of quality certification may help the public gain confidence and belief in both the cybersecurity and data governance of future systems.

Underlying this strategy is the continual need for government, with education and public information providers, to educate all age groups about the importance of privacy and security and the risks that might arise by sharing particular information:

- privacy and security – the role of data protection and informed consent
- failure of informed consent
- exposing lifestyle
- threat of hostile reconnaissance

Security and privacy issues should be addressed by including and informing the public from a position of verifiable evidence rather than speculation.

With increasing use of computer devices such as mobile phones, the take up by the public of any new data based-capability will be rapid once the public are made aware of the advantage to them.

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

The formation of the Alan Turing Institute is a very welcome development, which will focus considerable academic expertise on working to address sector-specific problems in a wide range of economic sectors.

The development work by the Catapults will contribute to ensuring that the foundations of a data-driven economy are in place. However, the approach to handling of data and the role of big data tends to be handled through a series of piecemeal initiatives and projects. For example, the treatment of data, in general, from the “Internet of Things” (IoT) is fragmented and lacks the clarity of data management and governance that is essential for the creation of long-term value.

In particular, this should include exploring new and sustainable business models for data trading and developing standards (with the British Standards Institution) for sector-specific and trans-sector standards for data to provide a firm foundation for data capture, trading and re-use. This standardisation process should address the entire data lifecycle and the need to provide metadata with the collected data so that its utility can be maintained in the medium and longer term.

Funding is also needed to promote the sharing of best practice. The *Connecting data* study identifies that there is considerable scope to learn from what is already being done in other countries, companies and sectors and thus accelerate the evolutionary process.

There are concerns among the engineering profession that funding is currently too focused on capital, and does not give sufficient sustainability to ongoing work. A greater proportion of funding needs to be available to enable researchers to progress in researching big data, rather than restricting it to capital spending. Revenue funding in R&D also increases the potential for results to be commercialised.

End of Submission