Introduction

1. This is a cross-engineering sector response, produced by the National Engineering Policy Centre, an ambitious partnership led by the Royal Academy of Engineering between 39 different UK engineering organisations representing 450,000 engineers. The response was developed by the Royal Academy of Engineering with active input from the Institution of Engineering and Technology, the Institution of Mechanical Engineers and BCS – The Chartered Institute for IT. The response is based on published evidence and on expert opinion from Fellows and Members of these organisations.

2. There are number of tensions that a national data strategy will need to resolve: for example, there is a tension between maximising the economic benefits from data and reducing harm to society. In the public sector, there is also a tension between individual interests – in particular, privacy and data protection – and collective, societal interests – in particular, the efficient and effective delivery of public services. The strategy will need to be clear about how it proposes to resolve these tensions, include resources to regulate effectively and provide the opportunity for genuine public debate.

3. The strategy should explicitly recognise data as an asset. Like other assets, its value depreciates with time. Data assets need owners, and maintenance to keep them relevant. If managed appropriately, data assets have high value. The emergence of many businesses with high valuations reflects this.

4. The potential for smaller companies, non-profit organisations and government to operate in a data-driven economy cannot be addressed in isolation from consideration of the power of the very large data companies and their control over large datasets. These companies have access to uniquely rich sources of data because of their power and reach, which they combine with data from their users. Their ability to attract the best machine learning and data science expertise makes the power imbalance more acute.

Objective 1. To ensure that data is used in a way that people can trust

Research area: opportunities and barriers to trust
1.1. How can organisations (private, public or third sector) demonstrate trustworthiness in their use of data?

5. Data may be commercially sensitive, or it may relate to individuals, with associated privacy requirements as legislated in the Data Protection Act 2018. Concerns about the sensitive nature of data can be a barrier to use of data, unless trustworthiness can be demonstrated. This is true whether an organisation is using the data that it holds itself, or whether data is being shared between organisations.

6. Organisations need to be trustworthy in the use of their data and be able to demonstrate this. The following factors contribute towards an organisation’s trustworthiness:

   - Compliance with legal and ethical requirements.
   - The application of a robust engineering approach, ensuring that data is assembled, structured and managed over its lifecycle so that it meets the various requirements, whether business, ethical or legal, for example.
   - The ability to guarantee the quality, provenance and timeliness of data.
   - Ensuring people with the necessary skills are managing the data.
   - The application of anonymisation or privacy-enhancing technologies, where needed.
   - Secure systems to capture, store and transmit data, bearing in mind that data can not only be accessed but also ‘spoofed’, such that data is falsified intentionally without the knowledge of the data recipient.
   - Ensuring that costs and benefits of collecting, storing and using data are fairly distributed.
   - Mechanisms to control and track how data is accessed and distributed. However, once copied and distributed it is extremely challenging to track data and where it is in its copied format.
   - Defining and agreeing how data is used, with enforcement mechanisms if an agreement is breached.
   - A user-centric approach, with easy-to-understand tools and policies. Privacy impact assessments are one example.
   - Ability for individuals to inspect the data held and control how it is used and when it is deleted.

7. A key consideration is how data is used. Data by itself is not that powerful. It is how it is analysed and presented as information. This is done by tools and applications and these are what the end users interface with. They must be trusted and efficient.

8. For personal data, trustworthiness may involve (but is not limited to) the following: that data is obtained with full consent from the data subject, after the data subject has been fully informed about what the data will be used for, who will have access to it, how long the data will be retained and how it will be destroyed (including any derived or transferred data), what other data it will be combined with, what insights into the data subject or others might be derived from the data, how it will be stored and in which jurisdictions, who it will be transferred to if anyone, what security will be in place to ensure that the data will not be accessed by unauthorised people while it is stored or transmitted, what

---

data breaches the organisation has experienced in the past, and what its processes are for notifying and compensating the data subject in the event of a breach.

9. It is challenging for an organisation to be able to put trustworthiness into practice, but also challenging to demonstrate it has done so and that its processing, use and sharing of data will not change in the future.

10. In certain applications, such as the use of CCTV to monitor public transport, a huge amount of effort goes into ensuring that such systems are compliant with Data Protection regulations, given the potential of the technology to identify individuals.

11. Understanding the provenance of data is particularly important for scientists and practitioners so that it is possible to judge the trustworthiness of the data. It is necessary to understand the conditions under which the data was collected and the assumptions inherent in it. Furthermore, information about the provenance of data is a key part of providing transparency about the operation of systems based on artificial intelligence technologies.

12. An understanding of how people and computer systems interact - ‘human-computer interaction’ - is also vital in building trusted systems.

13. Organisations could demonstrate trustworthiness using a number of means:

   o Being open about data wherever possible. For example, providing transparency around what data is held and by whom, as well as how it is managed and used, to allow external scrutiny.

   o Embracing and signing up to best practice principles around data management and use, which may exist at sector or industry level\(^2\).

   o Demonstrating compliance with legal and ethical requirements. Ethical requirements might be set out in frameworks or codes of conduct. Compliance toolkits could help organisations that are uncertain about how to meet these requirements.

   o Applying the relevant technical and non-technical standards and certifications as means of demonstrating adequate assurance. Technical standards for data might be around metadata, accuracy and timeliness, or around storage and cybersecurity. Non-technical standards might be for organisational processes. There is a need to develop accepted processes for the various sectors and applications.

   o Publishing audits of data management and use. For public companies, this includes enhanced reporting in company reports on the use and curation of data assets.

\(^2\) For example, the Gemini Principles reflect the guiding values around effective information management in the built environment. [https://www.cdbb.cam.ac.uk/Resources/ResourcesPublications/TheGeminiPrinciples.pdf](https://www.cdbb.cam.ac.uk/Resources/ResourcesPublications/TheGeminiPrinciples.pdf)
Developing consumer rating frameworks to provide transparency about the cybersecurity of connected consumer devices³.

Provide facilities for users to report issues of trust and concerns.

14. In the case of data sharing, a third-party organisation or ‘trusted intermediary’ may be involved, with its own business model and governance framework to be sustainable and to underpin trust. A range of organisations may need to be involved in the governance framework in order to underpin trust⁴.

1.2. How easy is it for the public to find about how information provided to or inferred about them by an organisation is being used?

15. It is not always easy for the public to find out how organisations are using their data. Explanations may be disingenuous, or alternatively incomplete.

1.3. Are organisations (private, public or third sector) using personal data in ways that may damage trust?

16. The age of personal data has changed the fundamental nature of customer relationships from one based on mutual trust and aligned interests between organisation and customer to one based on tension between organisation and customer, where personal data and the information that can be derived from it forms the basis of many business models.

17. There is a need for constant vigilance by the public and government to be aware of the latest capabilities for processing personal data and its use by organisations.

18. See also the answer to 1.4 below.

1.4. In what ways are companies making money from personal data? How profitable are these activities?

19. Companies are making money from personal data in the following ways:
   - Using personal data for their own purposes that were not made absolutely clear at the time of subject consent;
   - Sharing personal data with, or selling data, to other organisations without adequate privacy safeguards or protections.

20. A large part of the global internet economy is founded on the use of inadequately anonymised personal data for marketing and other commercial purposes. Many companies make large amounts of money from the use of data without explicit consent and in ways that data subjects do not expect, as was the case with Cambridge Analytica, for example. Facebook has demonstrated that it is possible to monetise people’s personal data and make a lot of money selling on aggregated data which provides a distinct profile of people’s habits and behaviour. It is core to their business model.

21. Other businesses such as energy utilities have different operating models. For example, in the energy sector, utility companies might be able to offer customers

³ For example, the 5Stars project is developing an assurance framework for assessing the cybersecurity of connected and autonomous vehicles to underpin a consumer rating framework that will give UK consumers transparency around a car’s potential cybersecurity risk, [https://5starsproject.com/](https://5starsproject.com/)

a reduction in their bills in return for the monetisation of their customers’ data. It would seem normal in a regulated industry for government to want any benefit gained by the utilities from wider use of this data to be shared with customers.

22. Consumer protection should go hand in hand with big data initiatives. Government must be proactive to protect society at large rather than only acting to do so once it has been demonstrated that people’s personal information is being exploited for others’ gain.

**Research area: concerns around trustworthiness**

1.5. Do people know how information provided to, or inferred about them by, an organisation (private, public or third sector) is being used, stored and shared?

23. The public are unaware, or have only limited awareness, about how information provided to an organisation, or inferred about them by an organisation, is being used, stored and shared.

24. This situation is being exacerbated by the continual introduction of new data analytics and processing techniques, including machine learning.

1.9. How far do existing protections, such as in the Data Protection Act, go in promoting transparency and trust? What, if anything, should the government do to further build trust?

25. One concern is that the internet transcends national and jurisdictional boundaries. It is challenging to regulate some of the large multi-national companies that misuse data.

26. There is a further tension, in that the UK government is keen to attract major players in the internet economy to locate offices in the UK; however, the interests of these organisations are often at odds with those of individual citizens, especially in respect of privacy and trustworthiness.

27. However, many UK citizens appear to be unconcerned about the potential for harm from their data.

28. If GDPR was strictly enforced, it is likely that many uses of data would need to stop.
Objective 2. To ensure that everyone can effectively participate in an increasingly data-driven society

Research area: inclusivity

2.4. What barriers to participation do different groups face? How are marginalised and vulnerable groups affected?

29. Current work on the ‘digitally left behind’ considers obstacles to effective interaction in a digital society. It divides groups of users according to whether they are confident or early adopters (the ‘elite’ community), uneasy with technology but can use some applications (the ‘coping’ community) or those whose are unable to engage. The latter group misses out on seeking and getting better jobs, accessing public services or buying products and services online. Obstacles to effective interaction in a digital society are complex and multi-dimensional, and include age, physical issues, technical complexity, financial limitations and trust.

30. The size of the digitally left behind community is uncertain. More research is needed to understand the underlying causes of the challenge, and to provide insight into key owners and stakeholders that could be involved in developing solutions.

Objective 3. To ensure that all businesses and non-profit organisations can effectively operate in an increasingly data-driven economy

Research area: technological developments

3.4. How and to what extent are small and medium businesses and non-profit organisations contributing to innovation in the data space?

31. SMEs lack the spare resources to investigate and invest in emerging technologies unless these are central to their business focus. SMEs tend to stick to what they know, find a differentiator from their competitors, and tend to pursue a new technology only if it appears vital. Inevitably, SMEs are followers in most areas of technology because their business focus is somewhere else.

32. Because of their numbers, SMEs probably produce more data than the big business players. This data is likely to be distributed and not centralised and in a uniform format. Data standards, by particular industry and process type for instance, are important here.

---

5 Bailey, D., Perks, M. and Winter, C. (September 2018), Supporting the digitally left behind, Ingenia online, Article – Issue 76, https://www.ingenia.org.uk/Ingenia/Articles/930f81f3-4d8a-4e76-b926-1a450d5e7d8f
33. Universities, as non-profit organisations, are contributing to innovation in the data space. There is a growing emphasis on open data and open source, partly driven by UKRI and requirements for data transparency in science. There are several large national initiatives for data repositories and data sharing. For example, JASMIN⁶ provides infrastructure for data analysis in environmental science. DAFNI⁷ is the Data and Analytics Facility for National Infrastructure which supports infrastructure research and strategic thinking for the UK’s long term planning and investment needs. The Consumer Data Research Centre⁸ enables access to data held by consumer-related organisations for the purpose of research. Academic facilities such as the Alan Turing Institute and Science and Technology Facilities Council also play a vital role in contributing to innovation in the data space.

3.5. Are small and medium businesses and non-profit organisations sufficiently benefiting from new emerging technologies such as AI and the Internet of Things (IoT)?

34. Many businesses, both large and small, are not yet benefiting from emerging technologies such as AI and the Internet of Things (IoT). The barriers are to some extent similar for any size of organisation, although smaller business will tend to struggle more to develop the capacity and capability to adopt new emerging technologies.

35. Making curated data more easily available would make it easier for companies to realise the most value from AI. There are many possible routes to making it available.

36. A major barrier to adoption of IoT is the lack of awareness and knowledge about how it can benefit organisations and how to go about adopting and implementing IoT. Technology selection is a challenge for organisations⁹.

37. The cost of developing, implementing and maintaining complex IoT products, services and infrastructures can be difficult to quantify accurately and can vary significantly with customised solutions, producing a key barrier to developing successful business models¹⁰. Although the entry costs of tools have been reduced, the effort required to use them is still considerable, especially where tailored apps are needed. Many of the tools to exploit data are financially beyond the reach of small businesses so shared services are needed.

38. Measures that build the AI and IoT ecosystems or enable ‘matchmaking’ between industry and suppliers of these products and services will help to address these barriers.

39. Continued support through the targeted programmes of UKRI/Innovate UK to SMEs seeking to develop Machine Learning or AI applications within their business operations is important.

---

⁶ JASMIN http://www.jasmin.ac.uk/
⁷ DAFNI https://www.dafni.ac.uk/
⁸ Consumer Data Research Centre https://www.cdrc.ac.uk/
40. Skills will always be limited in small businesses. Skills programmes targeted at SMEs, such as the Made Smarter North West pilot, will also be important, appropriately tailored to reflect sector- or application-specific needs.

3.6. How do businesses envisage that future technological developments will change how they use data?

41. Obtaining value from data is at the heart of emerging technologies such as IoT and AI and is a key business driver. Value will be enhanced through the convergence of different technologies, such as IoT and AI. For example, AI can usefully extract information from the large volumes of data generated by IoT.

42. As the volumes of data generated by IoT increase, robust approaches that optimise the usefulness of the data generated and the potential to create business applications are required. To maximise economic value, organisations must first identify the business needs that require data and then specify the requirements for data management accordingly. Good data management is at the heart of realising the value of technological developments.

43. There is a presumption that AI and machine learning can allow meaning to be extracted from noisy big data. This is a risky presumption as the meaning, provenance and quality of data sets may be ill defined and subject to erroneous assertions. Data engineering needs to precede collection and analysis to remove ambiguity and ensure that the relationships between data items are understood.

44. It is expected that there will be more seamless integration between data and computational modelling/analytics. In a sense this was the goal of eScience a decade or two ago, but there were many failures. Now these workflow technologies are coming of age.

45. Most data usage is today retrospective in that captured data is analysed and patterns established so that aspects of the future, according to these patterns, can be predicted and forecast. There needs to be an increased focus on predicting future events that are ‘new’. Existing data needs to be extrapolated in new ways. The challenge is complex but highly productive if shocks are to be avoided.

**Objective 4. To improve growth and productivity through the effective use of data across the economy**

**Research area:** productivity

4.1. **How is the effective use of data driving business productivity through increased efficiency?**

46. The Royal Academy of Engineering and IET’s report, *Connecting data: driving productivity and innovation*, set out the ways in which better use of data analytics and advanced connectivity would drive productivity and innovation across seven sectors, including advanced manufacturing, built environment, energy, transport, health, aerospace and defence, and insurance. The report explored the future opportunities for organisations and sectors to improve products and processes and to innovate using data analytics. 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. Sections 3.1 to 3.7 of the report set out the opportunities in the seven sectors.

47. The business benefits for built environment, transport and energy sectors are described in Section 4.5 of this document.

48. In sectors that operate physical assets, such as the oil and gas industry\(^ {12} \) and infrastructure sectors\(^ {13} \), the ability to fuse data streams provided by instrumentation systems with powerful, real-time systems simulations (or “digital twins”) is believed to offer benefits that include the rapid identification and remediation of production faults and material flaws, and the ability to forecast future performance growth and limitations.

4.2. What are the barriers to the potential productivity gains from more effective data use?

49. Technical, organisational, regulatory and legal barriers common to all sectors are present, particularly where innovative products and services challenge existing ways of working\(^ {14} \).

50. Many organisations do not know how to value and make best use of the data that they hold. They often do not have the necessary skills in their workforce.

51. There are technical challenges around developing high-quality datasets, reliably linking data sets that have originated in different contexts and ensuring that appropriate security and privacy mechanisms are established. There needs to be accepted processes for assuring the quality of data, or for managing its storage and usage, to underpin trust. Without this, data users may not trust data sufficiently to make decisions based on the data\(^ {15} \). Metadata describing the conditions and assumptions in effect when data is captured need to be recorded and associated with the resulting data set so that future users of the data can assess if the data is appropriate for their use.

52. Where personal data is used, there may be tensions between controlling access to data and business benefits. For example, in the energy sector, there is a question about whether consumers should be allowed to opt out or opt in to the smart meter system. However, the value of the dataset to a utility is less if only partial data is collected.

53. There are particular challenges for data-driven complex or safety-critical systems. In the energy sector, future risks of failure to deliver secure, reliable and stable operations will increase as the complexity of energy system operations increases. There is a need to create responsibility for developing and delivering the specific

---


\(^ {14} \) Royal Academy of Engineering and IET (2015), *Connecting data: driving productivity and innovation*, [www.raeng.org.uk/connectingdata](https://www.raeng.org.uk/connectingdata)

data handling architectures alongside their application into the underlying systems where it is intended the data will be utilised.\(^{16}\)

54. The right conditions need to be in place to allow productivity gains and innovation to occur, and these include appropriate legislation around ownership of data, and adequate access to data, whether open or proprietary.

55. Standards for achieving interoperability between datasets, devices and systems and for building trust in the use of data are also required. Non-proprietary standards are important for enabling innovation in areas such as database technology, data platforms and smart sensor networks. Standardised, open (published) data models and APIs will need to be promoted wherever feasible. Standardisation can help in strategic sectors such as Smart Cities, Energy, Transport and Health.

56. Standards can then be complimented by best practice frameworks that promote frictionless access to data sets together with their exchange and re-use, such as for secure exchange of non-personal data between platforms for B2B or business-government data sharing\(^{17}\).

57. The expected benefits to people, the economy and to government all depend on this data being available for use when and where it is needed. This requires wide coverage by communications technologies at a level and speed sufficient to offer realistic access to users, and thus on the appropriate infrastructure. This certainly does not exist sufficiently widely today and, as noted in the recent infrastructure review\(^{18}\), will require considerable attention if the benefits of data access are to be realised.

58. Individual organisations face challenges in creating new business models and developing the capability to capitalise on data analytics. The costs of applications/tools to exploit data and turn it into information may be prohibitive, particularly if the business case is not apparent. If organisations don’t know what to do with data, why exploit it? This is a hard challenge because organisations are at the early stages of understanding the value of data.

59. 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.

### 4.3. Are there best practices in particular sectors that others can learn from?

60. There are early pockets of excellent practice in data analytics in a number of different sectors 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

---

\(^{16}\) A Future Power System Architect might have this responsibility: IET, Future Power System Architecture Project, [https://www.theiet.org/impact-society/sectors/energy/energy-news/future-power-system-architecture-project/](https://www.theiet.org/impact-society/sectors/energy/energy-news/future-power-system-architecture-project/)

\(^{17}\) This could start with the identification of high-value public data sets followed by standardisation of the associated data models to then making them freely available via APIs in machine readable format to expedite seamless integration into a broad range of downstream services.

practice through all sectors of the economy, particularly those that are only now becoming reliant on data.

61. Examples of best practice across different sectors are provided in Connecting data: driving productivity and innovation (Section 3.8). Towards trusted data sharing: guidance and case studies provides ten examples of best practice in data sharing across a wide range of sectors.

4.4. How do firms develop expertise in their use of data?

62. Before a firm can effectively engage data scientists, it needs to have a reasonable degree of maturity in the way it manages information, including its governance and security. Without these underpinning information and data engineering disciplines considerable effort can be wasted on analysis of poor-quality data.

63. The core expertise in designing the storage, management and governance of data so it is in good condition to use is called information architecture. This is a skill that is in short supply and is rarely taught at university.

64. Data engineering is the expertise to integrate and deliver data to where it needs to be processed. The tools used by data engineers need to automatically collect lineage (provenance) metadata to ensure people can trace where the new copy came from.

65. Broad skills are needed for data analytics in individuals and across teams, which might include some combination of the following: statistics, machine learning, mathematical modelling, complex system modelling, informatics, internet skills, data science and computing. The ability to visualise data is a necessary part of extracting useful information, as well as the ability to articulate information to decision-makers.

66. There is a requirement for people who, although not technical experts, can identify the opportunities to use data analytics in order to drive productivity in business.

67. An understanding is required of the engineering issues to ensure that data management and use meets the necessary requirements including around business, privacy and security. An understanding of the sector or domain is also needed, 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 analysts who do not understand how the data is generated or how the data analysis should be interpreted.

68. In engineering sectors, engineers can develop data science skills alongside their own specialism, but may also find themselves working alongside data scientists who bring specialist skills, as well as an understanding of how data analytics may be applied to business in practice. The ability to work in multi-disciplinary teams, or to lead multi-disciplinary teams, are important.
69. Different skill sets are required to design and operate digitised physical assets, at every level. There is a lack of university and vocational training available to develop these skills.

70. 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. 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.

71. Executive education, CPD and on-line courses are all important means of upskilling. Businesses and other organisations will need to support employees in upskilling, which is demanding given the pace of technological change in areas such as database technology and data analysis techniques.

Research area: societal and environmental benefits of better data use

4.5. In what ways would better use of data provide environmental benefits around key issues such as climate change and biosecurity?

72. Better use of data in sectors that contribute to carbon emissions, such as built environment, transport and energy, has the potential to drive efficiencies and reduce carbon emissions as well as resulting in broader business benefits.

73. There are a number of sector-wide initiatives that aim to improve the way in which data is shared and used across sectors. These initiatives aim to address climate change and improve resilience. The initiatives include the Digital Framework Task Group for the built environment and the Energy Data Taskforce.

74. For the energy sector, the introduction of smart grids and utilisation of data to manage electricity system operations has the potential to be a significant tool within the UK’s overall efforts to deliver a net-zero emissions energy system.

75. Data from pilot and demonstration projects is vital for the development and deployment of new clean technologies. It is important that this data is accessible to enable collaborative learning and accelerate the development of technologies.

---

20 For example, Quantum Black, a data science consultancy, worked with Atkins and Arup on developing tools to monitor the construction of Crossrail. Section 3.8, Royal Academy of Engineering and IET (2015), Connecting data: driving productivity and innovation.
21 For example, Imperial College has recently launched an on-line MSc in machine learning: Imperial launches one of world’s first online Masters in Machine Learning (March 2019), https://www.imperial.ac.uk/news/190523/imperial-launches-worlds-first-online-masters/
22 Digital Framework Task Group, https://www.cdeb.cam.ac.uk/DFTG
24 Royal Academy of Engineering (2015), A critical time for energy policy.
One example is the sharing of test data from batteries in vehicles, to facilitate the development of battery technologies.

76. There is still a gap between what is achievable and current practice in environmental data analytics, despite some significant steps such as JASMIN (see answer to 3.4). The UK could and should be in a position to use its environmental data and models for science and policy analysis.

77. Specific benefits around the use of data in built environment, energy and transport sectors are set out below. These include environmental benefits, but also productivity benefits that are relevant to Question 4.1:

78. **Built environment**: Data has the potential to improve the design and performance of buildings and cities, leading to greater sustainability, more efficient use of space and a higher quality environment. For example:

   - **Procurement**: data feeds from smart cities could in future drive the performance required from an asset. Visualisation of data could assist clients in developing specifications.
   - **Design**: behavioural data from social media and sensors could be used to understand how built environment influences behaviour and to inform evidence-based design that results in a sustainable, productive and healthy built environment.
   - **Operations**: better energy data will allow the energy performance of an asset to be linked to its original construction contract.
   - **Maintenance**: predictive asset maintenance will be possible based on information about component and sub-system degradation.

79. **Energy**: Data from smart meters and other sources will enable an energy system capable of more integrated, flexible and complex transmission, distribution and trading. Energy system data collection, sharing and interoperability are critical for the development of smart grids. For example:

   - **Planning**: local strategic energy system planning will be improved if data from the energy system is combined with other types of data such as on building typologies, construction types and potential energy resources.
   - **Planning**: Accessible and good quality data is also crucial for energy system modelling and confident forecasting. As demand increases with wider electrification, and intermittent renewables and stored energy become a more significant part of the energy mix, it is likely that this will increase in importance.
   - **Operations**: implementing demand side response and management, allowing strategic management of high-power draining appliances (both domestic and industrial), is key to encouraging energy efficiency and demand reduction, and is one of the most productive ways to address energy and climate issues. This is being increasingly addressed through the

---

25 Royal Academy of Engineering and IET (2015), *Connecting data: driving productivity and innovation*, www.raeng.org.uk/connectingdata
work of groups such as the Energy Data Taskforce, whose report\textsuperscript{27} focuses on the importance of ‘presumed open’ data sharing between private suppliers to create whole system efficiencies.

- **Operations**: convergence across different energy networks will be possible, so that the energy system may be considered as delivering services to its end-users rather than as separate electricity, gas or heat networks.
- **Operations**: the operation of new types of energy generation, distribution and local storage will improve, including community energy systems and heat networks.
- **Operations**: optimal decision-making will be enabled including day/hour ahead forecasting and immediate operational management.
- **Operations**: better data will aid in setting tariffs and managing system carbon intensity.

80. **Water and smart cities**: Similarly to the energy system, better management of water distribution in cities in future will rely heavily on data. This will help reduce the risk of flooding in cities and enable more efficient use of water. However, there is still a reluctance for water companies to share data.

81. **Transport**: Data will 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. In particular:

- **Operations**: real-time, multi-modal information provided to passengers via apps\textsuperscript{28,29} will help them plan trips using the most appropriate mode and purchase tickets by mobile phone. New business models along with more sophisticated uses of data analytics will enable ‘mobility as a service’. Increased optimisation of routes and journey sharing will help to reduce congestion.
- **Operations**: vehicle tracking services are already allowing businesses to optimise fuel consumption and reduce running costs for their fleets.
- **Operations**: digital signalling using on-train sensors will allow trains to be controlled in real time, reducing congestion and allowing the railway to respond to peak demand.
- **Operations**: transport providers can monitor passenger movements and passenger crowding, for the purposes of improving efficiency and customer experience.
- **Maintenance**: predictive maintenance will allow transport providers to deploy resources more effectively in managing their assets. This will require integration across the supply chain to ensure that spare parts are located and maintenance services are delivered promptly.

82. While many of the emerging applications are from a single sector, new opportunities will arise as the result of combining data across sectors. For example, energy data could be shared with local authorities to provide indicators

\textsuperscript{27} Energy Data Taskforce (June 2019), *A strategy for a modern digitalised energy system*, https://es.catapult.org.uk/news/energy-data-taskforce-report/

\textsuperscript{28} For example, Citymapper uses real-time data to help users plan trips, www.citymapper.com

\textsuperscript{29} For example, Corethree combines data and mobile technologies to provide mobile ticketing services, bridging the gap between consumer and operator, www.corethree.net.
on fuel poverty, with manufacturers of electrical appliances to improve energy
efficiencies, with third parties to develop energy advice products for consumers or
with water companies to reduce their significant carbon footprint.

4.6. In what ways would better use of data provide social and health benefits
around key issues such as care of our ageing population and wellbeing?

83. Section 3.5 of Connecting data: driving productivity and innovation sets out the
opportunities for the use of data in health.

Research area: broadening data access

4.7. What kinds of data should businesses and non-profit organisations make
openly available? And why?

84. The kinds of data that should be made open will vary from sector to sector, and
should be driven by business or other requirements. A key question is, how would
additional data improve the productivity, efficiency, sustainability or resilience of
a system, business or sector?

85. For example, in infrastructure sectors, the opening up or sharing of operational
data will help to improve the efficiency and resilience, and allow regulators and
others to scrutinise performance. In the energy sector, data sharing between
utilities is needed to enable local demand management, since consumers in a
particular area are able to purchase their energy from many different suppliers.

86. Reference data is typically broadly reusable. This is the type of data that lists all
of the valid values for a data field. For example, all of the countries of the world,
or all of the bus stops in a city, or all of the valid postcodes. These valid values
may have other properties associated with them. The widespread use of common
reference data improves the quality of data and makes the data sets that use
them easier to link together.

87. Historical time series data that shows how particular measurements change over
time, such as weather, can help to detect trends and predict future events. Future
schedules and known events help organizations plan.

88. The value of open data to society and the economy is widely recognised, both in
the UK and abroad. However, open data is not ‘free’ data: robust data
management is needed to ensure data is fit-for-purpose, requiring skilled curation
over time, and is not without cost. Funding or business models will be need to
ensure that costs and benefits associated with data management and use are
fairly distributed between stakeholders.

89. In addition, there may be instances where legal, ethical, commercial, national
security or other reasons dictate that data is not made open. Instead, ways of
securely controlling access or limiting the way in which data is used may be
needed. There are many different approaches to controlled data sharing. These

approaches help to break down data silos and enable access to data and associated data services that add value. They require new workforce skills.

90. The concepts of open and shared data are potentially complex topics for many organisations. While there can be benefits, there can also be significant costs and risks. For example, data aggregation is poorly understood and may result in inferences being drawn that can be prejudicial to public safety and/or security or risk revealing information that is not eligible for production in line with the requirements of the Data Protection Act 2018.

91. It is vital that businesses and non-profit organisations, as well as public bodies, demonstrate appropriate governance in respect of the definition of data, its quality and integrity. Poor quality data is unlikely to be fit-for-purpose for business or investment decision-making.

4.8. Should government encourage businesses and non-profit organisations to make more of the data they hold open? If so, how?

92. Government can play several roles in helping facilitate the opening up of data. It already plays a role in setting standards and making government data accessible and usable. Government can also encourage private sector organisations to collect and release data, leading by example and sharing best practice.

93. Government might build on the successful use of R&D Tax Credits by exploring the use of Data Tax credits to incentivise data sharing.

94. Government is already funding initiatives such as the Energy Data Taskforce\(^{32}\) and the Digital Framework Task Group\(^{33}\) to encourage whole sectors to participate in data sharing. There is potential to extend this support to other sectors, such as manufacturing, and to ensure that the learning is shared between sectors.

4.9. Where appropriate, how might government encourage businesses and non-profit organisations to share more data they hold, where it cannot be made open?

95. Government can improve access to the skills that are required for identifying opportunities for data sharing, and for developing and implementing data sharing models. This requires business skills, and the soft skills required to lead, be part of multidisciplinary teams and work in partnership. It also requires technical skills for developing appropriate architectures and applications of technologies, and for data engineering and linkage.

96. Government can and should enable sharing where it is in the public interest. It can fund platforms and infrastructure for data exchange, as well as pilots for new models of data sharing.

Research area: growth and efficiency underpinned by public sector data

---

\(^{32}\) Energy Data Taskforce, [https://www.gov.uk/government/groups/energy-data-taskforce](https://www.gov.uk/government/groups/energy-data-taskforce)

\(^{33}\) Digital Framework Task Group, [https://www.cdbb.cam.ac.uk/DFTG](https://www.cdbb.cam.ac.uk/DFTG)
4.10. What is the best approach to valuing public sector data in order to reflect its potential to stimulate private sector growth and to offer wider public benefits (financial or non-financial)?

97. Without recognition of the value of data assets, decision-making on how best to develop, trade, protect and exploit such assets is potentially compromised.

98. However, it is a challenge to value data. 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.

99. 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 sensitive34.

100. As techniques for managing and processing data improve, the value of data also increases. There is a ‘data value chain’ whereby the value increases as data is transformed into information, knowledge and ultimately action, as insight is not enough by itself to generate value. The nature of the value chain varies according to the industry, the type of data and the way it is used. For example, in certain contexts raw data may have high value. In other situations, value would only accrue once a degree of analysis had been done on the data. There may be some uncertainty as to what value could potentially be realised as a result of the analysis. Data may be valued in different ways by different stakeholders – for example, each board member will have their own data needs as will each employee. The quality of data also influences its usefulness, and therefore its value35,36. An important aspect of determining value is to understand the provenance of the data, the processes that have created it and the assumptions made while it is being processed.

101. It is important to be able to value data in order to underpin the case for investment in managing and maintaining data.

102. At a series of joint roundtables between the Information Assurance Advisory Council and Royal Academy of Engineering37, the challenge of putting the value of data on company balance sheets was debated and alternative narrative approaches to valuing data that allow organisations to articulate the role of data in reaching business goals was explored. These techniques could be applied to public sector data to articulate the potential to stimulate private sector growth and to offer wider benefits.

103. In particular, scorecards could be used to identify factors that reliably indicate the health and value of an organisation’s information and data assets. A scorecard approach could potentially bring stakeholders together in a collaborative process to develop a mutual understanding of data and create a way of measuring value38.

---

34 Royal Academy of Engineering and IET (2015), Connecting data: driving productivity and innovation.
35 RAEng (2016), Roundtable on data as an asset (unpublished).
36 IAAC (2019), Valuing the information asset (to be published).
37 Discussions at these roundtables are reported in IAAC (2019), Valuing the information asset (to be published).
38 IAAC (2019), Valuing the information asset (to be published).
4.11. How can the public sector quantify, evaluate and weight these benefits in order to determine the terms on which the data could be made available?

104. At its simplest, this may be achieved through quantification of economic benefit in comparison to the cost of making the data available. However, factors that need to be taken into account in assessing costs include the costs and expertise required to:
   o review the nature and content of the data that is being published and provide a data model that describes the individual data items and their relationships;
   o provide ongoing curation of the published data, including providing updates and/or corrections;
   o additional protective measures required by individuals, organisations and physical assets as a consequence of revealing information that may be used to harm them.

105. Assessing the benefits of data publication and sharing is challenging. More work is needed to develop robust methodologies that can be used to justify the benefits against the often significant costs associated with publishing good quality, usable data.