National Data Strategy Consultation
Joint response from the UK National Academies

The Academy of Medical Sciences
The British Academy
The Royal Academy of Engineering
The Royal Society

Introduction

The Academies welcome the National Data Strategy, as representing very important progress towards embedding safe and well-governed data use to meet the needs of society. The Strategy covers many important areas, and in implementing the Strategy it is important that the following priority needs are achieved:

1) Enabling **data access and linkage** across government, on the basis of robust data infrastructure which includes interoperable data standards and architectures, and the use of privacy protections.

2) Embedding a culture of **data responsibility** across government and the wider system, including

- responsibility for data **quality** and accuracy including eliminating bias in data;
- responsibility to use data as a **public good**, for societal benefit;
- responsibility to ensure that its data is used in a socially responsible way with meaningful **public engagement** to build **trust** and trustworthiness.

3) Thereby, building a **trusted system** by embodying **trustworthy** practices. This requires meaningful engagement with the public, including those impacted by the use of data or new data-enabled ways of working, across diverse communities. It involves addressing the harms that can arise from data collection and use, as well as harnessing the opportunities. From this work, understand how the benefits of data can be maximized in an equitable way, and what behaviours are needed by stakeholders to embody trust across society.

4) **Shifting cultural barriers and risk aversion** to data sharing, especially in the public sector, by demonstrating that through good data standards, APIs and secure systems it can be shared safely.

5) Embedding the concept of **data as infrastructure**, and by that highlighting the need to:

- identify, value and maintain data **assets** and the infrastructures that support their use;
- **invest** in the costs of data collection, management and maintenance to maintain these assets – data curation and engineering is vital and requires skilled people and funding;
- ensure the **security and resilience** of that infrastructure, including using the best privacy-enhancing and security technologies;
- take a **strategic** approach to identifying those assets that may be needed in the future;
- ensuring the that the data infrastructure **serves everyone.**

6) Working to make data open in both the public and private sectors, by encouraging the adoption of a **duty to safely share data** for public benefit – such as in response to national and international emergencies, but also for wider societal benefit and to meet societal and policy challenges.

7) Urgently investing in the **skills** needed in the public and private sectors, across the UK, to enable the use of data.

8) Investment in and appropriate use of **digital twins** alongside other valuable data assets.

These are all needed to achieve the missions of the National Data Strategy and can be considered as measures of success in achieving these missions and in establishing the pillars of the Strategy. They should therefore form part of the monitoring and evaluation framework as the Strategy is implemented.

It is important that government is clear about the roadmap and timelines for implementing the National Data Strategy. The missions of the Strategy are of critical national importance – and there is significant social risk in enacting them without due care – so there must be accountability for their appropriate implementation. It is also essential to highlight the role of key bodies such as the UK Statistics Authority, the Office for National Statistics and NHS Digital who are well-placed to enact many of the missions of the National Data Strategy and to ensure that they have the powers and resources to do so. Finally, government must appreciate the significant financial investment needed to implement the Strategy and must be willing to making funding available to support the UK's data infrastructure for the public good.

Summary

Overall:

The Strategy sets out an important set of missions, resting on critical pillars. It could expand its missions by specifically aiming to enable the use of data to address societal and policy challenges; recast the pillars with a focus on data as infrastructure; and ensuring that the twin ideas of responsible data and data responsibility are threaded throughout the Strategy.

Access to data to inform the pandemic response has been challenging, with a lack of frameworks for safe and well-governed use of data across the public and private sector. However, a number of initiatives have been able to establish data systems at pace, and in implementing the Strategy, government should learn from these initiatives and seek to support them beyond the current emergency, considering them as a key part of the UK's critical infrastructure. This could include extending the powers of the Office for National Statistics to be a trusted processor of appropriate public and private data for the purposes of meeting societal needs.

To enable data to be used to effectively address societal and policy challenges, care should be taken to ensure that inequalities do not arise due to uneven data availability or access to computing across regions, communities and individuals. Care should also be taken to ensure that it is aligned with other strategies addressing such challenges, such as NHSX's health and social care data strategy.

There is unevenness in the availability of data and data science skills across the UK – though demand for skills is rising everywhere. The National Data Strategy should seek to ensure that businesses, public services and universities across the UK can attract the right skills and ensure that data to support public service provision is collected and analysed systematically across the UK.

Mission one: Unlocking the value of data across the economy

All sectors stand to gain from better data availability but, rather than focusing on sectors, implementation of the Strategy should prioritise societal and policy challenges such as reaching net zero. Focusing on challenges can also ensure that the implementation of the Strategy is contextualised, reflecting the diversity in kinds of data and purposes of use, and the diverse opportunities and risks presented by data collection, linkage and use.

Government should consider how it can encourage the adoption of a 'duty to safely share data' in emergencies and for the wider public benefit. Public research funders should better fund data collection and data sharing should be incentivised through appropriate recognition by the research community.

Government should consider data across all sectors as critical infrastructure and ensure that there is appropriate investment in maintaining that data and supporting the skills for using it, to ensure that it is used to serve needs across the whole of society, across the UK. That includes support for significant research datasets, such as those used in longitudinal studies.

Support should be introduced for schemes to enable SMEs to access data science skills, for example through pairing SMEs with data scientists in universities. Innovate UK and the Alan Turing Institute could have a role in supporting such schemes. Investing in accessible, ideally online training can provide support for a wide range of organisations.

Mission two: Maintaining a pro-growth and trusted data regime

Technologies can have a role in governance, and therefore investment in the development and use of privacy enhancing technologies can better enable safe use of data for economic, social and cultural benefit. The data protection framework should be reviewed regularly, reflecting changes in public attitudes, in order to maintain trust and trustworthiness. It should also consider the range of rights and controls that are needed to protect the interests of individuals, communities and wider society.

Being trustworthy and securing trust should be central to implementing the National Data Strategy. Many reports and public dialogues have shown the importance that trustworthiness plays in achieving wider strategic goals for the use of data. A key role for the CDEI is carrying out appropriate public engagement and developing best practice for public engagement relating to data and digital technologies. Such public engagement should inform best practice across the system to build and maintain trust.

Mission three: Transforming government's use of data to drive efficiency and improve public services

All of the broad areas of work identified in enabling better use of data across government are vitally important and interconnected. A focus in implementing the Strategy should be on carrying out 'pathfinder' or 'lighthouse' projects which seek to achieve progress across all areas of work and to give confidence that they can be achieved effectively. Again, focusing on challenges can also ensure that the implementation of the Strategy is contextualised, reflecting the diversity in kinds of data and purposes of use, and the diverse opportunities and risks presented by data collection, linkage and use.

Mission four: Ensuring the security and resilience of the infrastructure on which data relies

The value of kitemarks, and other ways for businesses to navigate the marketplace for data infrastructure and to invest in genuinely beneficial products, should be explored further. Appropriate operational frameworks, supported by technical guidance, are an important tool in risk management for data centres. Some form of 'testing and exercising' regime will be crucial to assess performance and resilience capacity.

Mission five: Championing the international flow of data.

The Academies support seeking EU 'data adequacy', as if the UK's data regulations are not aligned with the EU's, sharing data with researchers based in EU countries could become more difficult. This could have significant negative consequences for UK research and innovation, for the outcomes and safety of UK patients, and for the UK's contribution to global health, research and innovation efforts. Given the

interconnected nature of the challenges that data can help to address – from pandemic response to reaching net zero – the need for a critical data infrastructure has to be addressed at the international level, by reducing barriers to data sharing while appropriately managing the risks.

1. To what extent do you agree with the following statement: Taken as a whole, the missions and pillars of the National Data Strategy focus on the right priorities. Please explain your answer here, including any areas you think the government should explore in further depth.

The Strategy sets out an important set of missions, resting on critical pillars. It could expand its missions by specifically aiming to enable the use of data to address societal and policy challenges; recast the pillars with a focus on data as infrastructure; and ensuring that the twin ideas of responsible data and data responsibility are threaded throughout the strategy.

Missions and challenges: The missions of the National Data Strategy largely focus on the right priorities, coinciding with the priority issues identified above. It is essential to ensure each of these missions is pursued in a safe and well-governed way, building public trust in the means by which they are achieved. In addition to the listed missions, the National Data Strategy should focus on the use of data to address societal and political challenges – and indeed a focus on specific challenges is a way of carrying out the kinds of pathfinder or lighthouse projects that can demonstrate the possibility of safely unlocking the value and using it for public benefit; and can also enable government to better understand the different opportunities and risks relating to different kinds of data. Focusing on the cross-governmental and cross-sectoral sharing and use of data to reach the net zero target is an important challenge area that could be a major focus in the implementation of the Strategy. A crucial aspect of this will be sharing of data between government and industry and will require good industry engagement.

Pillars and priorities:

- Skills: The Strategy is right to focus on skills, with demand for data skills rocketing – as evidenced by the Royal Society report *Dynamics of Data Science Skills,* cited in the Strategy. There is a chronic shortage of data skills in the workforce, including within government. A people strategy is vital in order to implement the National Data Strategy. In addition, as the potential of data science continues to be realised it is clear that work on and with data requires not just skill in statistics and computing but also analytical skills, communication skills, and domain-specific knowledge across all subject areas to understand where and how data science techniques can be applied.
- Data as infrastructure: Following the introductory comments, the focus on *data foundations* would be better cast as a focus on *data infrastructure*. Thinking of data as infrastructure not only highlights the needs for quality, reliability and standards, it can also foster the approach of treating data as a critical national asset that serves the public benefit. It can highlight the value of data, but also the *costs* of maintaining data – and highlight how important investment in data maintenance is.

- Data availability: We have learned from the Covid-19 pandemic that timely access to data is of critical importance, but there are often barriers to accessing data or making it available. Implementing the Strategy should consider all means of enabling data use that allow the safe and appropriate navigation of often legitimate barriers to data availability. For example, this should include using privacy enhancing technologies to enable data use or learning across data sets, without moving, linking or sharing data.

- Responsible data: responsible data, and data *responsibility*, is a critical pillar in the Strategy. Fostering a mindset of data responsibility not only encourages important focus on the socially

responsible use of data, but should encourage data owners to acknowledge that they are responsible *for* their data, for its quality, accuracy, comprehensiveness, lack of bias and for its long-term maintenance. It can also focus on the responsibility that data owners have to *enable the use* of data for social benefit.

Managing tensions and dilemmas: There are a number of key tensions between these missions, eg between maximising the economic benefits from data (addressed in Mission 1: unlocking the value of data across the economy) and reducing harm to society (addressed in Mission 2: maintaining a pro-growth and trusted data regime). In the public sector, there is also a tension between individual interests – in particular, privacy and data protection (Mission 2) – and collective, societal interests – in particular, the efficient and effective delivery of public services (Mission 3).

These kinds of tensions and dilemmas were highlighted as a set of challenges raised by new uses of data, in the British Academy and Royal Society report *Data Management and Use: governance for the 21st Century*. The Academies argued for transparent and inclusive means for navigating trade-offs in data governance. The implementation of the National Data Strategy will require detail on how public dialogue with widest society will inform how to resolve the tensions that arise between unlocking the value of data and pursuing responsible data – including protection from data harms.

2. We are interested in examples of how data was or should have been used to deliver public benefits during the coronavirus (COVID-19) crisis, beyond its use directly in health and social care. Please give any examples that you can, including what, if anything, central government could do to build or develop them further.

Access to data to inform the pandemic response has been challenging, with a lack of frameworks for safe and well-governed use of data across the public and private sector. However, a number of initiatives have been able to establish data systems at pace, and in implementing the Strategy, government should learn from these initiatives and seek to support them beyond the current emergency, because they should be considered a key part of the UK's critical infrastructure. This could include extending the powers of the Office for National Statistics to be a trusted processor of appropriate public and private data for the purposes of meeting societal needs.

Access to both health data and data relating to everyday interactions was challenging during the pandemic. The Royal Society's RAMP and DELVE initiatives found that timely access to data has been the greatest barrier to providing the best possible scientific advice and has frustrated exploiting the UK's extraordinary data science capability to fully support the response to the pandemic. This includes access to data such as mobility and transport data, held largely in the private sector. (Details on the experience of enabling the use of data for Covid response, by activities convened by the Royal Society, are given in detail in a separate response to the Strategy from the Royal Society, submitted as an addendum to this response.)

Linking datasets provides better insights. For example, linking health data to non-health data sources could support understanding of the impact of COVID-19 on health inequalities and help inform and direct local responses. Therefore, it is essential that there is close alignment between the National Data Strategy and strategies for health data specifically.

The role of data in enabling resilience – of organisations, infrastructure or supply chains – has been highlighted during the pandemic. There are some positive examples of data use in this context of the pandemic, as follows.

Resilient supply chains

Data and digital technologies can help build resilience of supply chains, by helping to anticipate possible disruption and enabling better visibility and monitoring of the supply network. Data can inform decision-making around uncertainties, priorities and trade-offs. During the COVID-19 crisis, digital technologies played a role in enhancing the visibility of available capacity across critical supply chains which enabled robust assessments of data on vulnerabilities and resilience to be carried out.

For example, in food supply chains, there is a highly decentralised system of information and data. Through sharing data, it would be possible to test and model scenarios related to potential future shocks related to the UK leaving the EU, the COVID-19 pandemic, or cyber-attacks.

Data availability, skills and infrastructure have been integral to identifying and creating opportunities to make better use of data during the COVID-19 crisis. It is also critical that there is visibility and transparency about where information on supply chains is sourced, how these sources are managed, how they are analysed and communicated. These opportunities would benefit further from joined up policy decisions which span different regulatory regimes.

Infrastructure resilience

Another area is data sharing between infrastructure sectors to improve resilience. Given increased complexity and interdependence, resilience depends upon clear understanding and communication across sectors, organisations and stakeholders. Covid-19 has highlighted the importance of local information, networks, skills, and clear channels of communication, for example better identification, understanding and reach to those groups who are particularly vulnerable in different scenarios. Again, given the potential combined impacts of Covid-19, Brexit and winter weather, it will be vital to ensure effective and rapid data and information sharing.

Access to data is vital for modelling failures in interdependent infrastructure systems in order to understand the knock-on effects of these failures into supply chains, business interruption and the economy, which scale up as the disaster gets bigger i.e. there are multiplier effects.

The limitations of the NHS's digital infrastructure posed initial challenges for some testing sites due to the difficulties in sharing data related to COVID-19 testing with NHS and social care settings. The Test, Trace and Isolate system is underpinned by IT and digital capability which must be invested in and maintained to ensure it is fit for purpose. Strong digital infrastructure underpins the COVID-19 testing response and is vital to its success and expansion.

The role of data in enabling resilience is explored further in the National Engineering Policy Centre's recent paper on vulnerabilities associated with infrastructure interdependencies during the pandemic: *Winter is coming: risks for interdependent infrastructure*. The need for technologies such as data to underpin resilience was also emphasised in the Royal Academy of Engineering's submission to the Integrated Review of Security, Defence, Development and Foreign Policy.

Duty to safely share data

The experience of the Covid-19 response, both positive and negative, highlights the need for the private and public sectors to adopt a duty to safely share data in an emergency, and indeed for ongoing public

benefit. It is however essential that this sharing and linking of data is done in a safe and well-governed way to protect the rights of data subjects and commercial interests of companies. Extending the powers of the Office for National Statistics as a trusted data processor for appropriate types of data could be a key step in delivering that public duty safely.

3. If applicable, please provide any comments about the potential impact the proposals outlined in this consultation may have on individuals with a protected characteristic under the Equality Act 2010?

To enable data to be used to effectively address societal and policy challenges, care should be taken to ensure that inequalities do not arise due to uneven data availability or access to computing across regions, communities and individuals.

Action is needed to ensure that these proposals do not widen health, accessibility or social inequalities due to protected characteristics. Robust means are needed to identify areas where gaps in data or the way that data is collected might impact specific groups or communities

Clear and accessible information should be developed on the proposals and the implications they have for the public. People with protected characteristics should co-produce the proposals and the communications about the National Data Strategy.

Care should also be taken to ensure that the Strategy is aligned with other strategies addressing such challenges, including the NHSX's health and social care data strategy. The implementation of the National Data Strategy should also be informed by the important work of organisations and individuals that have carried out in-depth research and engagement on the inequalities that can arise or be exacerbated by data collection and use.

The Strategy should also consider inequalities arising due to different regional capacity to make beneficial use of data (see below). Inequalities due to variation in access to computing resources by regions, communities and individuals should also be considered in implementing the Strategy; as well as impacts on communities and individuals who do not engage with digital services – such as banking and telecommunications. Taking a data as infrastructure approach should be understood to involve a responsibility to ensure that the infrastructure serves everyone in society.

4. We welcome any comments about the potential impact the proposals outlined in this consultation may have across the UK, and any steps the government should take to ensure that they take account of regional inequalities and support the whole of the UK.

There is unevenness in the availability of data and data science skills across the UK – though demand for skills is rising everywhere. The Strategy should seek to ensure that businesses, public services and universities across the UK can attract the right skills and ensure that data to support public service provision is collected and analysed systematically across the UK.

Research by the Royal Society in *Dynamics of Data Science Skills* showed that demand for data science skills among employers is still strongest in London, but exhibiting strong growth in other regions. Regional breakdowns show the dominance of London for Data Scientist and Advanced Analysts, accounting for 58% of all postings in 2017/18. However, for Data Scientists and Advanced Analyst job vacancies, growth was larger (relative to the base amount) in Northern Ireland (563%), the North West (269%) and the East of England (250%).[1]

Where regional skills gaps exist, universities with good industry links have a key role to play in developing appropriate professional training. Employers have a role in upskilling the workforce by training existing employees, particularly those at risk of losing their jobs through automation, and can work with universities to co-produce training. By working in collaboration with employers, universities can potentially address regional skills gaps and address productivity needs. This could involve working across professional disciplines to understand the type and level of data science skills that will be needed by professionals in fields such as law, healthcare, and finance. There is a clear role for organisations such as Innovate UK to support collaboration between universities and businesses to grow skills across the UK. Examples can be learned from the collaborative upskilling projects undertaken with SMEs and regional businesses by the National Innovation Centre for Data.

As highlighted in the Academy of Medical Sciences' report, *'Our data-driven future in healthcare'*, the government should work to ensure that the benefits of data-driven approaches are felt equitably across the UK population. Addressing existing regional inequalities will require infrastructure and educational support for companies, universities and healthcare institutions across the country to enable them to seize the opportunities offered by data-driven approaches. For example, the recent Academy of Medical Sciences and Health Data Research UK workshop *Realising patient and NHS benefits from health and care data – from policy to practice* highlighted that non-research hospital trusts may be disadvantaged in seizing these opportunities as they may lack the existing infrastructure or expertise required. This could result in missed opportunities for research and innovation. Regional inequalities can also arise through differential levels of data collection in other sectors. For example, if much more data is available about transport in London, due to initiatives such as TFL's data store, this will lead to research focusing on areas which are data rich. Further considerations include the potential bias if approaches towards investment in data infrastructure assess spending in terms of 'value for money', which tends to bias towards areas where more people live, or where wealthier people live.

Mission one: Unlocking the value of data across the economy

5. Which sectors have the most to gain from better data availability?

All sectors stand to gain from better data availability but, rather than focusing on sectors, implementation of the Strategy should prioritise on societal and policy challenges such as reaching net zero. Focusing on challenges can also ensure that the implementation of the Strategy is contextualised, reflecting the diversity in kinds of data and purposes of use, and the diverse opportunities and risks presented by data collection, linkage and use.

All sectors, across the whole of the UK, stand to benefit from greater use of data. The role of data could more be more helpfully considered in relation to tackling specific societal challenges or driving specific outcomes, which may cut across many sectors and require cross-sectoral sharing of data. Indeed, using data to address societal and policy challenges could helpfully be elevated to a core mission of the National Data Strategy, with a greater role for the 'lighthouse projects' by Cabinet Office and the ONS as a means for addressing all pillars of the Strategy. Taking this approach will also help to ensure that, in implementing the Strategy, the specific opportunities and potential risks in using diverse kinds of data are properly understood and addressed.

Data for net zero: The Royal Society's (forthcoming, 3 December) report on *Digital technology and the planet* highlights that data can help reduce emissions and achieve net zero by underpinning applications and services across sectors. Several studies showed that existing digital tech applied across sectors

could contribute nearly a third of the 50% reduction in emissions necessary by 2030. In addition to reducing carbon consumption, digital technologies also have a role to play in improving operations and research. Data will also increasingly play a role in enabling low-carbon and resilient infrastructure systems and in the transition to the future energy system, as highlighted in the National Engineering Policy Centre's paper *Beyond COVID-19: Laying the Foundations for a net-zero recovery*.

'Data for net zero' should be a key focus in the effort to improve data availability – and could be a specific mission of the Strategy. The National Digital Twin programme is doing important work towards an Information Management Framework. But there needs to be a wider, concerted effort towards the digitalisation of the net zero transition, including improving the availability of relevant data, from across sectors. Systems approaches to complex policy challenges, that highlight the interconnectivity between different sectors, technologies and areas of policy, can help to identify where effective data sharing will be needed. The National Engineering Policy Centre's paper *Net zero: a systems perspective on the climate challenge* sets out the central role of a systems approach in addressing this complex area of policymaking.

Data for health and care: The Academy of Medical Sciences' report, *'Our data-driven future in healthcare'* sets out principles for the development and deployment of health data-driven technologies that will maximise benefit for society while respecting the views of the public and building trust. The Royal Society and Academy of Medical Sciences workshop on *AI in Health and Social Care, from Bench to Bedside* highlighted the potential impact of AI in improving healthcare, from diagnostics to supporting social care. But it also highlighted deficiencies in the data systems that underpin health care, for example, the varying digital maturity of systems across the country, which means that data is very often messy and in different formats.

Further challenges include improving infrastructure resilience, or the levelling up agenda. There could be links between the National Data Strategy and other sector-specific strategies or missions.

While all sectors would benefit from greater data availability, the way in which they will benefit is likely to differ. The use of data may change underlying business models or practices (such as in the insurance sector) and open up the sector to entirely new entrants with implications for competition. The role of government will need to vary according to the existence and nature of any market failures, but also the potential economic, societal and environmental opportunities and risks.

It will be useful to draw on learning from government-supported initiatives such as the Centre for Digital Built Britain / Information Management Framework and Energy Data Taskforce, and their relevance to other sectors. This might include understanding the barriers to adoption across the sectors in which the frameworks have been developed. There is potential to learn lessons across sectors, and between public and private sectors. The challenge will be to spread examples of best practice through all sectors of the economy, particularly those that are only now becoming reliant on data.

6. What role do you think central government should have in enabling better availability of data across the wider economy? If yes, what is it? If not, why not? How does this vary across sectors and applications?

Government should consider how it can encourage the adoption of a 'duty to safely share data' in emergencies and for the wider public benefit. Public research funders should better fund data collection and data sharing should be incentivised through appropriate recognition by the research community.

In its *Machine Learning* report the Royal Society argued for the need for an 'amenable' data environment to make use of machine learning technologies. As part of this a need was identified for policy frameworks or agreements which make data available to specific users under clear and binding legal constraints. This idea was taken up in the Hall and Pesenti review, which called for exploration of the 'data trusts' model to support legal frameworks for data access.

In parallel, the Royal Society and British Academy *Data Management and Use* report called for ways to navigate the tensions between incentivising innovative uses of data while ensuring that such data can be traded and transferred in mutually beneficial ways. This concerned enabling the sharing of data across commercial organisations for the wider promotion of innovation while protecting commercial interests.

The Royal Academy of Engineering's report, *Towards trusted data sharing: guidance and case studies* illustrated multiple technical, economic and governance issues - highlighting the friction involved in data sharing - and set out a practical checklist to help organisations navigate through the issues.

Work on creating frameworks for such sharing of data should be a major focus for the implementation of the National Data Strategy. Existing frameworks for health data, such as those set out in the Academy of Medical Sciences' report 'Our data-driven future in healthcare', and in the Office for Life Sciences' Life Sciences Sector Deal 2, could support new frameworks applying to other sectors or types of data.

Government should consider how it can encourage the adoption of a 'duty to safely share data' for societal benefit. This could be by extending the powers of the Office for National Statistics to require, where possible, that private companies enable appropriate data they hold to be used for public benefit, in ways that do not pose a risk to privacy of individuals and communities, or risk to the commercial interests of those companies. Enabling researchers to access such data especially in response to emergencies or societal challenges will greatly improve the ability to unlock the public value of data.

Ministers should engage with the leaders of large companies holding publicly valuable data assets to enable their safe use for research that can improve public services – such as research into transport systems – building on the provisions of the Digital Economy Act that allow the ONS to access such data for its purposes. Government has an opportunity to develop licenses across all sectors in which the duty to safely share data can be specified.

In the research sector, incentives are needed for sharing data, to deter a 'data hugging' mindset. UKRI should be supported in its efforts to treat data as research infrastructure, with appropriate funding to support its maintenance and use. Health Data Research UK's Health Data Research Hubs will provide further insights into potential models for how this can be achieved, and institutions such as the Alan Turing Institute are well placed to support this.

7. To what extent do you agree with the following statement: The government has a role in supporting data foundations in the wider economy. Please explain your answer. If applicable, please indicate what you think the government's enhanced role should be.]

Government should consider data across all sectors as critical infrastructure and ensure that there is appropriate investment in maintaining that data and supporting the skills for using it, to ensure that it is used to serve needs across the whole of society, across the UK. That includes support for significant research datasets, such as those used in longitudinal studies.

If government adopts a *data infrastructure* approach it should lead government to consider the data assets that already exist, and how these are maintained. The UK's data assets include significant research datasets that should be maintained and used in the public interest. As an example, the ESRC have supported the establishment of the Consumer Data Research Centre, based at University of Leeds and UCL. These collect a wide array of consumer data—such as loyalty card data—in a secure and

controlled way. There is great potential in these datasets, but awareness and access to them within government appear limited.

Making the best use of research data includes distinguishing between longitudinal and cohort study data. Great value could be derived from linking these, but an overly restrictive approach to administrative data makes this very difficult at present. For example, while great value could be derived from linking schools data with that from the Millennium Cohort Study, gaining access is burdensome presenting a barrier to timely use. An audit of such highly valuable data resources should be conducted to identify what already exists and how better use could be made of them. Facilitating such linkages and ensuring safe access should be a priority for Government.

The government has a key role in ensuring fairness, transparency and trustworthiness are maintained. If the data system, and those working in it, are not trusted, it will undermine all efforts to capitalise on the promise of data-driven approaches. Instances of bad practice come with the risk of harming the reputation of the entire system and not just those involved. This is particularly important in instances of private companies working with NHS data, as this has been shown to be an especially sensitive area in the views of the public. Public dialogue work commissioned by the Academy of Medical Sciences' in 2017, and by Understanding Patient Data and the Ada Lovelace Institute in 2019 have shown that while the public are generally very supportive of their health data being used for public benefit, where private companies are involved they have heightened expectations for transparency and privacy.

Government can lead by example and share best practice, encouraging private sector organisations to collect and maintain high quality data. Where government carries out 'pathfinder' or 'lighthouse' projects to connect and use data, lessons can be learned about what it is possible to do in a well-governed way. As argued in the Royal Society's *Protecting Privacy in Practice* such government-led projects can also make use of the technologies that enable secure and privacy-preserving use of data. Government carrying out these lighthouse projects can also stimulate research and development in privacy enhancing technologies.

Ensuring the UK has the highly skilled workforce to deliver on the promise of data-driven approaches should be a priority for the government, as it is not guaranteed that this training will happen to the level required without this support. In addition to training, viable career routes in research need to be maintained given the demand from the private sector.

Government can improve access to the specific skills that are required to ensure that data foundations or infrastructure are robust, for example:

- 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.

- 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. Skills in engineering and cleaning data are in high demand and a great deal of resource is spent on data cleaning. The Royal Society's report on *Dynamics of Data Science Skills* showed a steep rise in the demand for these skills and an associated increase in salary (higher than for similar roles).

8. What could central government do beyond existing schemes to tackle the particular barriers that small and medium-sized enterprises (SMEs) face in using data effectively?

Support should be introduced for schemes to enable SMEs to access data science skills, for example through pairing SMEs with data scientists in universities. Innovate UK and the Alan Turing Institute could have a role in supporting such schemes. Investing in accessible, ideally online training can provide support for a wide range of organisations.

SMEs, universities and research institutes are potentially at a disadvantage when trying to use data effectively compared to very large companies or institutions – and even more so public and third sector organisations. This is particularly true for health data which requires high standards in data security and governance in order to gain access to datasets as highlighted in the BioIndustry Association and Medicines Discovery Catapult's 2018 report, *State of the Discovery Nation 2018*. This has infrastructure and administration implications that may be too costly or arduous for smaller companies and institutions.

SMEs cannot compete for talent with larger companies. They also have very short timescales for a return on investment in skills. A positive approach is pairing SMEs with data scientists based in universities to solve their problems collaboratively – pairing data scientists with domain experts that 'own' the data and benefit from its use. Innovate UK has a role in supporting collaborations between SMEs and universities to support this knowledge sharing.

It may also be difficult for these organisations to achieve the high data standards in order for their data to be included in national or international endeavours. However, not meeting all criteria for high data standards does not mean that these data sets are not useful or do not have value. Therefore, criteria for meeting these standards may need to take into consideration the restrictions these organisations may have and seek ways to include them when appropriate.

Business support programmes to encourage adoption of better data use that are targeted at SMEs, such as the Made Smarter North West pilot, will be important, appropriately tailored to reflect sector- or application-specific needs. Any interventions need to be tailored to meet the needs of SMEs, and communicated to them in a way that demonstrates their usefulness for their specific situation.

Discussions at a recent joint workshop between the Royal Society, British Academy, Open Data Institute and Ada Lovelace Institute looking at the potential for use of data by charities and not for profits argued that central government should provide more opportunities for organisations to receive resources and financial support to improve their data practices, through organisational change and access to training.

Mission two: Maintaining a pro-growth and trusted data regime

10. How can the UK's data protection framework remain fit for purpose in an increasingly digital and data driven age?

Technologies can have a role in governance, and therefore investment in the development and use of privacy enhancing technologies can better enable safe use of data for economic, social and cultural benefit. The data protection framework should be reviewed regularly, reflecting changes in public attitudes, in order to maintain trust and trustworthiness. It should also consider the range of rights and controls that are needed to protect the interests of individuals, communities and wider society.

Technology has a role as part of the governance toolbox. As the Strategy notes, the Royal Society's *Protecting privacy in practice* report outlined how a set of emerging technologies, privacy enhancing

technologies (PETs), can help use and share data while protecting sensitive information. These technologies can provide new ways to address the goals of data protection regulations, such as the need for appropriate safeguards - although this should not detract from the need for legal frameworks and for assessing whether a given use of data is ethical in the first place.

Protecting privacy in practice identified a need to support organisations to become intelligent users of PETs. It recommended that Government, public bodies and regulators raise awareness further and provide guidelines about how PETs can mitigate privacy risks and address regulations such as GDPR. For example, the Information Commissioner's Office (ICO) should provide guidance about the use of suitably mature PETs to help UK organisations minimise risks to data protection, and this should be part of the ICO's Data Protection Impact Assessment guidelines. Such guidance would need to cover how PETs fit within an organisation's overall data governance infrastructure, since the use of PETs in isolation is unlikely to be sufficient. The ICO has been updating its guidance on anonymisation including case studies on PETs.

Because of the fast-moving context, periodic reviews are needed to ensure the relevance of the data protection framework to new innovations in the field. Considerations of shifting public opinions and maintaining public trust are essential areas that should be reconsidered periodically in the context of revising regulations. Public engagement and dialogue should be an ongoing process to ensure that any shifts in opinion are recognised and can be addressed.

The British Academy and Royal Society also highlighted the importance of thinking beyond data 'ownership' and considering the rights and controls that are needed to protect individuals, communities and wider society, in its workshop with techUK on *Data Ownership, Rights and Controls*. A number of individuals and organisations are active in understanding the rights of groups in relation to data, and this should be reflected in the evolution of the data protection framework.

11. To what extent do you agree with the following statement: the functions for the Centre for Data Ethics and Innovation (CDEI) should be Artificial Intelligence (AI) monitoring, partnership working and piloting and testing potential interventions in the tech landscape? How would a change to statutory status support the CDEI to deliver its remit?

Being trustworthy and securing trust should be central to implementing the National Data Strategy. Many reports and public dialogues have shown the importance that trustworthiness plays in achieving wider strategic goals for the use of data. A key role for the CDEI is carrying out appropriate public engagement and developing best practice for public engagement relating to data and digital technologies. Such public engagement should inform best practice across the system to build and maintain trust.

The Royal Society and British Academy's *Data Management and Use report* recommended governance framework for data management and data use should perform three broad categories of functions. These may be carried out by a variety of public and private actors:

- Anticipate, monitor and evaluate
- Build practices and set standards
- Clarify, enforce and remedy

Despite the range of actors already carrying out some of these important governance functions in their specific sectors or domains, it was argued that a new body – effectively the Centre for Data Ethics and Innovation – steward the landscape as a whole. The role of the CDEI closely mirrors those functions.

While taking a cross-landscape point of view, the CDEI must be sensitive to sectoral differences and must consider sector-specific implications when giving high-level advice. The medical research and healthcare sector often has unique requirements and challenges, as highlighted in a recent roundtable report on AI in healthcare from the Academy of Medical Sciences, the Medical Research Council and the National Institute for Health Research.

All activities that the Centre undertakes must involve and be informed by the public. It should explore the use of AI, data-driven and digital systems, and algorithms in the context of public views and benefits to society, and should engage in meaningful public engagement to do that. It should share examples of good practice to enable other organisations to do so effectively. The Centre should seek to play a role in reducing inequalities and promote the use of digital systems in ethical, equitable and societally beneficial ways.

Issues surrounding data that the Centre may wish to scrutinise, include, but are not limited to: the audit and integrity of data; Intellectual Property; liability; data privacy; data access; and provenance or context, which may be a key component to discussions around its utility, integrity and reliability. More generally, one way in which the government could increase trust in its own handling of data and use of AI would by an independent body actively monitoring and holding public sector bodies to account. This may be CDEI or another body such as the Information Commissioner's Office or the National Audit Office, if given appropriate powers.

We recognise that a statutory footing will be valuable if it creates a clear line of communication between the Centre and Government and gives legitimacy and credibility to the Centre's advice. That said, the Centre's independence is key to its value and credibility. A statutory footing must not compromise this independence and assurances will be needed about how it retains this.

Mission three: Transforming government's use of data to drive efficiency and improve public services

12. We have identified five broad areas of work as part of our mission for enabling better use of data across government. We want to hear your views on which of these actions will have the biggest impact for transforming government's use of data.

- **Quality, availability and access**
- **Standards and assurance**
- **Capability, leadership and culture**
- **Accountability and productivity**
- **Ethics and public trust**

All of the broad areas of work identified in enabling better use of data across government are vitally important and interconnected. A focus in implementing the Strategy should be on carrying out 'pathfinder' or 'lighthouse' projects which seek to achieve progress across all areas of work and to give confidence that they can be achieved effectively. Focusing on challenges can also ensure that the implementation of the Strategy is contextualised, reflecting the diversity in kinds of data and purposes of use, and the diverse opportunities and risks presented by data collection, linkage and use.

The COVID-19 pandemic has shown just how important **timely access to quality data** (point 1) is in order to use data for public benefit. Better insights could be derived from linking datasets, harmonizing

survey data, facilitating comparative and pooled analyses, and ensuring cross-national access to data. For example, linking health data to non-health data sources could support understanding the impact of COVID-19 on health inequalities could help direct local responses.

Access to good data also will play an important role in nurturing the development of data skills. It can ensure that data scientists get necessary experience with 'real world' problems that is so important in data science. But more importantly, this will enable the use of data science skills for public and commercial benefit.

However, this is not independent of the other concerns. Common standards are essential to ensuring that data is shareable and appropriately linkable – though both accessing and linking data must be well-governed with checks to ensure that it is in the public interest. Ethics and public trust are also absolutely central to the availability of and access to data, and require careful consideration. Loss of trust in any one area has the potential not just to undermine in individual actors or institutions, but could undermine the whole National Data Strategy.

These issues are therefore inextricably interconnected, and the implementation of the Strategy could focus on securing the use of data in addressing specific societal or policy challenges. Setting these up as 'pathfinder' or 'lighthouse' projects can enable learning on how *all* of these areas of work can be addressed, across diverse areas of data use, and diverse types of data.

The Strategy should acknowledge the difference between data to support operations and services and data to inform policy, and ensure that there is appropriate engagement to secure trust in both of these respects.

Mission four: Ensuring the security and resilience of the infrastructure on which data relies

14 - What responsibilities and requirements should be placed on virtualised or physical data infrastructure service providers to provide data security, continuity and resilience of service supply? How do clients assess the robustness of security protocols when choosing data infrastructure services? How do they ensure that providers are keeping up with those protocols during their contract?

The value of kitemarks, and other ways for businesses to navigate the marketplace for data infrastructure and to invest in genuinely beneficial products, should be explored further. Appropriate operational frameworks, supported by technical guidance, are an important tool in risk management for data centres. Some form of testing and exercising regime will be crucial to assess performance and resilience capacity.

The Royal Society reports on *Protecting Privacy in Practice* and on *Progress and Research in Cyber Security* argued that standards and kitemarks are needed for quality assurance and to increase 'buyer confidence' in privacy enhancing technologies, and in cyber security. Currently privacy standards are unclear and guidelines are scarce. Even though there is a lot of research on standards and processes, currently they are not mature enough for cross-sector agreement on best practice. In the UK, the reviewing of PETs and provision of kitemarks by a trusted authority such as the National Cyber Security Centre (NCSC) would give more confidence to companies and their customers. Trustworthy standards and appropriate guidance will further drive a culture change that goes beyond a 'sticking plasters' approach and would build upon the 'privacy-by-design' approach embodied in GDPR. There are roles for the ICO and the National Cyber Security Centre to provide advice and assurance on the systems that are effective for a given purpose.

Clouds or other data centres should enable users to have guaranteed knowledge of where the data is stored (so there could be a UK data cloud where the provider ensures that it is stored and processed in the UK). Data storage providers should ensure access to the data they hold has clear service level agreements. These might be available 100% of the time, or 99% of the time. Again, there should be certification of the quality of data providers – some form of kite marking to agreed standards around security of the holdings.

15 - Demand for external data storage and processing services is growing. In order to maintain high standards of security and resilience for the infrastructure on which data use relies, what should be the respective roles of government, data service providers, their supply chain and their clients of such services?

Government has a duty to set standards for dealing with redundancy, duplication and the robustness of data, which data service providers are required to demonstrate they can meet.

16 - What are the most important risk factors in managing the security and resilience of the infrastructure on which data relies? For example, the physical security of sites, the geographic location where data is stored, the diversity and actors in the market and supply chains, or other factors.

There are a whole range of risks, all of which need to be addressed in the round for each situation. In the Royal Academy of Engineering report, *Cyber safety and resilience: strengthening the digital systems that support the modern economy*[5], vulnerabilities identified included the following:

- Poor quality components and the way that they are integrated into communications networks compromise the cyber safety and resilience of systems. The trustworthiness of software is also a concern.

- The supply chain is now considered to be susceptible to a range of hardware-based threats. Counterfeiting and the emerging threat of hardware Trojans may introduce modifications to hardware. With the globalisation of supply chains, the design and manufacture of today's electronic devices is now distributed worldwide, through overseas foundries, third party intellectual property (IP) and third party test facilities. Many different untrusted entities may be involved in the design and assembly phases and it is becoming increasingly difficult to ensure the integrity and authenticity of devices.

- IoT is a communications infrastructure that may be a target for attack in its own right, but it also is bearer or store for data. The security of data at rest or in transit is an important consideration. Security is needed to protect its integrity and availability and to reduce the risk that it may be used for hostile purposes.

- There remains the human risk of insiders; from hacktivists and criminals as well as unhappy employees, and of course human error.

Appropriate operational frameworks, supported by technical guidance, are an important tool in risk management for data centres. Operators should integrate cybersecurity into their and the supply chain's overall risk management systems. The appropriate competencies are also needed to ensure that

operational systems are implemented effectively. Operators, vendors, designers and regulators each have their own competency set as they have different responsibilities.

Trustworthiness of companies should be built on a code of best practice similar to climate change obligations. A "comply or explain" style of enforcement could work well in this area.

17. To what extent do you agree with the following statement: The government should play a greater role in ensuring that data use does not negatively contribute to carbon usage?

Achieving net zero is a priority challenge and it is essential that data from all sectors is used actively to reach that target, and that digital technologies do not contribute disproportionately. Government has a role in requiring data about emissions from all sectors including the tech sector.

The Royal Academy of Engineering's report *Towards trusted data sharing: guidance and case studies* highlights that best engineering practice is a vital part of realising the opportunities and managing risks, with its focus on the interface between technical systems, people and organisations. Data must be assembled, structured and managed over its lifecycle so that it meets business or other requirements, for which a robust engineering approach is needed[6]. As part of good data management, it will increasingly be necessary to ensure that the carbon impact of data use is minimised, for example by ensuring that only the data that is needed is collected and processed. This becomes more important as technologies such as the Internet of Things are adopted, with the potential to collect huge volumes of data. Considerations such as whether the data is being kept, or whether it might be securely destroyed, are relevant here.

The National Data Strategy rightly points out that issues remain around a lack of transparency from providers, in particular sustainability reporting related to specific services. We welcome the fact National Data Strategy plans for government to make sustainability of data a key aspect in its procurement and supply chain. Further to this, and as argued in the Royal Society's *Digital Technology and the Planet,* Government also has a role in ensuring tech companies share publicly data about the energy consumptions of their digital systems and products, including embodied and use phase emissions, in particular from data centres.

Regulators should develop guidance about the energy proportionality of digital applications. Such guidance could set out key questions to consider when developing or deploying digital technologies. Where there are options to use less energy-intensive approaches, guidance should make this clear. For example, the Financial Conduct Authority should provide guidance on the energy intensity of blockchain-based applications used in financial systems. Green computing approaches should be part of a research and innovation mission for the digitalisation of the net zero transition

In addition, government can help to promote the use of data to reduce resource use and enable decarbonisation, and a focus on data use for net zero could be an important 'pathfinder' or 'lighthouse' project[7].

Mission five: Championing the international flow of data

19. What are your views on future UK data adequacy arrangements (e.g. which countries are priorities) and how can the UK work with stakeholders to ensure the best possible outcome for the UK?

Given our interconnected world (e.g. supply chains, COVID-19 transmission, international net-zero efforts etc.), all suggestions in the report for creating a critical data infrastructure also need to be addressed at the international level. Reducing the barriers to international data sharing and integration should be the aim, though we appreciate that there are regulatory and cultural challenges that make this more difficult than it is at the national level.

The Academies support seeking EU 'data adequacy' and breaking or drifting from this adequacy could have significant negative consequences for UK research and innovation, for the outcomes and safety of UK patients, and for the UK's contribution to global health, research and innovation efforts. It is important that the UK's own data adequacy regime does not undermine this data adequacy, which might happen if it permitted onward transfer of data about EU citizens to third countries. This could be detrimental to our use of data for research.

There is great potential value in using international data to inform UK approaches. However, significant international divergence on definitions and standards present a significant barrier to making the robust comparisons required. The ONS and ESRC should work with other national statistical agencies and research agencies to enable work that properly enables international comparisons of important social phenomena.

Sources:

Academy of Medical Sciences: *'Our data-driven future in healthcare'* (2018)

Academy of Medical Sciences and Health Data Research UK workshop *Realising patient and NHS benefits from health and care data – from policy to practice*

Academy of Medical Sciences and Royal Society: *From Bench to Bedside* (2019)

Academy of Medical Sciences, the Medical Research Council and the National Institute for Health Research: *Artificial intelligence and health* (2019)

British Academy: *The Right Skills: Celebrating Skills in the Arts, Humanities and Social Sciences* (2017)

British Academy and Royal Society: *Data Management and Use: governance for the 21st century* (2017)

British Academy, Royal Society and techUK: *Data Ownership, Rights and Controls* (2018)

DELVE (Data Evaluation and Learning for Viral Epidemics) *Data Readiness: Lessons from an Emergency* (2020)

National Engineering Policy Centre Supply chain challenges, lessons learned and opportunities (July 2020)

National Engineering Policy Centre Beyond COVID-19: laying the foundations for a net-zero recovery (November 2020)

National Engineering Policy Centre (September 2020), Infrastructure Resilience Roundtable: Ensuring resilient national infrastructure systems (to be published)

National Engineering Policy Centre *Winter is coming: risks for interdependent infrastructure* (October 2020)

Royal Academy of Engineering *Cyber safety and resilience: strengthening the digital systems that support the modern economy* (2018)

Royal Academy of Engineering (2018), Towards trusted data sharing: guidance and case studies.

Royal Society: *Progress and Research in Cybersecurity* (2016)

Royal Society: *Machine Learning: the power and promise of computers that learn by example* (2017)

Royal Society: *Dynamics of Data Science Skills* (2019)

Royal Society: *Protecting Privacy in Practice: the use and limitations of privacy enhancing technologies in data analysis* (2019)

Royal Society: *The Data Governance Landscape* (2020)

Royal Society: *Digital Technology and the Planet* (forthcoming 2020)

Submitting organisations

The Academy of Medical Sciences:

We are the independent body in the UK representing the diversity of medical science. Our mission is to advance biomedical and health research and its translation into benefits for society.

The British Academy:

The British Academy is the UK's national academy for the humanities and social sciences. We mobilise these disciplines to understand the world and shape a brighter future.

The Royal Academy of Engineering:

The Royal Academy of Engineering is a charity that harnesses the power of engineering to build a sustainable society and an inclusive economy that works for everyone.

The Royal Society:

We are the independent scientific academy of the UK, dedicated to promoting excellence in science for the benefit of humanity.