Response to

Policy paper on the National Data Strategy [1]

On behalf of the UK Computing Research Committee, UKCRC.

Prepared by: Professor Chris Johnson,
Pro Vice Chancellor, Engineering and Physical Sciences,
Queen's University Belfast.
c.w.johnson@qub.ac.uk

The UK CRC is an Expert Panel of all three UK Professional Bodies in Computing: the British Computer Society (BCS), the Institution of Engineering and Technology (IET), and the Council of Professors and Heads of Computing (CPHC). It was formed in November 2000 as a policy committee for computing research in the UK. Members of UKCRC are leading researchers who each have an established international reputation in computing. Our response thus covers UK research in computing, which is internationally strong and vigorous, and a major national asset. This response has been prepared after a widespread consultation amongst the membership of UKCRC and, as such, is an independent response on behalf of UKCRC and does not necessarily reflect the official opinion or position of the BCS or the IET.

Questions on the framing of the strategy

We want to ensure that we produce a forward-looking strategy that takes into account public opinion and delivers real change. These questions will help to inform future work that the government will take in this space. It will provide evidence for the government to target areas for intervention in future policy.

Q1. 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.

> **Somewhat agree.**
>
> **In addition to the existing pillars, we would advocate a fifth element "Trust" which is the intended effect of many aspects of "Responsibility" but which is not always achieved even when data is used in an ethical, lawful manner. Together trust and responsibility should be sufficient to ensure that the public do not turn to unreliable sources for "fake news".**

The pillars and missions, arguably place insufficient emphasis on openness as an expectation for public funded repositories. "Publicly funded data should be publicly available", says ukdataservice.ac.uk:

https://www.ukdataservice.ac.uk/get-data/other-providers/open-data

(funded by ESRC). data.gov.uk provides a search facility for open datasets

https://data.gov.uk/
https://guidance.data.gov.uk/#data-gov-uk-technical-documentation

The National Data Strategy could establish a commitment to making more such data openly available. The UK does not rank highly for openness:

https://www.oecd.org/gov/digital-government/open-government-data.htm

We are below the OECD average in these rankings. One reason for this is that we have tended to focus on a mixed set of objectives where the value of data derived by public funding means that it is not available to the public. These concerns are reinforced by Mission 1 to "unlock the value of data across the economy". Once clear reading of this statement is to sell data, not to make it maximally available and useful. The Ordnance Survey data is infamous case in point: it is publicly funded but not publicly available. Some is:

https://www.ordnancesurvey.co.uk/business-government/tools-support/open-data-support

but most is not:

https://odimpact.org/case-united-kingdoms-ordnance-survey-opendata.html

This creates significant overheads for UK researchers who are also funded by the public; making it easier to derive data rich interventions for the populations of other countries than it is for our own citizens.

NB: For question 2, we are only looking for examples outside health and social care data. Health and social care data will be covered in the upcoming Data Strategy for Health and Social Care.

Q2. 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.

Data was used in a host of ways during the coronavirus crisis to help companies respond in resilient ways to rapid changes in market conditions, for example following changes in lockdown rules and consumer behaviour.   UK Computing Research assisted in many of these developments – for example, by laying the foundations for the digital twins that helped UK energy distribution companies conduct the contingency planning to ensure the network could cope as more people spend more time at home.   Similar examples can be derived for transport, food distribution etc where computational modelling and operations research helped UK companies anticipate and respond to unprecedented challenges.

Q3. 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?

The proposals focus on data and could useful be extended to include the information derived from the processing of that data through, for instance, the analysis of statistical correlations or the putative identification of patterns within that data.   These processes often help optimise services towards central tendencies/"central values" in a distribution at the expense of outliers, which in many cases may correspond to individuals with a protected characteristic but could equally apply to the very young, the elderly etc. While the data may itself reflect the diversity of the population it is the use to which it can be put that introduces equal concerns.

There are also concerns about more systemic implications of the proposals as stated here – for instance, the Pitt review into UK flooding documents cases in which the emergency services could not reach at-risk individuals because local NHS staff felt it would be a breach of confidentiality even though there were justified concerns for the safety of those individuals from the rising flood waters.

Q4. 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.

Building on the answer to Question 3, if data derived from one region is used to identify an optimisation process that is then applied in another very different region the results are unlikely to yield the same benefits. There may also be aspects of a particular context that make other factors more important in one region than in another, leading to

correlations being ignored simply because data was dismissed as irrelevant to any initial analysis. Greater research is required to develop appropriate processes and procedures that can help both private and public sector organisations to identify and mitigate such effects in a manner that is proportionate and justified by the intended use of the data.

We are proposing the creation of a framework to identify where we can and should make data available in the wider economy. There are a number of ways the government can intervene to achieve this goal – including as a collaborator, steward, customer, provider, funder, regulator and legislator. Using Policy Lab's Style of Government Action could be helpful in thinking about the next few questions.

These questions will provide an opportunity for the government to scope out areas of focus for the data availability framework.

Data availability: For data to have the most effective impact, it needs to be appropriately collected, accessible, mobile and re-usable. That means encouraging better coordination, access to and sharing of data of appropriate quality between organisations in the public sector, private sector and third sector, and ensuring appropriate protections for the flow of data internationally.

Q5. Which sectors have the most to gain from better data availability? Please select all relevant options listed below, which are drawn from the Standardised Industry Classification (SIC) codes.

	Accommodation and Food Service Activities
	Administrative and Support Service Activities Agriculture, Forestry and Fishing
	Arts, Entertainment and Recreation
X	*Central/ Local Government inc. Defence*
X	*Charity or Non Profit*
	Construction
X	*Education*
	Electricity, Gas, Steam and Air Conditioning Supply
X	*Financial and Insurance Activities*
X	*Human Health and Social Work Activities*
X	*Information and Communication*
X	*Manufacturing*
	Mining and Quarrying
X	*Transportation and Storage*
	Water Supply; Sewerage, Waste Management and Remediation Activities
	Wholesale and Retail Trade; Repair Of Motor Vehicles and Motorcycles
X	*Professional, Scientific and Technical Activities*
	Real Estate Activities
	Other

Q6. What role do you think central government should have in enabling better availability of data across the wider economy?

The government's role should be promoting and supporting open data that can focus social impact through citizen engagement; for instance, pothole-registering services; or as a timely example, mapping the need for meals for poor children during school holidays with the restaurants etc that are volunteering provision.

Q6a. How should this role vary across sectors and applications?

Existing market mechanisms support the commercial exploitation of data sources but public bodies and third sector organisations often find this difficult – especially where an agile approach is required to respond to changes in society and the environment.

Data foundations: The true value of data can only be fully realised when it is fit for purpose, recorded in standardised formats on modern, future-proof systems and held in a condition that means it is findable, accessible, interoperable and reusable. By improving the quality of the data we are using, we can use it more effectively, and drive better insights and outcomes from its use.

Q7. 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 is a significant user and producer of data – lead departments should be actively engaged with the industry bodies/research organisaitons that help establish appropriate innovative and accepted interchange formats so that public bodies are not left behind by, or actively delay, transitions to more innovative and open approaches.

Q8. 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?

Some of government's own procurement practices often act as barriers to small and medium sized enterprises. In some cases, this stems from calls that require companies to commit to archaic data formats with short time scales – the costs of which cannot easily be assessed without significant due diligence. Similar comments can also prevent researchers from determining whether public data can be used to validate new forms of analysis/modelling that might yield significant public benefits. These barriers are exacerbated when single, large preferred suppliers sustain data monopolies.

The Smart Data Review in 2019 consulted on ways to make evolving schemes more coordinated across banking, finance, telecoms and energy. The focus of Smart Data is

customers asking their providers to share information about them with third parties who then use this data to offer innovative services to consumers and SMEs.

Q9. Beyond existing Smart Data plans, what, if any, further work do you think should be done to ensure that consumers' data is put to work for them?

Great care should be taken with the use of language in this consultation. Phrases such as "ensure that consumers' data is put to work for them" create significant concerns when consumers have little idea about the mechanisms that are used to determine which agencies are best placed to provide the greatest benefits in "putting their data to work for them".

- "The focus of Smart Data is citizens asking their providers to share information about them with third parties". Policy should empower consumers to own and take good care of information about them, not simply encouraging its sharing with third parties where the track record has been notably poor.

Q10. How can the UK's data protection framework remain fit for purpose in an increasingly digital and data driven age?

Active engagement between CSAs and UKRI can help to ensure that lead Departments are aware of opportunities to develop and apply innovative approaches to data management. Progress across government is extremely varied with some CSA and Departments exploiting a range of opportunities – not only within AI and ML but also with open formats – to maximise public benefits but also to encourage exchange between commercial organisations in developing new market opportunities. Others seem largely ignorant of the issues addressed in this consultation. We note the links between this question and our response to question 18.

In section 7.1.2 we lay out the functions of the Centre for Data Ethics and Innovation (CDEI), set up in 2018 to advise the Government on the use of data-driven technologies and AI.

Q11. 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?

      Strongly disagree
      Somewhat disagree
      Neither agree nor disagree
X   *Somewhat agree*
      Strongly agree

Q11a. How would a change to statutory status support the CDEI to deliver its remit?

> *Unless the Centre works with other regulatory organisations it will struggle to have much impact. Equally, even if its status was changed without that cooperation they may struggle to understand the significance of particular data sources within particular industries just because of the ubiquitous and embedded nature of data driven enterprises across all sectors.*

The government is going to set an ambitious package of work in this space and wants to understand where we can have the biggest impact.

Q12. We have identified five broad areas of work as part of our mission for enabling better use of data across government:

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

We want to hear your views on which of these actions will have the biggest impact for transforming government's use of data.

> *Responses to previous questions have identified the influence of CSAs in engagement with UKRI and leading researchers – others have argued that government procurement practices should avoid data monopolies that often lead to cost escalation and legacy infrastructure. Close involvement with the industry bodies that establish interchange standards is also necessary post-Brexit.*

Q13. The Data Standards Authority is working with a range of public sector and external organisations to coordinate or create data standards and standard practices.

We welcome your views on which if any should be prioritised.

> *The infrastructure on which data relies is the virtualised or physical data infrastructure, systems and services that store, process and transfer data. This includes data centres (that provide the physical space to store data), peering and transit infrastructure (that enable the exchange of data), and cloud computing that provides virtualised computing resources (for example servers, software, databases, data analytics) that are accessed remotely. We welcome standards that enable companies and private customers to determine whether or not a particular offering is value for money in terms of the non-functional attributes (reliability, security, trust etc) identified in this consultation.*

Irrespective of where the DSA focuses, there must be some means of assessing whether or not their engagement has had any net effect both on the standards and then on the UK industries/public bodies affected by them.

Q14. 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?

This was addressed in Q.13 – the European Network and Information Security Agency has published a range of guidance in this area and this could be refreshed by, for example, the NCSC working with the DSA and then promoted in a manner that can be understood by commercial and private end users.

Q14a. 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?

In many areas of the UK this is almost impossible – except where financial data is concerns or where companies have undertaken audits as part of GDPR (and more rarely NISD) compliance. There are contractual and technological barriers – as well as the costs of conducting such exercises. In some cases, lead government departments have acted through the NISD competent authorities (eg HSE with BEIS) to tackle these issues but progress has been very mixed across UK industry.

Q15. 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?

Above all, we would welcome a statutory obligation for any company holding data "owned" or "controlled" by another party to provide a clear and auditable mechanism by which that party can regain the information within a specified time limit and a specified (hopefully open) format.

The increasing costs of ransomware attacks should also be a focus for this consultation which has directly affected the expectations mentioned in the previous paragraph.

Q16. 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 many complex factors that interact in this space. In some cases, critical UK data has been put at risk by physical co-location (for example, with the Buncefield petrochemical farm).   In other cases, the lack of diversity and of patching undermined service provision (for example, the impact of Wannacry on NHS services). A key issue here is that we cannot expect every customer of data services to be equally informed about the nature and extent of these risk factors and hence the suppliers of data services should accept a duty of care, similar to that embedded within the physical provisions of the 1974 Health and Safety at Work Act.

Q17. 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?

       Strongly disagree
       Somewhat disagree
       Neither agree nor disagree
       Somewhat agree
X   *Strongly agree*

This has traditionally had little impact on companies and public bodies beyond the HVAC and processing costs associated with data centres. There is a generation that now rightly expects higher standards than we have arguably become used to. The duty of care mentioned in Question 16 could be extended to the impact of storing the data in terms of the UN SDG priorities which would align well with the evolving policies across government in the run up to COP26.

Q18. How can the UK improve on current international transfer mechanisms, while ensuring that the personal data of UK citizens is appropriately safeguarded?

We will seek EU 'data adequacy' to maintain free flow of personal data from the EEA and we will pursue UK 'data adequacy' with global partners to promote the free flow of data to and from the UK and ensure it will be properly protected.

This seems a proportionate approach, however, we have little evidence that enforcement actions have been sufficiently policed and this has undermined public confidence in the UK bodies that have an interest in maintaining the pillars of this policy.
The EU arguably has the most coherent data protection rules, supported by the decision in the Schrems/Safe harbour replacement case. There is a danger that the UK will lose some of the protections as we conduct bi-partite negotiations with countries that have a different perspective on these issues.

Q19. 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?

Data adequacy should form a core component of on-going trade negotiations as many of the countries that make greatest use of data about our citizens are also our closest trading partners – for obvious reasons. Some of these partners have not traditionally accepted UK influence on how that data may be used even though they have strong reciprocal expectations in that respect.