Response to the consultation on

The National Data Strategy

Submitted by
Prof. Derek McAuley, Prof. Lilian Edwards and Dr. Jiahong Chen of
Horizon Digital Economy Research Institute, University of Nottingham

2 December 2020

1. Horizon[1] is a Research Institute centred at The University of Nottingham and a Research Hub within the UKRI Digital Economy programme[2]. Horizon brings together researchers from a broad range of disciplines to investigate the opportunities and challenges arising from the increased use of digital technology in our everyday lives. Prof. McAuley is Director of Horizon and Principal Investigator of the EPSRC-funded DADA[3] (Defence Against Dark Artefacts) project, addressing smart home IoT network security, and its acceptability and usability issues, the ESRC-funded CaSMa[4] (Citizen-centric approaches to Social Media analysis) project to promote ways for individuals to control their data and online privacy, and the EPSRC-funded UnBias[5] (Emancipating Users Against Algorithmic Biases for a Trusted Digital Economy) project for raising user awareness and agency when using algorithmic services. Prof. Edwards is Professor of Law, Innovation & Society (Newcastle University) and a member of the Alan Turing Institute. Dr. Chen is Researcher Fellow of Horizon, currently working on the DADA project. We are happy to be contacted for further discussions and for our response to be published in full.
2. Before responding to the individual consultation questions, we would like to set out a number of general principles that should inform the ongoing planning and implementation of the government's next steps following the NDS. The relevance of these principles will be further elaborated in our responses to individual questions below. These principles include:
3. **Promote Data Quality not just Data Quantity:** Any collection, transformation or use of data should have regard to the level of accuracy, up-to-dateness, relevance and representativeness of the dataset (see response to Q13).
4. **Comply with Existing Rules:** Data-related requirements in legal frameworks currently in force, especially data protection law, should as a first priority be upheld and enforced effectively and before recourse to guidance or ethics (see response to Q10).
5. **Protect Public Data Alongside Openness:** Public sector data, especially sensitive or personal data, should only be made public where its security, provenance and accountability can be guaranteed. As best practice, wherever possible data should not be transferred in bulk but shared via principles of decentralisation and interoperability (see response to Q14).

---

[1] http://www.horizon.ac.uk
[2] https://epsrc.ukri.org/research/ourportfolio/themes/digitaleconomy/
[3] https://www.horizon.ac.uk/project/defence-against-dark-artefacts/
[4] http://casma.wp.horizon.ac.uk
[5] http://unbias.wp.horizon.ac.uk

6. **Govern Public Data Shared with the Private Sector:** Any rules restricting or guiding use of public sector data regarding access, sharing and reuse should also apply when this data is shared with or used by private sector bodies including sub-contractors as far as possible. These conditions can be applied and enforced via procurement processes (see response to Q6).

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.

7. ○ Strongly disagree
   ○ Somewhat disagree
   ◉ Neither agree nor disagree
   ○ Somewhat agree
   ○ Strongly agree

Q2. We are interested in examples of how data was or should have been used to deliver public benefits during the coronavirus (COVID-19) pandemic, 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.
For question two, 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.

8. No answer.

Q3. If applicable, please provide any comments about the potential impact of the proposals outlined in this consultation may have on individuals with a protected characteristic under the Equality Act 2010?

9. Rachel Coldicutt has coordinated a response to this question, with a list of strategic and tactical recommendations that we support: https://docs.google.com/document/d/1r51kaNcB6PAka679ooOUv5Wv3g-ixCUPP8hXMCLuwoI

10. In addition to that response, we would also like to highlight the importance for the NDS to take into account, as well as the protected characteristics operating in current English equality law, the duty on public bodies to consider socio-economic disadvantage when making decisions in Section 1 of the Equality Act 2010. This was enabled in Scottish law but not, so far, in the rest of the UK. One obvious example is the controversies around GCSE and A-Level grading earlier this year, where the algorithms were criticised for having treated students from certain demographics unfairly. Some of such categories are not necessarily protected under the Equality Act 2010 and yet were of serious public interest. Other contexts may include uses of data for employment, child welfare, benefits, fraud detection or criminal justice purposes, in all of which cases trained algorithms may unfairly and disparately target or disadvantage certain socio-economic groups. Socio-economic discrimination is not always co-existent with racial or other protected forms of discrimination and it would set an example if the use of public sector data, especially when shared with private sector bodies, to be held to a higher standard which considered these potential detriments.

11. We therefore call for stronger safeguards and monitoring measures to be put in place to avoid both short- and long-term detrimental effects to certain social segments with protected characteristics,

and those at socio-economic disadvantage alike. We suggest the appointment of an independent body (possibly similar to the National Data Guardian in the health and care sector) to oversee the arrangements on data uses in public projects and procurement processes, as well as to evaluate the net effects of the NDS, whether positive or negative, on different social groups. To enable such a function, there should be heightened transparency requirements for uses of data in public and private sectors, including, for example, record-keeping duties to document data provenance, algorithmic models and impact assessment.

Q4. We welcome any comments about the potential impact of the proposals outlined in this consultation on the UK across all areas, and any steps the government should take to ensure that they take account of regional inequalities and support the whole of the UK?

12. No answer.

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.

13. No answer.

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

14. While we support the principle of access to public sector data and its reuse, to prevent abuses of public sector generated data, which has been created at state expense and which may impinge greatly on the welfare of citizens and consumers, we propose the **Govern Public Data Shared with the Private Sector** principle noted above. Specifically, there should be a general legal requirement that rules governing data access, sharing and reuse which would have pertained to public sector data (e.g. rules derived from data protection, Freedom of Information and Public Sector Data Reuse Directive) must be included as terms in the procurement process where such data is outsourced by agreement with private sector parties including sub-contractors along with an impact assessment which shows how these rules are to be met post-transfer. In addition, the government should also promote appropriate data good practices in the private sector. To ensure accountability, we in particular recommend transparency duties in cases where individuals or the society may be adversely impacted by reuses of public sector data in the private sector, or where substantial profit has accrued to private sector contractors as a result of such transfer and reuses. Such duties have been proposed by a number of bodies in respect of use of data (not just personal data) to train machine learning algorithms (e.g. those proposed by AlgorithmWatch, Ada Lovelace Institute: https://algorithmwatch.org/en/governing-platforms-final-recommendations/ and https://www.adalovelaceinstitute.org/wp-content/uploads/2020/10/Transparency-mechanisms-explainer-1.pdf).

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

15. No answer.

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.

16. ○ Strongly disagree
    ○ Somewhat disagree
    ◉ Neither agree nor disagree
    ○ Somewhat agree
    ○ Strongly agree

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?

17. No answer.

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?

18. We suggest that the government should look beyond traditional centralised data accumulation and data mining, towards alternative business and technical models which may achieve more innovative results with less risk of data breach and more social benefits. Horizon has been a pioneer for over a decade in developing decentralised "edge" computing models for deriving results from data without exposing that data to potentially risks of insecurity, privacy invasion or siphoning off into the private sector (see e.g. the Databox project). For a private-sector example, comparison and recommendation services do not always need to gain direct access to user data; instead, the analysis can be performed entirely on the user's terminal device, with user data transferred to and stored on the device only. Approaches like this can promote competition without the risk of oversharing personal data. This approach should be invaluable to the use and reuse of public sector data to create value for citizens without alienating sensitive data to the private sector. The government should support demonstrative projects to show that such innovations are both technically feasible and financially sustainable.

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

19. Post-Brexit, the UK's data protection legal framework should continue to focus on maintaining a world-leading level of protection for personal data, and indeed should ascribe resources to more effectively enforcing existing data protection laws, something which has come under questioning scrutiny especially in the disrupted COVID-19 period. The principle **Comply with Existing Rules** should form a key part of the NDS, not just because it is in line with the NDS's explicit goal of "creating a fairer society for all", but also because it is crucial for maintaining the UK's international trade gold-standard brand in relation to transfers of personal data. Contrary to some advocacy materials, the principal Western global markets are now for trustworthy data products rather than deregulated unsafe products (see e.g. the European approaches against centralised, data-invasive contract tracing apps and preference for decentralised ones); and multi-national corporates selling services into the UK, the EU *and* elsewhere will wish to meet EU data protection standards in a harmonised way, rather than creating different versions with different support, customer service and regulatory issues, for different segmented markets. In short, any attempt to water down GDPR protection after Brexit should be repelled not simply out of respect for the UKs history of human

rights and the rule of law but because it will be a losing strategy in the battle for markets in a connected globalised data driven world.

20. Robust mechanisms must be established to prevent and address the negative effects of data uses, notably the requirements of DPIA and data protection by design and by default. These exist in current law but are often unenforced or regarded as impractical. In particular, DPIAs must be published by default, publication must be mandatory for the public sector and consultation with affected data subjects must become both routine and public. There is also a need for data protection regulatory oversight to be better integrated with other types of review, notably equality and labour standards oversight. Although steps have been taken along this line (e.g. cross-regulator meetings), there is a need for more formalised frameworks, joint enforcement actions and sharing of expertise and resources.

21. Maintaining a high level of data protection, supported with effective enforcement, is crucial for business. Strong data protection forms the foundation of consumer trust and ethical innovation. Recent developments in data protection regimes in other jurisdictions have shown that the global data protection standard is moving up closer to – not away from – the regulatory model represented by the GDPR, which proves the internationally-recognised social and economic values of a strong data protection law. Crucially, free flow of personal data across border, which is essential for the digital economy, may be impeded by a lowered level of data protection in the UK leading to a non-adequacy finding (see response to Q18).

Q11. To what extent do you agree with the functions set out for the Centre for Data Ethics and Innovation (CDEI) - AI monitoring, partnership working and piloting and testing potential interventions in the tech landscape? Please explain your answer.

22. ○ Strongly disagree
    ○ Somewhat disagree
    ◉ Neither agree nor disagree
    ○ Somewhat agree
    ○ Strongly agree

23. We are not convinced these functions are either clearly defined, adequate or that overall, the CDEI is necessary rather than desirable. The definition of "AI" is contested (indeed CDEI itself speaks of data-driven technologies which is far wider) and there needs to be a debate about whether this body is intended to be limited to machine learning and near-future innovation and social impact, or has a role in respect of more philosophical concepts and if so, what impact these interventions are having on industry. A number of bodies in the UK landscape already overlap with the apparent functions of CDEI both in the data/innovation, and "AI" world and elsewhere (the Digital Catapults, Alan Turing Institute, Office for AI, Leverhulme Centre for the Future of Intelligence, the Strategic Artificial Intelligence Research Centre, the Oxford Internet Institute, etc). A review needs, we suggest, to question if the CDEI adds value for money to an existing plethora of quango-like and (not always) public money supported bodies, especially given our response below to Q11a.

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

24. We suggest that (see our response above to Q11) before considering a statutory remit for CDEI, first evidence is needed to show that resources have been put into good use. It needs to be

demonstrated that its role is both clearly defined and needed, that it has effectively performed its role in the past, and that it has a clear function for the future. Without such evidence we cannot say if it will deliver its new functions set out in the NDS in the future whether in a statute-backed form or not. In particular we are in doubt whether the CDEI has notably influenced commercial practices.

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 any actions you think will have the biggest impact for transforming government's use of data.

25. No answer.

Q13. The Data Standards Authority is working with a range of public sector and external organisations to create a pipeline of data standards and standard practices that should be adopted. We welcome your views on standards that should be prioritised, building on the standards which have already been recommended.

26. The principle of **Promote Data Quality Not Just Data Quantity** is of paramount importance for ensuring data is used truly for the benefits of the society. Further standards should be developed to support data collectors, data brokers and data users to evaluate and improve data quality, and to decide for what purpose a dataset is fit. Certain metrics are especially important as there have been examples of how data can be misused/abused when it is inaccurate, out of date, mistakenly labelled, repurposed for inappropriate uses, or simply non-representative of what it claims to represent. Depending on the sector, the development of the standards should focus on these aspects. Importantly, since the uses of data cannot be determined before having a good understanding of the content of the data, the DSA should first conduct a survey of what data is available (and of what quality) in a sector prior to commencing work of standardisation for that sector.

Q14. What responsibilities and requirements should be placed on virtual or physical data infrastructure service providers to provide data security, continuity and resilience of service supply?

27. The principle of **Protect Public Data Alongside Openness** should be mandated for all these categories of actors. One promising way to implement this principle is to require data to be made available through APIs, rather than releasing the whole dataset. This can improve security, accountability, decentralisation and interoperability of data sharing, not to mention for large datasets, it is already prohibitively expensive to transfer and keep a copy – for example the Met Office data.

28. One benefit of using APIs is that it allows better control over who has access to what data, especially when such data can be potentially sensitive for safety or security reasons. Using such control afforded by APIs also means that record-keeping can be conducted in a more sophisticated, automated manner, which can not only reduce the risk of data abuses, but also enables better

transparency and auditability of data uses. The API model also allows data to be stored in a decentralised way, reducing the risks of large-scale data breaches or data loss, and of power concentration resulting from data centralisation. Lastly, data sharing via APIs also enhances interoperability across systems. In fact, the API approach has been chosen in other jurisdictions as the secure way to make data open, such as in the EU's European Strategy for Data, and Singapore's Smart Nation initiative.

29. We are however also aware that the API model can be potentially exploited to restrict access to data, which has been seen in the area of handling FOI requests. We therefore urge the government to lay down clear rules around what is permissible when it comes to accessing public sector data via APIs. Also, the design of the API protocols should involve the industry, experts, regulators, civil society groups and other stakeholders.

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?

30. No answer.

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?

31. No answer.

Q16. What are the most important risk factors in managing the security and resilience of the infrastructure on which data use 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.

32. A core design principle of the Internet was federated (independently operated) interoperable systems for the purposes of resilience against attack; a functioning competitive market of interoperable services offering heterogeneity and fault isolation would significantly limit the damage caused by a service failure or cyberattack.

Q17. Do you agree that the government should play a greater role in ensuring that data does not negatively contribute to carbon usage? Please explain your answer. If applicable, please indicate how the government can effectively ensure that data does not negatively contribute to carbon usage.

33. ○ Strongly disagree
    ○ Somewhat disagree
    ◉ Neither agree nor disagree
    ○ Somewhat agree
    ○ Strongly agree

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

34. As highlighted by industry as well as academic experts, ensuring the free flow of personal data between the UK and the EU will be crucial for all sectors across the economy. The Court of Justice of the EU's rulings in key decisions, such as *Schrems* and *Schrems II*, have made it clear that data protection is a matter of fundamental rights, and any international agreements concerning data transfers to outside the EU/EEA can be challenged before the Court. This means, the validity of the EU's recognition of the UK as a safe third-country for data transfers, whether in the form of an "adequacy decision" or as part of a future trade agreement, will depend on the UK's fulfilment of its ongoing obligations to maintain a high standard of data protection both internally and regarding arrangements for onward transfers to other countries (such as trade deals with other economies).

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?

35. We do not feel this is the right question. The UK both as a trading partner and as a world leading centre of industry and innovation must decide the level of human rights protection and trustworthiness it wishes to pitch its brand. The UK must also have regard to its own internal debate on fundamental liberties and its own common law requirements. A decision cannot simply be made to accord with the EU at the top end or the US, China or APEC at the lower end. Fundamental legal protections cannot be traded like commodities or disposed like assets and liabilities; they need to be reviewed as part of social and constitutional as well as economic fabric of the country.