Written evidence submitted by the Ada Lovelace Institute (DDA0038)

About the Ada Lovelace Institute

The Ada Lovelace Institute (Ada) was established by the Nuffield Foundation in early 2018, in collaboration with the Alan Turing Institute, the Royal Society, the British Academy, the Royal Statistical Society, the Wellcome Trust, Luminate, techUK and the Nuffield Council on Bioethics. Ada is an independent research institute with a mission to ensure that data and AI work for people and society.

We are working to create a shared vision of a world where data and AI are mobilised for good, to ensure that technology improves people's lives. We take a sociotechnical, evidence-based approach and use deliberative methods to convene and centre diverse voices. We do this to identify the ways that data and AI reorder power in society, and to highlight tensions between emerging technologies and societal benefit.

Introduction

The Ada Lovelace Institute welcomes the opportunity to contribute to the Science and Technology Committee's inquiry into the right to privacy: digital data. Our submission is grounded in the expertise we have built in data protection, use, governance and management. We have cited our research throughout the response and included links to specific relevant research in an annex.

Our response focuses on the first two prompts of the inquiry, covering the benefits and barriers to data sharing, and the extent to which data issues are appropriately addressed in recent government strategies and consultations, where we highlight some existing barriers that are not fully addressed by Government proposals. Specific points regarding health data sharing and the role of the CDEI have been incorporated into these points.

Summary

1. Benefits and barriers

Effective use of data can make significant direct contributions to society.[1] Better use of data can bring benefits including to:

- enable scientific breakthroughs and rapid responses to emerging challenges
- create more accurate modelling and forecasting to support evidence-based policy
- create real-time feedback on services
- personalised triage
- make public services more efficient and effective, both in terms of delivery and user experience.

Indirectly, the use of data can bring social benefit as resulting growth and innovation support a healthy economy. Data sharing can enable social connections and research collaborations across fragmented, globalised populations.

---

[1] Margetts, N. and Dorobantu, C (2019) 'Rethink government with AI' *Nature* Available at Rethink government with AI (nature.com)

In health specifically, the research platform OpenSAFELY enabled the access and analysis of over 50 million NHS patient records. This analysis of COVID-19 risk factors by health status and social category enabled targeted public health measures, both in the UK and internationally.[2]

At present however there are a number of barriers to effective data sharing in the public sector. Joint research with Ada and the Royal Society examining data-sharing initiatives during the pandemic identified the following issues to effective public-sector data sharing:[3]

a. Missing and poor-quality data.
b. Limitations to data access, data-sharing agreements and 'data readiness'.
c. Uncertainty about legal compliance and perceptions of legal barriers.
d. Cross-disciplinary cultural differences creating friction or barriers.

2. The extent to which data issues are appropriately addressed in recent Government strategies

The Government has demonstrated a welcome focus on setting the direction for innovation and data sharing as a core pillar of a post-Brexit economy. While we agree with their diagnosis of some of the issues, we disagree with some of their solutions, and see some gaps in achieving an effective, well-governed data-sharing ecosystem.

The Government approach will need to strike the right balance, to draw together data where there are clear public benefits while limiting extractive or harmful practices.[4] Decisions about data sharing currently take place against a backdrop of a power imbalance, where a small number of market-dominant organisations hold an immense about of information and power in shaping societies.

This can pose a particular challenge as the ambition of Government is to share data across the public sector, and between public and private-sector entities: Governments and the public sector across the world have historically struggled to engage with and govern private-sector use of data, for example the data breach between Royal Free and Google's DeepMind.[5] Examples of harmful practices, particularly from platforms, have demonstrated that innovation and growth are not always aligned with social value and triggered public anxiety about how data may be repurposed.

To achieve a balance, we make five recommendations to strengthen the approach outlined in the proposals referenced in this inquiry, to build the right ecosystem to ensure the Government is successful in their ambition to increase data sharing for public good.

1. Improving public trust

Having public trust is an essential pre-requisite to increasing data sharing. However, this trust in public-sector data is 'tenuous' and there is evidence of a trust deficit. considering data sharing across sectors.[8] The Government must undertake in-depth,

---

[2] Williamson, E.J., Walker, A.J., Bhaskaran, K. et al. (2020) OpenSAFELY: factors associated with COVID-19 death in 17 million patients. Nature. OpenSAFELY: Factors associated with Covid-19 death

[3] Ada Lovelace Institute (2020) Learning data lessons: data access and sharing during COVID-19 | Ada Lovelace Institute

[4] For example, the super-complaint brought against data sharing between the police and the Home Office; and concerns raised about public health data from test and trace being shared with police

[5] For example, the data breach between Royal Free and Google's DeepMind from the ICO

nationwide public dialogue and deliberation to ensure data -sharing agreements are informed by understanding of public attitudes.

2. **Clarifying GDPR requirements to protect against 'over-compliance'**
   The interpretation of the law has been identified as a barrier to data sharing, as SMEs and public institutions adopt unnecessary risk aversion. Regulators and Government need to offer clarity, guidance and standardised approaches to data sharing to support using data with confidence.

3. **Strengthening independent regulation**
   The public see strong regulation as a pre-requisite to protecting against harm, however, existing regulation has not been able to do this consistently. The Government should strengthen regulatory powers and coordination to support public trust and balance ambition of greater data sharing with trusted, independent oversight and enforcement.

4. **Introducing Biometrics legislation**
   The Government should legislate for the particular risks posed by biometrics and create a comprehensive legal framework for their governance.

5. **Further research and development**
   The Government should invest in further research and development to take a leading role globally in its approach to data governance on three aspects:

   a. Defining innovation to enable greater competition and value-creation.
   b. Developing structures for data stewardship that include public participation in their governance.
   c. Piloting and standardising stronger mechanisms for anticipating social impact, for example through AIAs, building on algorithmic transparency registers.

Detailed Evidence

1. Public trust

The results of decades of exploitative practices in the private sector has led to an erosion in public trust in the use of data.[6] High-profile examples of data use that the public did not see as legitimate have led to increasing concerns about public use of data – a recent report from the Centre for Data Ethics and Innovation described trust in public sector data sharing as 'tenuous'.[7]

In the health sector, as more data has become 'health' data, there has been an increase in private actors without a concomitant increase in adequate transparency and public confidence in social value creation. [8] Public concerns increase if data is being perceived to be sold, or individuals lack clarity about how their data is used.[9]

---

[6] Half of Britons surveyed in Doteveryone's *People, Power and Technology* survey felt they had little agency when it came to the use of their data online, and felt pessimistic about the impact of technology on their lives and on society in the future. Doteveryone (2020) Introducing People, Power and Technology: The 2020 Edition – doteveryone

[7] CDEI (2020) Addressing trust in the public sector data use.

The recent example of the GP Data for Planning and Research (GPDPR) initiative (which resulted in over three million people opting out of data sharing) warns that public anxiety about weak data governance can lead to withdrawal from data sharing even in sectors that are highly trusted, minimising opportunities for socially beneficial innovation. This erosion of trust undermines the potential collective benefit that could be extracted through responsible, trustworthy data stewardship in the public interest: we should not be complacent that people will continue to accept their data being as available as it has been over the last two decades, as the public become more engage and concerned about data use.

There must be effective mechanisms in place to understand not only average public opinions, but the views of those usually disadvantaged or least trusting of the use of their data, or the systems that use those data. Many of those who benefit least from health data uses,[10] are most at risk of oversurveillance,[11] and are also disadvantaged by poor data collection and infrastructural practices.[12]

The Government must go beyond communicating the data benefits, and undertake in-depth, nationwide public dialogue and deliberation, to ensure data-sharing agreements are informed by nuanced and dynamic understanding public attitudes, and data agreements are viewed as legitimate in context.

2. GDPR interpretation and 'over-compliance'

Ada research with the Royal Society on data access and sharing during the pandemic found organisations, particularly SMEs and public-sector institutions, can perceive the GDPR framework as difficult to apply or to understand how best to work within, often leading to unnecessary risk aversion.

The simplicity of GDPR as a risk-based framework is viewed as a benefit from the perspective of organisations that are well equipped with the legal and operational capacity to document and validate compliance. But for smaller organisations and individuals without access to legal support, the simplicity translates into uncertainty, which in turn can promote overcompliance or risk-averse approaches to using data.

As the data landscape becomes more complex, it brings together data from the public and private sectors, bringing in different players and data-sharing agreements. This leads to a lack of clarity about organisational roles – particularly when using non-health data to make health inferences. [13] Even across the public sector, there are cultural barriers and practical challenges to setting up data-sharing agreements.[14]

---

[8] Powles, J. and Hodson, H. (2017) Google DeepMind and healthcare in an age of algorithms. *Health Technol.* **7,** 351–367 (2017). https://doi.org/10.1007/s12553-017-0179-1

[9] See for example lawyers acting on behalf of openDemocracy demanding transparency on Palantir database https://www.opendemocracy.net/en/ournhs/uk-government-could-face-legal-action-over-huge-secretive-health-database/

[10] UK Statistic Authority (2021) *Leaving no-one behind: how can we be more inclusive in our data*. Available at: https://uksa.statisticsauthority.gov.uk/publication/leaving-no-one-behind-how-can-we-be-more-inclusive-in-our-data-executive-summary/pages/1/

[11] Institute of Race Relations (2021) *A threat to public safety: policing, racism and the Covid-19.*available at A threat to public safety: policing, racism and the Covid-19 pandemic - Institute of Race Relations (irr.org.uk)

[12] See the Advisory Committee on the Framework Convention for the Protection of National Minorities National minorities and COVID-19: inequality deepened, vulnerability exacerbated - Newsroom (coe.int)

There appears to be an assumption within Government that the current legal data protection regime stifles growth and innovation, but we have not found credible evidence to support that. Instead, there is evidence that it is the interpretation of the law that is a barrier to COVID-19-related research or innovation, rather than the law itself. This is in additional to a number of other practical barriers including data quality, culture and incentives as findings of the Alan Turing Institute's major review drawing on over 100 experts using data science during the pandemic demonstrate.[15]

We therefore recommend the UK's approach could provide greater institutional confidence in data sharing through greater clarity and guidance on existing regulation; standardised approaches to data sharing to give confidence to public and smaller private organisations about data sharing.

3. Strengthening independent regulation

Across extensive public research, we have found consistently that the public want more, not less, regulation to trust the use of their data.[16] There is no blanket social license for the use of data in the public interest, even during emergencies.[17] The public want clear information up front on data practices, and clarity on the boundaries of data use, rights and responsibilities, even during times of crises.

The current regulatory system is not strong enough to address existing challenges or reassure the public of their privacy, so any ambition to increase data sharing should include strengthening independent regulatory capacity. Because the primary goal of the GDPR is not to restrict data use, but to allow purposeful data processing, it has not yet been effective in curtailing the potentially harmful aspects of the data economy and its power imbalance, or in prompting a move away from the commodification of human behaviour and activity by large corporates.

CDEI plays a valuable role in the data governance ecosystem but does not mitigate the need for independent regulation.

We suggest the UK's approach could provide greater public confidence in data sharing through strengthening regulatory powers, capacity and coordination, and stronger enforcement mechanisms to ensure individual and group privacy is balanced with economic and societal benefits, and greater data sharing is matched with trusted, independent oversight and enforcement.

---

[13] Ada Lovelace Institute (2020) *The data will see you now* https://www.adalovelaceinstitute.org/report/the-data-will-see-you-now/

[14] Ada Lovelace Institute and the Royal Society (2020) *Learning data lessons: data access and sharing during COVID-19* Learning data lessons: data access and sharing during COVID-19 | Ada Lovelace Institute.

[15] The Alan Turing Institute (2021) *Data science and AI in the age of COVID-19* Available at: https://www.turing.ac.uk/research/publications/data-science-and-ai-age-covid-19-report

[16] See forthcoming Ada Lovelace Institute policy briefing on public attitudes to data governance

[17] Ada Lovelace Institute (2020) No green lights, no red lines (public perspectives on COVID-19 technology) and Confidence in a crisis? (public engagement on COVID-19 technologies)

4. Biometrics legislation

Biometric technologies pose particular risks and harms to privacy and free expression, and may lead to discrimination (both through their differential accuracy for different demographic groups, and through their uses).

The Science and Technology Committee has previously identified the need for a deeper understanding of public attitudes and clarity of legal governance of biometrics. In response, Ada conducted a nationally representative survey on UK public attitudes towards facial recognition technologies (Beyond Face Value) and convened the Citizens' Biometrics Council, a council of 50 UK adults to learn and then deliberate on biometrics in greater depth. Both the survey and the citizens' council highlighted public support for stronger safeguards on biometrics technology.

To assess the efficacy of existing safeguards, we commissioned an independent legal review by Matthew Ryder QC ('the Ryder Review'). The forthcoming review finds that the current legal framework for governing biometrics is not fit for purpose, and that the accountability mechanisms in place are fragmented and ineffective. The Review identifies the need for a new, technologically neutral, statutory framework to govern biometrics. To ensure that legislation is enforced, the Review suggests the establishment of a national Biometrics Ethics Board.

Based on our findings of public support for stronger safeguards, and the legal review findings that current safeguards are not fit for purpose, we recommend that Government passes new legislation to govern the use of biometrics. This primary legislation should account for the use of this technology for both identification and categorisation and should apply to uses by the public and private sectors.

5. Further research and development

There are opportunities where the Government could offer global leadership to define a post-Brexit approach to data governance.

***We recommend further Government research and development on three aspects*:**

I. **Defining 'value-creating' innovation which enables greater competition**
Innovation may be *value creating* (in opportunities, reduced costs, improved efficiency and effectiveness for example), but it may also be extractive or destructive. The acquisition of Fitbit by Google is an example of extractive innovation in that it increases profits with no expectation that consumers or citizens would share in the benefits. Recent revelations about some of Meta's business practices might be critiqued as 'destructive' innovation.[18]

Government should develop their definition of what UK innovation means in a post-Brexit era. Promoting genuine innovation and competition will require curtailing extractive practices in the data economy. There is an opportunity and need to undertake data reform that enables innovation amongst SMEs and increases competitiveness. This goes beyond existing Government proposals, which we predict will entrench existing ad-based business models used by big

[18] See fuller comments by Professor Tommaso Valetti on responsible innovation available at https://www.adalovelaceinstitute.org/event/responsible-innovation/

tech companies.

II. **Developing structures for data stewardship which include public participation in their governance**

Our long-standing call is therefore for ways to manage research and innovation in ways that redistribute power back to the public.[19] Examples include the developmental work on data trusts by the ODI, as well as Ada's contribution on legal mechanisms for data stewardship (with the AI Council),[20] and participatory data stewardship.[21] The Government should undertake to support further work to build and pilot models these models and ensure social value is generated from data uses.

We have outlined a framework for how participation might work in our report, Participatory Data Stewardship.[22] In particular, the Government should consider public deliberation and dialogue initiatives to inform the development of data policy and guidance, such as the public dialogue on the ethics of location data commissioned by the UK Geospatial Commission.[23] As well as one-off deliberative exercises, standing structures for public involvement should also be supported, for example through public involvement in ethics approval panels for data use, similar to Genomics England's participant panel.[24]

Controls for what data can be shared and when may be the bedrock of good data governance, but regulation must be clear on how any data sharing rules are applied. There may be instances when privacy-preserving methods (such as the use of trusted research environments and continuing use of data access committees such as iGARD[25]) would help allay fears of unregulated data access and sharing. These need to be balanced against other situations where more open data sharing might be more beneficial to patients, for example in the case of rare diseases. It is also important to remember that mandating mechanisms such as anonymity of data, are not necessarily seen as an adequate safeguard to reidentification to secure societal benefit.[26]

Looking forward to newer technologies – digital twinning and synthetic data – active personal data collection may shift. New AI techniques that generate artificial data with similar statistical properties to collected data could still have the ability to derive health insights and make predictions but would not be governed as personal data. These will no doubt complicate the debate about privacy in

---

[19] Hall and Pesenti (2016) Growing the Artificial Intelligence Industry in the UK available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf
[20] Ada Lovelace Institute (2021) Exploring legal mechanisms for data stewardship
[21] Ada Lovelace Institute (2021) Participatory data stewardship,
[22] Ibid.
[23] Geospatial Commission (2021) Public Dialogue on location data ethics
[24] The Participant Panel | Genomics England
[25] Independent Group Advising on the Release of Data (IGARD) - NHS Digital
[26] Current data sharing and collection within public bodies is governed by various frameworks including Information Governance and the Caldicott Principles our most recent work on the ongoing need for regulatory approaches for innovative practice, together with separate yet ongoing work with the NHSX's National Chest Imaging data-sets

data and bring forward a need to explore these issues early on – in the design phases, in deployment of technologies and in scaling up from one health domain to another, and crucially, even when open access data is used. The sheer amount of data often needed in machine learning, and the complexity of such models and systems, can make it difficult to define clearly where privacy-preserving methods must be applied.

III. **Piloting and standardising stronger mechanisms for anticipating social impact, for example through AIAs, building on algorithmic transparency registers**

Our upcoming research with the NHS AI Lab looks at the potential for algorithmic impact assessments (AIAs) as an approach to ensure collective, societal, economic and environmental impacts are considered at an early stage in the development of data-driven technologies.[27] It is the first such research to explore AIAs in a data-access process, showing their potential as part of a data governance and algorithmic accountability toolkit. To further evidence these approaches, and understand how they work best, more pilots will be needed, with transparent publication of their results.

While AIAs may anticipate social impacts, correcting power and information asymmetries will require a comprehensive approach and a toolbox of interventions combining legal, technical, market and economic measures.[28]

January 2022

[27] For more on our forthcoming work, and to access it when published in early February 2022, see: https://www.adalovelaceinstitute.org/project/algorithmic-impact-assessment-healthcare/

[28] The Ada Lovelace Institute is exploring through its Rethinking Data work what types of interventions would enable positive transformations for data use and regulation. Publication of the final output from this work is expected in 2022.

Annex

On data stewardship: responsible and trustworthy data use, governance and management:

- Exploring principles for data stewardship, September 2020
- Exploring legal mechanisms for data stewardship, March 2021
- Participatory data stewardship, September 2021
- Independent review of the governance of biometrics data, forthcoming.

On public attitudes and experiences of data-driven technologies:

- Beyond face value (survey of public attitudes to facial recognition), September 2019
- No green lights, no red lines (public perspectives on COVID-19 technologies), July 2020
- Confidence in a crisis? (public engagement on COVID-19 technologies), August 2020
- Citizens' Biometrics Council March 2020 (public deliberation on biometrics)
- The data divide (survey of public attitudes towards and experiences of COVID-19 technologies), March 2021
- Public dialogue on the ethics of location data (public dialogue, in partnership with the Geospatial Commission)
- The role of good governance in building public trust in data-driven responses to public health emergencies, forthcoming

On transparency and accountability for algorithms and AI:

- Transparency mechanisms for UK public-sector algorithmic decision-making systems, October 2020
- Algorithmic accountability for the public sector, August 2021
- Supporting the development of transparency mechanisms, ongoing
- Supporting AI research ethics committees, forthcoming
- Algorithmic impact assessment for healthcare, forthcoming.

On the use of data and data-driven technologies as part of the COVID-19 pandemic response:

- Exit through the App Store? (rapid evidence review of COVID-19 technologies), April 2020
- Learning data lessons: data access and sharing during the COVID-19 pandemic, January 2021. With Royal Society
- What place should COVID-19 vaccine passports have in society? February 2021
- Checkpoints for vaccine passports, May 2021.