Follow up Written Evidence Submitted by
Professor Sir Ian Diamond, National Statistician
(DDA0071)

Thank you for inviting me to give evidence to the Science and Technology Committee on 8 June for your inquiry *the right to privacy: digital data*. I hope you and the Committee accept my sincere apologies for the technical difficulties on the day that meant I was unable to contribute to the session as much as I would have liked. Please see enclosed our answers to the outstanding questions which I hope will assist the Committee with its report.

What progress is being made in addressing weaknesses across the UK's data ecosystem? Which areas and actions would you prioritise to accelerate further progress?

As we discussed in our written evidence for this inquiry[1], the Office for National Statistics (ONS) is leading in the delivery of an Integrated Data Service (IDS) in collaboration with partners across government as part of the National Data Strategy (4.2.1.). The IDS will enable faster and lower cost analysis for government through streamlined data access processes and departmental collaboration which will transform ways of working. It releases the power of data to enable government reform based on improved policy decisions. Under robust security and ethical protocols, that again we elaborated on in our written evidence, and through a Trusted Research Environment, the IDS will enable analysts to access and analyse linked data and give them the ability to disseminate a range of data.

On areas to prioritise to accelerate further progress, for the ONS to build a service that is fit for purpose, we need partners across government to feed in their requirements on data, as well as their technical and analytical needs.

How deep seated are risk-averse approaches to data sharing? What is driving this and what role can you play in addressing it?

The Committee is right to consider this issue; as we noted in our written evidence, when it comes to data sharing "*we could do more still if various barriers to data sharing were eased, particularly a mindset shift from internal department boundaries and towards risk management at government level, including recognising the risk of valuable data not being shared.*" We always balance the potential of data sharing for improving policy and outcomes with stringent ethical and legal safeguards to keep the public's data safe.

There are multiple drivers to a general risk-averse mindset within Government: data governance, data quality, and a focus on risk over reward.

On data governance, while initiatives like the Central Digital and Data Office's Playbook team have helped resolve differences in opinions on some key data shares and the IDS will bring a more comprehensive approach to cross Government data

---

[1] https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2022/03/Alison-Pritchard-ONS-to-Greg-Clark-Science-and-Technology-Cttee-privacy-and-digital-data-Jan-2022.pdf

sharing, friction still remains. Ownership of data is dispersed and fragmented, with different stewardship frameworks adopted across departments.

Secondly, rapid growth of cross-government data sharing has resulted in increase in demand at a time where the administrative systems that generate data are undergoing profound transformation to improve automation and consistency of data. This, coupled with challenges in recruitment of data professionals, can create a gap between the demand for data to be shared and bottlenecks in the preparation of the data.

Lastly, it is of vital importance to ensure public trust when we are sharing data, but Government should explicitly encourage data sharing and champion its benefits to ensure risk does not outweigh the potential rewards.

More broadly, data sharing between the public and private sector is hindered by an unequal risk profile: the organisation that shares the data assumes (or perceives itself to assume) legal risk, while the recipient reaps the benefit. Work is ongoing to understand how legal frameworks can be adapted to mitigate this risk.

The ONS is leading data sharing efforts across government and has over 400 agreements in place to share data for statistical and research purposes. We are the data hub for government and are working to streamline the existing landscape and build the IDS, to provide a mechanism through which to share data, while proactively engaging with privacy groups and the public about the development, to ensure the new service is supported.

What further steps, if any, are needed to ensure public trust on the sharing of their data, especially with private companies? What are the best mechanisms for engaging with the public on data sharing and how can their views be taken into account?

Secure access to data for the private sector, through the Secure Research Service, was introduced by the ONS in 2015 following a full public consultation, and extensive engagement with privacy groups, to ensure that such use was acceptable and demonstrably in the public interest. They must demonstrate the public good of their research, ethical considerations, and adhere to the accreditation conditions of any other project (including transparency).

The importance of public engagement cannot be understated, and maintaining public trust in how the ONS, Government and Accredited Researchers use data is critical to the continued use of these data for analysis. We do this in a number of ways, including:
- Being transparent around data usage.
- Proactively seeking public and privacy group engagement on our use of data, and provision of secure access to others through the Integrated Data Service.
- Building awareness of the ONS with the public.
- Working closely with other government and research organisations to ensure alignment of public-facing data messaging.
- Routinely developing and publishing case-studies, to show the positive outcomes and impact of data sharing and use.

- Being transparent about the safeguards we have in place to ensure we use data securely and ethically.

Key to this is involving the public in decision-making. For example, through my Data Ethics Advisory Committee[2], which provides independent ethics advice to the ONS, across government and beyond, and is made up of data ethics experts and lay members who provide an independent public voice to the Committee. The work of the Data Ethics Committee, and various partner committees such as the Research Accreditation Panel[3], is transparently available to the public through published minutes, blogs from members and registers of activity.

The outreach campaign involved in the 2021 Census was an example of engaging with a wide range of community and representative groups, as well as citizens directly, to build understanding of why the numbers matter, how we keep personal information safe, and build confidence in supplying their data. We are now taking this successful approach and applying it to all of our public engagement work.

We take an insight-led approach and are confident in our understanding of the public's attitudes, barriers and motivations. We will continue to regularly engage with audiences, through quarterly surveys and 6-monthly focus groups, to ensure we keep up to date on any shifts.

This understanding, for example, includes that the public have higher levels of acceptance for the use of data for the 'public good' than for commercial gain, and that both COVID-19 and GDPR seem to have had a positive impact on attitudes and behaviours in relation to data sharing. Sharing personal information is thought to be an inevitable part of modern life, with perceptions that people have little choice in sharing their data, but there is a growing need for reassurance that (linked) admin data will be anonymised. There is a very low awareness about both admin data collection, linkage and use, but this leads to broad acceptance once the concept is introduced depending on why and how it is being used. Finally, the high level of awareness and trust in the ONS, as evidenced in the most recent Public Confidence in Official Statistics report[4] where 89% of respondents able to express a view trusted the ONS and 75% of respondents were aware of the ONS, means we are well positioned as trusted data guardians.

Furthermore, close relationships and collaboration with business, charities, public sector and many other organisations are useful for taking a wider range of views into account, and we recently established the National Statistician's Expert User Advisory Committee (NSEUAC) to bring us even closer to our users and further build their diverse needs into our approach.

[2] https://uksa.statisticsauthority.gov.uk/the-authority-board/committees/national-statisticians-advisory-committees-and-panels/national-statisticians-data-ethics-advisory-committee/
[3] https://uksa.statisticsauthority.gov.uk/digitaleconomyact-research-statistics/research-accreditation-panel/
[4] https://uksa.statisticsauthority.gov.uk/news/public-confidence-in-official-statistics-continues-at-high-level-uk-statistics-authority-reacts-to-latest-survey/

Professor Goldacre told us of the risk of using pseudonymised health data and advocated its eradication. How much of a problem is pseudonymisation across the rest of the UK's data ecosystem and how can any risks be dealt with?

There are trade-offs when it comes to pseudonymisation. The ONS is a world leader in trying to mitigate both risks: the risk to privacy that Professor Goldacre highlights, and the risk of lower quality research. We have done so by investing in the Five Safes Framework. The wider controls within a TRE and the Five Safes Framework combine to reduce the risk of reidentification and enable the safe use of detailed microdata.

Should Trusted Research Environments be networked and if so, how can this be done safely and effectively?

While in principle Trusted Research Environments (TREs) can be networked, we would not recommend this as this would increase security risk without improving access to, or use of, data.

However, TREs must work closely together to ensure that, where there is a need to combine data held in different environments (for example to create a UK-wide Integrated Data Asset, from data held separately in the four nations, or combine Health and Administrative records), data can be securely transferred, linked, and made available for analysis. This requires that the organisations responsible for the TREs develop the technical mechanisms, and related Information Governance, to enable such collaboration and to so safely, securely, and ethically, in line with established technical, security and capability standards for processing of data for research purposes.

The value of such an approach is recognised by all major TREs in the UK, with good examples of success during the COVID-19 pandemic, and we continue to work closely together. For the ONS, this is an essential component of the IDS.

How should science research be defined for the purposes of data sharing?

The Digital Economy Act 2017 (DEA) Code of Practice does not explicitly define this but does say that the research powers support: *"helping researchers and policy-makers build a better understanding of how people live their lives, their patterns of need and use of different services and the resultant outcomes, to support the design and delivery of more effective and efficient public services."*[5]

It explains that data can be used if the purpose "*serves the public good*", meaning the use must do one or more of the following:
- provide an evidence base for public policy decision-making.
- provide an evidence base for public service delivery.

---

[5] https://www.gov.uk/government/consultations/digital-economy-act-part-5-data-sharing-codes-and-regulations/research-code-of-practice-and-accreditation-criteria#part-2-accreditation-criteria

- provide an evidence base for decisions which are likely to significantly benefit the economy, society or quality of life of people in the UK, UK nationals or people born in the UK now living abroad.
- replicate, validate, challenge or review existing research and proposed research publications, including official statistics.
- significantly extend understanding of social or economic trends or events by improving knowledge or challenging widely accepted analyses.
- improve the quality, coverage or presentation of existing research, including official or National Statistics.

We would highlight that the focus of this framework is on the use and the public benefits deriving from the research, regardless of professional background or academic discipline.

Do you have any concerns regarding the Government's proposals regarding the use of AI?

While giving evidence to the Committee on 8 June, we briefly discussed AI and its challenges for transparency[6]. I noted then that I would like to see the various groups thinking about the ethics of AI convened by government to create "*a set of clear ethical procedures.*" I also highlighted the need for good regulation of AI by regulators with an understanding of both the models and the data being used.

What progress is being made on digital skills in government and what role are you playing in improving them?

We have recognised the need to increase data skills and awareness in government, and through our Data Science Campus, we aim to build data science capability across the public sector. This work is aligned to the National Data Strategy and is achieved through a range of programmes and community engagement. In 2021/22 the Data Science Campus has provided data science learning to over 6,000 people across a wide variety of organisations and experience levels.

At the top of government, the Data Science Masterclass for Senior Leaders is being rolled out to senior leaders, predominantly permanent secretaries and Senior Civil Servants with plans to expand across the public sector. Over 4,500 learners have registered across 48 departments and agencies, with further cohorts planned. We have also developed a core data science foundational curriculum and learning pathway that is delivered across the public sector, including the Civil Service Fast Stream. This is an optional part of the Fast Stream curriculum and mandatory for 300 generalists, thereby increasing the data skills of future leaders.

The Campus also support more technical training such as the specialist Data Science Graduate Programme, a two-year programme covering core data science skills available to existing civil servants and new joiners. We are also working with academic partners to deliver a government designed master's programme (MDataGov) and to provide placements and secondments for students at all levels.

---

[6] https://committees.parliament.uk/oralevidence/10352/pdf/

Finally, we run accelerator and mentoring programmes. In 2021 the accelerator programmes in data science and data visualisation had a cohort of 57 from across a variety of public sector organisations. Our Community Programme supports the data science specialist community across government and recently delivered a hugely successful Data Science Festival which had over 90 speakers, presentations, a hackathon, training sessions, community and social events delivered virtually to over 3,000 participants.

Through my role as Head of the Analysis Function, we aim to integrate analysts in all facets of government. The proliferation of data has expanded the scale of the task but also means analysis can have a greater impact than ever before.

To take advantage of the opportunities presented by analysis (and avoid pitfalls), it is necessary for analytical skills to extend beyond traditional professional boundaries. By operating as a function, professional analysts will act as the catalyst for other professions contributing to analytical insight. This will be built on multi-disciplinary partnerships and developing capability.

To that end, we recently undertook an audit of the analysis skills within the Policy Profession, which highlighted where we need to focus our attention on continuing to mature these skills. We will use the recommendations in the report to continue work together to drive enhanced analytical skills across government.

I hope this is helpful to the Committee, and please do let me know if I can assist further.

(27th June 2022)