Office for National Statistics

UK Statistics Authority
1 Drummond Gate
London
SW1V 2QQ

Alison Pritchard

Deputy National Statistician and Director General, Data Capability

Rt Hon Greg Clark MP
Chair, Science and Technology Committee
House of Commons
London
SW1A 0AA

31 January 2022

Dear Mr Clark,

I write in response to the Science and Technology Committee's call for evidence for its inquiry '*the right to privacy: digital data'*. Our submission focuses on the questions raised in the call for evidence, including:

- The potential benefits, including to research, of effectively using and sharing data between and across Government, other public bodies, research institutions and commercial organisations, and the existing barriers to such data sharing.
- The ethics underpinning the use and sharing of individuals' data in health and care contexts.
- The extent to which appropriate safeguards and privacy are applied in the usage and sharing of individuals' data.
- The effectiveness of existing governance arrangements.

This inquiry is particularly relevant to the Office for National Statistics (and, in turn, the UK Statistics Authority) as we progress in two areas: data sharing and data ethics. As detailed in the annexed evidence, the benefits of data sharing are extensive; our accredited Secure Research Service (SRS) currently ensures researchers can access data safely, in line with the five safes framework, and we hope to harness the potential of data sharing further through the Integrated Data Service (IDS), a cross-government initiative we are leading.

Regarding data sharing barriers, negotiations to secure agreement to share data, and the associated work to put in place the legal, security and other arrangements required, can be protracted and complex, including with willing partners. We believe there is potential for elevating the consideration of risk for data sharing beyond silo departmental perspectives to better reflect the value of that data use elsewhere.

When it comes to data, the importance of individual privacy and security cannot be understated, and we elaborate on the many safeguards we put in place to protect this. Finally, following the Public Administration and Constitutional Affairs Committee recommendations in 2019[1], the Authority has taken a leading role in data ethics,

[1] https://committees.parliament.uk/publications/7137/documents/75480/default/

including through the National Statistician's Data Ethics Committee (NSDEC), and we explain this further in the annexed note.

I hope this is useful, and please do let me know if we can provide further evidence or discuss directly with the Committee.

Yours sincerely,

Alison Pritchard
Deputy National Statistician and Director General for Data Capability, Office for National Statistics

Office for National Statistics written evidence '*the right to privacy: digital data*', January 2022

Summary

The Office for National Statistics (ONS) has demonstrated the value of data sharing and access to data over the course of the pandemic and will go further through the Integrated Data Service (IDS), all the while within the parameters set out in the Digital Economy Act 2017 (DEA). However, 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 are ambitious and radical in our pursuit of greater data access for better analysis (and therefore, better policy); with this in mind, our ethics framework is extensive and carefully considered. In addition to our ethical principles, which we enact through both a self-assessment tool for researchers and the National Statistician's Data Ethics Committee (NSDEC), we follow the required legal safeguards, and together these ensure the ONS is well-equipped to use, link and share data for the public good.

The potential benefits, including to research, of effectively using and sharing data between and across Government, other public bodies, research institutions and commercial organisations, and the existing barriers to such data sharing

Effective data use and sharing has many benefits to government and society. An integrated, collaborative approach to data use enables in-depth and efficient analytical and research practices across organisational and sector boundaries. This directly informs government and society, supports the formulation of cost-effective, transparent public policy, and ultimately, delivers better outcomes which can be evaluated.

It is important to note, as we will do in more detail in other sections of this submission, that we only use and share data for research purposes and ensure that data is anonymised as early as possible. Building and ensuring public trust in data sharing is the most important factor when considering the benefits and barriers, and we currently do this through the application of ethical principles and legal safeguards. We are absolutely committed to increasing levels of public engagement and involvement to improve awareness, understanding and trust. We are developing an engagement approach with other key partners so the public not only understand the framework that is in place to safeguard their data but are also alive to the huge benefits that effective data use and sharing can bring.

The ONS is leading a cross government initiative to develop the IDS, which will bring together ready-to-use data to enable faster and wider collaborative analysis for the public good. It will provide the facility to fully exploit the opportunities for safe and secure access to data provided for in the DEA. There is a clear need for such a service: the COVID-19 pandemic illustrated the potential for government and public services to use and share data to help and protect people. When data are shared effectively, the speed at which analysis can be done means time-critical policy issues can be understood and addressed quickly; the ONS demonstrated this on a regular basis throughout the pandemic in partnership across government and with other organisations, using ONS's trusted research environment (TRE), the Secure

Research Service (SRS). Recognising the importance of TREs to research, we have agreed to support and participate in HDR UK's pan-UK Data Governance Steering group to help streamline data governance approaches for better data linkage for health data research.

The IDS will build on the success of the SRS, which has successfully and securely hosted de-identified data for over 15 years but is reaching its capacity while becoming increasingly costly to run. The IDS will reshape the way that data users share assets and offer controlled access to integrated de-identified data assets, a broader range of data to address policy analysis and evaluation and will work with the Central Digital & Data Office (CDDO) to ensure common data standards, governance, and quality measures. The IDS is a fully cloud-based platform, which will enable connectivity with data where they are held, reducing the friction caused by sharing data multiple times and the complexity of multiple data sharing agreements. This will organically increase the appetite to share data, as it will not need to be moved across the governments data estate and will therefore reduce cost burden on the supplier.

The IDS programme has just concluded its first year, with its most recent success the conclusion of the cloud procurement activity, which led to Google Cloud Platform (GCP) being awarded the contract to take forward the next stage of its development. Migration from the existing provider will conclude in spring and allow for a userbase of 40 across government and the devolved administrations. Throughout 2022, capability, data available and projects will significantly increase, and its userbase will be broadened and increased to around 1000 users by autumn, to include users outside of government.

As part of IDS's private beta, which launched in September 2021, the service launched 3 collaborative projects with partners across government to further demonstrate the benefits of data use and sharing, as well as having an environment to collaboratively run analysis. These projects, on key topics such as climate change, remain ongoing, and the programme is currently working across government to develop a portfolio of projects with a view to significantly scaling up use of the service in 2022, aligned with the aforementioned numbers.

In the meantime, government can and are already linking administrative data to inform policymaking, using powers from the DEA which facilitates the linking and sharing of de-identified data by public authorities for research purposes. For example, Longitudinal Educational Outcomes (LEO) is a de-identified, person level administrative dataset that brings together education data with employment, benefits, and earnings data from DfE, HESA, DWP and HMRC. We used this data to publish initial findings understanding earnings outcomes for free school meals students[2], and Ofsted is using it to look at the impact of quality of schooling on long-term labour market outcomes.

Other examples of linked datasets include the Data First programme, which links administrative datasets from across the justice system and beyond for research such as investigating racial bias in court case outcomes in England and Wales. Finally, Growing up in England (GUiE), which links 2011 Census data from the ONS with

[2] https://www.ons.gov.uk/releases/educationandsocialmobilityunderstandingearningsoutcomesforfreeschoolmealsstudentsinitialfindings

educational attainment data from DfE will be used by the London School of Economics to build up new quantitative evidence on Gypsy, Roma and Traveller young people. These projects give an idea of the real potential and benefits that can be gained when we effectively use and share data across government and beyond, and the IDS should only make these analyses easier in the future.

Opportunities to accelerate

While the legal right of access to data is an important enabler, many prospective data suppliers face difficulties associated with the provision of their data, especially when datasets are to be shared for the first time. There are several challenges to data sharing most data suppliers experience: risk appetite; capacity; legal clarity (including GDPR) and governance.

In terms of risk appetite, when we know the environment is safe and secure, there is still too much weight on the risk of data sharing, as opposed to the very real risk of policy harm where valuable data is not being actively used and shared. There needs to be a cultural shift in approach, which the IDS should contribute to, and the Committee's support will also be useful in this regard.

When we talk about capacity, one issue is personnel turnover, particularly when new decision-makers have differential understanding of legal frameworks or approaches to risk management in the context of data sharing, which can introduce further delays and blockages.

Lack of legal clarity can also play a role. For example, the current approach in data handling is to favour individual departmental boundaries, which is not suited to multi-departmental use of data. Particularly where infrastructure and governance are compliant with data protection legislation, it should be possible to treat such organisations as non-third party. Onward sharing of linked datasets is often another challenge a data supplier may face. Data are often brought together from separate departments to undertake topic-based analysis in areas such as education, labour market or health. If these linked data are then shared, the data governance models of the supplier departments may differ, adding further complexity due to the lack of standard data governance across the government.

These obstacles are not uncommon to most government departments, public bodies, and commercial organisations when it comes to data sharing. The ONS actively supports data suppliers and works with them to explore all options that enables the sharing of data and reduces the burden on the supplier, but there can still be tensions.

The IDS should help begin the shift from this mindset, but we need buy-in and recognition of the opportunity data sharing can present to do this with great success. Happily, the IDS should also mean that other traditional complexities such as the lack of a common gateway for sharing and lack of incentive to share data will soon be rectified. For example, the IDS will look to incentivise departmental sharing through wider access to a broader range of government data from partnering departments. The IDS data stewardship model considers the importance of data security and has the appropriate ethical controls to ensure privacy are at the heart of its design.

The ethics underpinning the use and sharing of individuals' data in health and care contexts

The importance of data ethics in the use and sharing of individual's health data cannot be understated as we enable this data to be used in ever more radical, ambitious, inclusive and sustainable ways, as set out in our strategy[3]. It is also crucial that the UK Statistics Authority (the Authority) guarantees public trust and acceptability and reduces potential harm to individuals involved in research; both of which can only be ensured through the application of ethical principles.

The Authority has developed the following ethical principles, which all ONS research projects must comply with, including uses of health and care data for statistical purposes:

1. The use of data has clear benefits for users and serves the public good.
2. The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately.
3. The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.
4. Data used and methods employed are consistent with legal requirements such as Data Protection Legislation[4], the Human Rights Act 1998, the Statistics and Registration Service Act 2007 (SRSA) and the common law duty of confidence.
5. The views of the public are considered in light of the data used and the perceived benefits of the research.
6. The access, use, and sharing of data is transparent, and is communicated clearly and accessibly to the public.

The extent to which appropriate safeguards and privacy are applied in the usage and sharing of individuals' data

Data Acquisition

The SRSA (as amended by the DEA) provides the ONS with statutory powers to acquire data for the production of public good statistics. These powers are governed by numerous safeguards to ensure the safety and confidentiality of personal information. For example, the ONS must ensure the data it receives is both necessary and proportionate for the intended purposes and all acquisitions must be GDPR compliant. This is set out within a published Code of Practice[5] and statement of principles[6]. The principles require that the ONS must work collaboratively with suppliers to reach a point of mutual agreement.

[3] https://uksa.statisticsauthority.gov.uk/statistics-for-the-public-good/

[4] "Data Protection Legislation" means the full, applicable data protection framework as set out in the Data Protection Act 2018. This encompasses general processing (including the General Data Protection Regulation and the applied GDPR).

[5] https://www.gov.uk/government/publications/digital-economy-act-2017-part-5-codes-of-practice

[6] https://www.gov.uk/government/publications/digital-economy-act-2017-part-5-codes-of-practice/statistics-statement-of-principles-and-code-of-practice-on-changes-to-data-systems

The separate but complimentary Code of Practice for Statistics[7] is based on three pillars: trustworthiness, quality and value. Data governance[8] is a key principle of the Code, requiring the ONS to follow their statutory obligations around data collection, confidentiality, and data sharing.

The primary function of the ONS is the production and publication of official statistics for the public good. All data collected and acquired by the ONS may only ever be used for its statutory functions and never for decisions about an individual or business. Data that is collected is anonymised/de-identified at the earliest possible opportunity as the ONS' focus is never on specific individuals but on the broader application of data to produce statistics that reflect society as a whole.

Before obtaining any data from another organisation, an agreement is put in place between the relevant parties that establishes how the sharing will take place. These agreements ensure the confidentiality of the data is maintained by prescribing secure transfer methods, storage requirements, secure access control and retention periods. Such agreements work to ensure adherence to the safeguards set out in both legislation and relevant Codes of Practice referred to above.

Legal safeguards

The SRSA contains safeguards to ensure that we do not disclose personal information. Namely, information that identifies a particular person (whether living or deceased) or a body corporate must not be disclosed by the ONS unless there is a statutory exemption that applies.

In accordance with the DEA, the ONS may make certain data it and other organisations hold available to others for research purposes. This is achieved through the accredited platform the SRS[9]. The SRS provides a safe setting, as part of the Five Safes Framework[10], to protect data confidentiality. The framework is a set of principles adopted by a range of secure labs, including the ONS, which provides complete assurance for data owners. The Five Safes are:

- Safe People: Only trained and accredited researchers are trusted to use data appropriately.
- Safe Projects: Data are only used for valuable, ethical research that delivers clear public benefits.
- Safe Settings: Access to data is only possible using our secure technology systems.
- Safe Outputs: All research outputs are checked to ensure they cannot identify data subjects.
- Safe Data: Researchers can only use data that have been de-identified.

[7] https://code.statisticsauthority.gov.uk/wp-content/uploads/2018/02/Code-of-Practice-for-Statistics.pdf
[8] https://code.statisticsauthority.gov.uk/the-code/trustworthiness/t6-data-governance/
[9] https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme
[10] https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme#the-five-safes

All data processing by the ONS is guided by our Data Strategy[11] *'to mobilise the power of data to help Britain make better decisions'* and is fully supported by a range of data policies, principles and standards.

Before undertaking projects that require the large-scale processing of personal data or the processing of special category personal data, we undertake data protection impact assessments (DPIAs). This ensures that privacy is built into the design of a project from the outset and allows us to identify and mitigate any privacy concerns as early as possible.

Prior to the publication of statistical information, thorough disclosure control methodology[12] is applied to ensure that we do not reveal any personal information. This requires a careful balance of risk and utility of the data, i.e. ensuring that the data is still functional for users, but that sufficient safeguards and measures have been applied to eliminate the risk of disclosing personal or sensitive data.

The ONS is committed to full transparency in relation to its processing of identifiable data. Extensive privacy information is made available to data subjects, whether through dedicated privacy information for survey respondents or more generally. We have a published log of data acquisitions on the ONS website[13]. Every effort is made to ensure data subjects are aware of how to contact us for further information or to exercise their data subject rights.

Our ONS Security framework both governs and operates a security strategy, security principles and a security policy framework, which incorporates and references appropriate recognised security standards and guidance from within UK Government (Cabinet Office, National Cyber Security Centre [NCSC], Centre for Protection of National Infrastructure [CPNI]) and international standards.

We have a dedicated data protection officer who advises the organisation in line with legislation and the organisation's data protection policy, supported by an experienced and qualified team that manage information rights requests and staff training in security and data protection.

We also lead several committees that focus on scrutinising how data is used across the organisation including the Data Governance Committee and the National Statistician's Data Ethics Committee (NSDEC).

The effectiveness of existing governance arrangements, e.g. the Centre for Data Ethics and Innovation

There has been much interest in data ethics over the last couple of years. Much of this work has focused on the 'theory' of data ethics and has fallen short of making applied decisions about the ethics of the use of data for analysis.

The Authority's work on data ethics, through the UK Statistics Authority's Centre for Applied Data Ethics, has focused on supporting analysts to apply high level ethical principles/frameworks to their work and providing them with the support to be able to do this efficiently, to enable timely research that is ethically appropriate. This has

[11] https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy
[12] https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol
[13] https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/sourcesofdata

involved creating and user supporting an ethics self-assessment tool, which empowers researchers to themselves apply the Authority's data ethics principles to their research projects. This enables researchers to identify and mitigate against ethical risks in their projects with support from the UK Statistics Authority's Centre for Applied Data Ethics' expert user support services, published guidance and training. This tool has had a significant impact and has been widely used since its inception, with 258 projects last year using the tool, including those from academia, the ONS, other government departments, and devolved administrations.

In circumstances where research teams require additional independent ethical advice and support to mitigate against significant ethical risks identified via the ethics self-assessment process, we also provide expert ethical advice through NSDEC which provides transparent and timely ethical advice and assurance to the National Statistician. Over the last year, 17 projects including the COVID-19 Schools Infection Survey (ONS), a study to assess the algorithmic feasibility of COVID-19 specific vocal biomarker detection (JBC) and research to determine the population-level relative risk of hospitalisation or death that COVID-19 presents to people with different socio-demographic characteristics and co-morbidities (ONS), have been escalated to NSDEC to seek expert independent advice and assurance on those most pressing ethical issues that the ONS and the wider research community has faced. Our experience has taught us that providing this applied data ethics support plays a vital role in empowering analysts to use data in innovative ways that are ethically appropriate and efficient.

Office for National Statistics

January 2022