

ADR UK Response to National Data Strategy Consultation

December 2020

About ADR UK

ADR UK (Administrative Data Research UK) is an Economic and Social Research Council (ESRC) investment initially from July 2018 to March 2022. The partnership is supported by £59 million drawn from the National Productivity Investment Fund (NPIF) via the ESRC, part of UK Research and Innovation (UKRI).

Its objective is to transform the way researchers access the UK's wealth of public sector data, to enable better informed policy decisions that improve lives. By linking together administrative data held by different parts of government and making it available for researchers, we are enabling vital research that has the potential to lead to more effective public services, in areas from improving education and healthcare to tackling crime.

ADR UK is made up of three national partnerships – ADR Scotland, ADR Wales, and ADR Northern Ireland – and the Office for National Statistics (ONS), coordinated by a UK-wide Strategic Hub. Each partnership is made up of academic and government partners, as well as dedicated secure services through which approved researchers can access de-identified administrative data.

To find out more, visit [the ADR UK website](#).

The below forms our response to the National Data Strategy's second consultation (December 2020). This accounts for the interests of all ADR UK partners, including those within each of the devolved nations – ADR Northern Ireland, ADR Scotland and ADR Wales – and on a UK-wide level.

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.

ADR UK **somewhat agrees** with the statement. There is more detail required on what government will do practically to ensure all departments commit to helping deliver the vision of the National Data Strategy.

In our [response to the initial consultation](#) for the Strategy in July 2019, ADR UK stressed the following three key points:

- 1) **There is a need for a greater appreciation of the necessity to improve data use within government for the purpose of research.** Greater investment and commitment are needed to ensure that data is handled properly and that the potential benefits of the government's wealth of data for research are maximised, particularly for informing policy decisions. This is necessary to avoid data 'missed use' as well as the consequences of data misuse.
 - We are pleased to see a commitment to driving greater use of data in Pillar 3 of the Strategy, 'Data availability', as well as under Section 6 of the Strategy, 'Availability: ensuring data is appropriately accessible'. In particular, we welcome the Strategy's commitment to *"drive use of the Digital Economy Act (2017) powers, as well as addressing barriers to data sharing more widely"*, as well as the third of five *"concrete and significant opportunities for data to positively transform the UK"*, namely: *"Increasing the speed, efficiency and scope of scientific research"*. This point is inextricably linked to each of the four other opportunities listed, particularly the fourth: *"Driving better delivery of policy and public services"*, which lies at the heart of ADR UK's work enabling greater access to administrative data for policy-relevant research.
 - We would stress, however, that beyond life sciences – which are noted as having *"some of the clearest examples of the benefits to society"* (section 2.3) – economic and social research is equally important, encompassing every aspect of our day-to-day lives, from education to social justice and beyond. Facilitating greater access to all forms of public sector administrative data for research in the public interest is therefore essential to drive progress and improve experiences in all areas of life.
- 2) **There is a need for clear overarching governance structures** to emphasise, coordinate and enforce the need for better use of data across government as a whole, including the ethical and responsible use of data. This governance structure should incorporate the use of research evidence for policy, perhaps in a similar manner to the governance of evaluation in policy through HM Treasury's [Magenta Book](#).
 - We are pleased to see this addressed under section 4.2.3, 'Productivity and accountability', but would appreciate an involvement in how such governance structures could be designed and implemented. There is significant expertise across the ADR UK partnership that the teams implementing the Strategy could usefully draw on.

- 3) **There is a need for clearer structures for enforcing the ethical and trustworthy use of public data, with a focus on public engagement.** Increasing public trust in government's use of data, and raising awareness about the benefits of better sharing of data, is essential for the success of government data initiatives.
- We are pleased to see this is somewhat covered under pillar 4, 'Responsible data', and under section 7.
 - However, although the Strategy demonstrates a clear commitment to transparency and public trust, this appears to be predominantly focused on a one-way dissemination of information. While transparency and public trust are essential, particularly for increasing public understanding around how and why data is collected and used, the Strategy has far less focus on the need to explore and properly account for the interests and concerns of the public. This should be done via public consultation and public participation activities which **directly seek the views and input of relevant data subjects**. The Strategy stresses the importance of "*a regime that reflects what people really care about*" (section 7); however, it does little to describe how 'what people really care about' will be established.
 - To fully understand people's interests and concerns, **the specific communities and demographics relevant to initiatives making use of their data should be consulted, but also ideally involved in the development of those initiatives**, via established participatory activities such as public panels, workshops and focus groups in which views are gathered and accounted for. For example, ADR UK currently consults members of the public on aspects of our work via [Public Panels](#) in both Wales and Scotland, and consults relevant community representatives such as NGOs, VCOs and practitioners via stakeholder workshops and project-specific **Community Representative Panels**, to ensure the research being conducted meets the specific needs of data subjects.
 - We suggest there needs to be a stronger focus on **accessibility, accountability and impact**. It should be clear in the pillars that the data the government is proposing to make safely and securely available belongs to the public, and as such the public needs to be specifically considered at all levels and stages of the Strategy's implementation, beyond the consultation stage. It is critical that this is not the last time that bodies representing publics are engaged in the use and development of data and its accompanying infrastructure.
 - More recognition could also be given to the **Digital Economy Act's [Research Code of Practice](#) and its 'public interest' requirement for research project accreditation**. Established practices are now in place to review proposals against public good and ethical standards, carried out by an independent [Research Accreditation Panel](#) established by the UK Statistics Authority, and its proceedings and decisions are published. Whilst the framework may be complete for administrative data research, the model it provides might be useful to address similar issues in health and social care research.

We are also pleased to see a focus on **data skills** in section 5: 'Data skills for a data-driven economy and data-rich lives'. The development of data skills is essential to maximising the potential of data for public good, which is why each ADR UK partner is involved in training

researchers to use the administrative datasets made accessible within our Trusted Research Environments. Moving forward, this focus also aligns with ADR UK's upcoming establishment of a new **Data and Public Policy Centre for Doctoral Training (CDT)**, which will fund doctoral students to address significant public policy and public service challenges. This will be supported by funded research fellows to ensure there is senior academic expertise to nurture the doctoral students and maintain the talent pipeline. In addition, ADR UK is working with ESRC to ensure ESRC-funded PhD students are trained to use ADR UK datasets. The doctoral students within the CDT will also have access to a range of other data sources from programmes across UK Research & Innovation (UKRI), to strengthen the pipeline of researchers trained to analyse complex datasets.

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.

The impacts of the Covid-19 pandemic will not be limited to public health, though that is the sharpest and most immediate catastrophe. We are already witnessing a large-scale loss of jobs and the impact is not only on the economy as a whole, but on people and communities. These effects are likely to ripple for some time, impacting across education, employment, poverty and deprivation, and mental health. **The strategic and widespread use of data will be crucial in order to understand and mitigate these various but intertwined impacts of the pandemic.** This means that researchers, policymakers and service providers will not only need access to data, but the ability to link large-scale datasets together – specifically, provision should be made to link health data to other datasets across the UK wherever this is not currently possible.

ADR UK has a number of examples of how data has already been used to deliver public benefits during the Covid-19 pandemic:

Scotland

The Urban Big Data Centre – directed by Scottish Centre for Administrative Data Research Co-Director Nick Bailey – used new forms of data to understand social distancing/compliance and economic impacts (use of CCTV image analysis for pedestrian activity; traffic sensors for road traffic; Google and Apple mobility products; Zoopla for housing market activity; Airbnb for tourism activity).

Wales

During the height of the pandemic, ADR Wales carried out **analysis of the Shielded Patient List in Wales as part of a [collection of rapid response projects](#)** to help inform thinking at government level. Around 130,000 people were on the shielded list in Wales by July 2020, with over 92,000 people living with those asked to shield, resulting in an estimated 223,000 people in Wales living in shielded households. The analyses used de-identified, linked data to look at the composition and characteristics of shielded households, the situation of children living within shielded households, what access shielded patients had to outdoor space during the pandemic and also the number and percentage of teachers and teaching assistants shielding in Wales. Findings from these publications directly informed the One Wales response to Covid-19 which feeds into Wales Technical Advisory Group (TAG) and UK SAGE updates. This work has allowed ADR Wales to flag unintended consequences of Covid-19 restrictions on other areas of life and for specific groups, made possible only by linking together different data sources.

In addition, **ADR Scotland and ADR Wales have made requests for data to support management of the pandemic for non-health purposes.** Both the Scottish and Welsh governments have made cases to the Department for Work & Pensions (DWP) that articulate their short, medium and long-term requirements for data to support them in managing the Covid-19 pandemic. These requests are being managed by the ADR UK Strategic Hub and the Office for National Statistics (ONS), as they are related to ongoing work being taken forwards with DWP and HMRC to test ways of linking data across multiple ADR UK data infrastructure sites. Further details of the cases made to DWP by the Scottish and Welsh Governments for access to DWP benefits data to support their management of the pandemic can be provided to the National Data Strategy team on request.

Although there is some engagement at a working level with DWP on taking this work forward, this may not result in action in time for any practical benefit on the management of the pandemic to occur. A rapid shift in culture for opening up access to data, which the National Data Strategy acknowledges in general terms as being an issue, would be required.

There are other areas in which progress between ADR UK and DWP has been slow; for example, the **Race Disparity Audit (RDA) dataset**. This dataset brings together 2011 Census data with Benefits and Income data from DWP and HMRC at a record level, creating a large and informative research resource. There are many more insights to be gained through increasing access for further analysis, which would support further insights into the Covid-19 pandemic response, if the data were updated. This could include, for example:

- Research into benefits, benefit reform and interactions with benefits;
- Longitudinal analyses of outcomes for vulnerable populations;
- Household and housing analysis.

This audit has already helped to inform our understanding of differences between ethnic groups, and to identify those public services where disparities are diminishing and those where work is needed to develop effective strategies to reduce disparities between ethnic groups. HMRC and ONS have already given their approval for this data to be made available to external researchers, via the ONS Secure Research Service (SRS), and we continue to work collaboratively with DWP to gain their approval. This is a good example of the kind of process that would benefit from being accelerated in order to enable timely access to data to support research in the public interest.

Northern Ireland

Academic partners in Northern Ireland have expressed concern that the research community as a whole is experiencing a **barrier to delivering needed public benefit during the Covid-19 pandemic due to limited remote access to data**. While the centralised collation of and easy access to health data is positive, it does not currently extend to datasets outside of health and social care, which means opportunities to deliver public benefit from data usage are being missed. The need for harmonised data across all four administrations of the UK forced some uncomfortable comparisons for Northern Ireland, but has ultimately been useful in demonstrating precedents from across the UK from which Northern Ireland can learn. Work is now being taken forward by government partners in Northern Ireland to assess options for improved access going forward in light of this.

In general, **better availability of data from mobile phones could also make analysis of mobility and social activity across the UK much more effective**. Several of the pandemic modelling groups were looking for this and some was made available by phone companies, albeit partial and limited. Government could take a more strategic approach to identify data needs and use of existing powers to bring this into public use. For example, mobile phone providers are regulated so could be required to provide suitably anonymised but highly granular extracts for public sector use as part of their licence. Public transport operators are licensed and could be required to provide similar extracts for public sector use.

There can be an assumption that, because research and analysis outputs are non-disclosive, the input data can also be non-disclosive. However, the best research and analysis value arises from the use of highly granular data, which may be personal data. The GDPR and Information Commissioner's Office (ICO) guidance says this is consistent with the legal framework for scientific research and statistics, but government departments need a better understanding that data inputs need to be highly granular, and can be, in order to create sophisticated analysis.

Furthermore, the Covid-19 pandemic has reinforced the fact that **an understanding of social and economic circumstances is important for a proper understanding of health**, and vice versa. To more fully understand the characteristics of those who are more likely to contract and suffer serious health harm or death as a result of Covid-19 – particularly those communities disproportionately affected by the disease – we need better access to administrative records to

link testing, health and socioeconomic data. Health and socioeconomic factors are intrinsically linked, and the way data is used to tackle the pandemic should reflect this; for example, to ensure that testing sites are located in and easily accessible by the most deprived communities.

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?

If done without proper academic rigour and due consideration for people with a protected characteristic under the Equality Act 2010, it is possible some data linkage projects could lead to policies or practices that have disproportionate negative or discriminatory unintended impacts on such individuals and communities – for instance, because of pre-existing bias in the recording of source data if this is not adequately corrected for. This risk is more relevant in relation to operational uses of data, but is a concern we must also be mindful of in the research space.

ADR UK is committed to ensuring this does not happen as a result of the research we enable; for example, by **funding methodological evaluations of linkage quality and ‘missingness’** in the data. Also, by facilitating academics funded to carry out foundational research on linked datasets to engage with organisations able to represent the interests of people with a protected characteristic via **project-specific ‘Community Representative Panels’**. These panels are currently being set up for a number of our projects, with membership including relevant NGOs, VCOs and practitioners with the knowledge and expertise to allow them to consult with and represent these groups. We recommend that similar forms of engagement are encouraged or mandated across government data initiatives within scope of the National Data Strategy; a failure of one initiative to engage adequately or be sufficiently mindful of such issues could impact negatively on trust in government data use across the board.

However, it is also vitally important to consider the potential impacts of the Strategy *not* being delivered on individuals with protected characteristics. In the context of ADR UK’s work enabling access to public sector administrative data for research to inform policy and practice, for example, **academic researchers can add immense value to how data is used and interpreted**. The latest techniques can be used to assess data linkage results in the context of protected characteristics: where there are differences in data linkage match rates by these protected characteristics, analytical techniques can be put in place so that every subsequent user of this data (inside or outside government) can take account of these differences. Without this type of analysis, these groups can be mis-represented in results, leading to poor decision making.

Research using linked public sector administrative data has huge potential to tackle the disadvantages faced by groups with protected characteristics. For example, the ADR UK-funded and Ministry of Justice-led **Data First** project will provide valuable insight into the

experiences of vulnerable individuals who use the justice system. This insight can be used to help address inconsistencies across the system and make it work better to support everybody. In addition, the Race Disparity Audit is an example of how statistics and research can explore and describe inequalities, enabling policy departments to discover the differential effects of their policies and providing benchmarks for tackling them.

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?

We encourage the government to remember the **particular history and deprivation present in Northern Ireland** and the additional heightened sensitivities around data usage this may create. This reality is one reason why ADR UK partner Administrative Data Research Centre Northern Ireland (ADRC NI) puts public engagement at the heart of its work, and encourages the government to do the same around any data access initiatives in Northern Ireland. Opening access to and use of data has the potential to unlock unique solutions to the particular circumstances of Northern Ireland, but this must be done in such a way as to bring the public along in order to maintain the social licence of data usage.

It is important that **access to data is levelled up across the different regions of the UK**. Access to data should be geographically dispersed so that our knowledge generation economy is not overly focused in London and the south east. Local authorities need access to linked data for policy analysis as much as Whitehall does, to empower local decision making, and geographical location should not be a barrier to research. ADR UK is working hard to increase geographical access to the data we curate by rolling out further remote access via Assured Organisational Connectivity (AOC) agreements to access data within the ONS SRS, as well as through the respective data centres within the devolved administrations.

As part of the ADR UK investment, micro safe settings known as SafePods are currently being installed at universities throughout the UK. These operate under the same policies and procedures as other physical safe settings and provide accredited researchers with access to multiple participating data centre datasets, including ONS and the UK Data Service, widening the geographical availability of researcher data access.

In addition, ADR UK is supporting the Joint Biosecurity Centre (JBC) and the Ministry of Housing, Communities and Local Government (MHCLG) by funding a team that includes academics and ONS staff to enable localities in England to access detailed, local Covid-19 data and other datasets relevant to the management of the pandemic. This involves the creation of **'Local Data Spaces'** in the ONS SRS. These Local Data Spaces will facilitate safe and secure access to granular Covid-related data and help provide greater analytical opportunities and insights on the pandemic for Local Areas (LAs). This will further open up the potential of data to be impactful on a local level.

Scotland, Wales and Northern Ireland are particularly disadvantaged when UK Government departments do not provide household level social and economic data for research carried out in those nations to inform devolved policy. Any socioeconomic regional level of analysis needs household-level data as its building-block, so data points can be associated with Unique Property Reference Numbers and ONS National Statistics Geography, enabling the construction of data analysis outputs at a geography of executive action. The pandemic has shown that the data must be as granular as the level of decision-making, and usually this is at LA level. Even then, LAs need a sub-LA breakdown, usually down to Middle Layer Super Output Area (MSOA).

Mission one: Unlocking the value of data across the economy

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.

Better data availability is likely to benefit all sectors, but we have highlighted those the work of ADR UK itself will likely assist the most:

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

A key role of central government should be to ensure that **public sector administrative data is safely and securely made available to support research in the public interest**, leading to evidence-based policymaking and better decisions. To achieve this, there needs to be a central champion and incentives and/or penalties to drive a shift in data sharing culture across government. **Building public confidence and trust**, and using the data for public benefit, is key to this process. Central government must be a leader in developing a culture of safe data use, working across its own departments and encouraging the development of networks of data access and linkage to answer pressing questions to improve people's lives.

A key component of the Strategy needs to be around **tackling the conservatism of data owners**. Where there are legal gateways to support data sharing and linkage in the public interest, the default expectation should be that data should be shared; for example, when it has the potential to save time, money or lives, without any accompanying negative impacts. There should be an expectation that a public service would consistently measure (using data) what it delivers in terms of equity of access, quality of and outcomes from services offered.

More could be done centrally to **support individual departments to simplify their decision making processes** so that, whether or not a data share is seen as 'top priority' by the data owning department(s), a decision can be made on whether it should be shared. At present, it is often the case that a department will only make a decision if it is seen as 'top priority', because the decision making process is arduous for all involved. This means that lower priority decisions (which may be 'top priority' for another department or organisation) are unable to progress.

A possible solution to this is for **'how to' guides** to be published – by the Cabinet Office, DCMS or ONS with the support and advice of ADR UK and other stakeholders – to help data owners properly understand the legislation – particularly the 2017 Digital Economy Act – and make it clear what needs to be considered when sharing data, and what support is available. The ADR UK-funded ONS SRS, for example, is set up to support departments to enable access to anonymised data for approved researchers both within and outside government, to drive evidence-based policy and, ultimately, better outcomes. In addition, **skills development is important**, and departments should ensure that they have people in place to support activities related to data sharing and linkage, to support informed decisions being taken in a timely way.

Fundamentally, administrative data held by public bodies relates to interactions by members of the public or businesses with that organisation. The government should do more to maintain the social contract to utilise this data for research and decision making. Overall, **public opinion agrees that data should be used for work in the public good**, but there remain concerns about private sector organisations' motivations to use data. **Transparency around the use of the public's data and how it contributes to the public good is important.**

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

We are not aware of any reasons this role should vary across sectors and applications – if the issues around departmental risk aversion to data sharing and perceived complexity of the decision making process are solved, then this will benefit all sectors.

In addition, [public attitudes work](#) shows people are most supportive of data use in sectors and for applications where it is perceived to bring the greatest potential/tangible rewards, such as protecting people’s health or improving their income. This means government should be equally, if not even more proactive in promoting data use in these areas, even if they may traditionally have perceived these types of data as being more sensitive or contentious.

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.

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree**

To note, when piloting the most promising mechanisms to develop strong data foundations, ADR UK’s experience suggests **you should not necessarily focus on the perceived ‘quick wins’, if you want to make a tangible difference in this space**. Instead, we would urge the organisations tasked with delivering this Strategy to work with the departments and organisations that are most willing to work with you. For example, in ADR UK’s experience, helping a government department such as the Ministry of Justice solve some really fundamental issues around data access, quality and linkage by funding them to deliver a programme such as [Data First](#) has reaped far greater benefits and in a quicker timeframe than perceived quick wins, where all that was required was for data owning departments to agree to open up access to pre-existing datasets for research, when the data already exists in the ONS SRS.

Central Government must also make provisions to **ensure development of the safe and secure data foundations will add to rather than break down public trust**. Data misuse in one sector can fuel data mistrust in another sector; the use of data in the private sector can impact how the public perceive the use of public sector data. It is therefore incumbent upon government to require that any sector looking to use its own data or to access government-held public data be kept to a series of agreed-upon principles and rules, for which there is ample evidence including, for example, the [Caldicott Review](#) and ADR UK’s own [literature review](#) on public attitudes to data sharing.

'Building trust in statistics' began with the 1998 Green Paper 'Statistics: A Matter of Trust' and, leading to a new Framework for official statistics and the Statistics and Registration Services Act 2007, has transformed public confidence in official statistics. It took time, organisational change and legislation, but the process provides a model of what works in building public trust in data.

Without a proper infrastructure for data that includes legal, ethical and public trust considerations, it will be **very difficult to operate a data business**. The experience of open banking shows that, to move the market to innovate and put data to work for consumers, investment is needed. Investment will pump prime the creation of trusted, safe environments which implement data foundations and, subsequently, innovators and researchers can gain access to data to create the insights and services which will benefit consumers.

The Open Data Institute and Alan Turing Institute are good examples of government-supported institutions now having an important role in the wider economy. Such institutions should continue to be funded in order to ensure data foundations can be strengthened and maintained.

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?

This question relies on the assumption that 'consumers' are homogenous, and all want companies to be using their data in a personalised way. This assumption should not be operationalised without **further work into public attitudes on the commercialisation of their data**. Consumer attitudes change across industries and with experience. This should not be a one-time decision for all industries.

ADR UK and ESRC know from public engagement and consultations that there is an expectation that data held about persons should be used not only for the delivery of a primary service (such as payment of a pension), but also in the service of a wider public good. A secondary purpose such as scientific research and statistics is not just accepted by the public, it is expected by the public. We know that there is public support for secondary statistics and research use of personal data when assurances can be made about security of data, and that independent assessments are made of what constitutes 'the public good'. Government can go some way to meeting these expectations by publishing a data management plan for all secondary uses of the data it obtains from the public for its primary functions.

Mission two: Maintaining a pro-growth and trusted data regime

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

The GDPR and the 2018 Data Protection Act (DPA) are supportive of the better use of data for creating better evidence for policy and research in the public good. The provisions for scientific research and statistics are particularly valuable. However, they are still **poorly understood and have not penetrated far enough** into the thinking of departmental Data Protection Officers, Chief Technology Officers, Chief Data Officers, and Senior Information Risk Owners. The outcome is a lack of untabulated data available for integration into the information systems and products of their own department and for others.

An important feature of processing for scientific research and statistics is the shift in the biographical focus of the data. Data used as an input for these purposes may have had the data subject as the biographical focus, making these data personal data when they were processed for the administration of a public service. However, evidence for policy does not have any particular individual as its focus – by definition, policy relates to social groups, geographical regions, sectors of the economy, distributions of illness or disease, environmental conditions, and combinations of these factors. The processing of the data no longer 'relates' to any particular individual, and the data, in this context, is no longer personal data – even if the data is still held in a highly granular form. This is too rarely considered in departments when assessing if processed data is personal data.

Knowing what is *not* personal data is as important as what *is* personal data. When the risks are so high and the rewards are so low, there is an understandable reluctance in Data Controllers to conclude that a collection of personal data has been processed and shared in such a way that those data are no longer personal data. An understanding of what **"re-identification is reasonably unlikely"** really means is essential. The advice of the ICO is that account should be taken of not only the design of the data, but also of the context within which it is used, which is critically important to the use and re-use of data outside the framework of the GDPR. For as long as a public authority considers its untabulated data to be personal data due to a theoretically possible but extremely unlikely re-identification scenario, there will be little growth possible.

A much better understanding of the meaning of personal data *within the context of trusted research environments* is needed across government if we are to take advantage of the digital data avalanche to improve evidence for policy and public discourse. In addition, it is important to have a clear statement from central government/the ICO that access to data via trusted, safe research environments is compatible with the DEA and GDPR.

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.

Strongly disagree
Somewhat disagree
Neither agree nor disagree
Somewhat agree
Strongly agree

We somewhat agree with the statement. The CDEI should also maintain **public engagement** as a key function of its work, particularly when **determining what constitutes 'public interest' in data sharing and use** as part of its further work [addressing trust in public sector data use](#). Determining what constitutes public interest cannot be done without exploring public views on the matter. Public attitudes towards the public interest of data use is an area which ADR UK also hopes to explore further.

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

The CDEI would gain a number of valuable attributes if given statutory status. Other relevant organisations with such a status are able to issue binding Codes of Practice; to have a regulatory function; to be referenced in other legislation; to be answerable directly to Parliament; to issue their own annual reports; to have their own independent budget allocations; and to be led by a person appointed under the public appointments procedure. These attributes have proved valuable to, for example, the UK Statistics Authority, the Information Commissioner's Office and many others.

These attributes are fundamental to being able to deliver a challenging remit within Government with enhanced scrutiny of the effectiveness of that remit through transparency with the public and Parliament. If, for example, the positive potential of new data technologies in public services is to be realised, it may be essential that an independent body on a statutory footing is in place to design and impose standards and procedures – and to answer to the public and Parliament on what those standards and procedures are and how they are being implemented.

This is of interest to ADR UK, as a broader remit for the CDEI could include ethics and innovations in the research use of administrative data, removing one of the barriers to better access to data; at present, every data owner has their own interpretation and position with respect to the ethical and innovative use of data for research.

Mission three: Transforming government's use of data to drive efficiency and improve public services

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.

Each of the above listed areas of work are intertwined and together are crucial for success in enabling the better use of data across government. Two that we would consider paramount, especially given our partners' experience in this area, are: **Quality, Availability and Access**; and **Ethics and Public Trust**.

Ethics and public trust are underpinned by **standards and assurance**, which should be universal and independently monitored across the most rigorous international standards. The government needs greater transparency around how and why data is used, and a campaign to build trust, which is and has been easily eroded by incidents of data being lost, misused or otherwise inappropriately handled within both the public and private sectors. The Digital Economy Act 2017 (DEA)'s provisions for Accreditation of research purpose, person and place, for example – including an ethical review and transparency of decisions, all carried out independently under a Code of Practice – is a model which could be replicated in other disciplines.

However, as stated under Mission 2 – *"This regime will not be overly burdensome for the average company, nor will it be unnecessarily complex or vague... [or feature] undue regulatory uncertainty or risk"* – it should also be considered that when excessively complex, the measures associated with these five areas of focus may have an adverse effect on the ability of organisations to make the best of use of data for the public good. Is it important to ensure that these areas of work are **designed in such a way that they enable more efficient use of data, rather than create unnecessary hurdles** to maximising its potential for public good.

We have found that when a new standard and/or service is introduced with cross-government effect, the response of individual departments is to consider this to be an additional barrier, rather than a replacement for local actions and decisions. Thus, the intended effect of greater efficiency and more uniform and consistent decision making is not achieved, and in fact the procedures become even more protracted. For instance, even with the commencement of the research provisions of the DEA, no department of government has stood down its own decision

making procedures and documentation for research access to data; instead, the same procedures are maintained and the DEA procedures are seen as additional. To researchers, the DEA is an additional hurdle, because none of the pre-existing hurdles have been removed despite this being the intention of the legislation. For the National Data Strategy to be deliverable, **work should be done to systematically remove those pre-existing hurdles** to enable the originally intended effect of the DEA to be fully realised.

In addition, **wherever possible, government activity should have an independent research component** to help establish the effectiveness of the policy. The [HM Courts & Tribunals Service \(HMCTS\) response](#) to a report by Dr Natalie Byrom, Research Director at The Legal Education Foundation, on '[Digital Justice: HMCTS data strategy and delivering access to justice](#)', is an example of good practice. A small investment for the ADR UK-funded Data First programme within the justice reforms will have a very positive effect on the deployment of those reforms.

The [CDEI report into biases in algorithmic decision making](#) makes a recommendation that more use should be made of the expert analysis and high quality data found in the ONS SRS, and that the expert (and human) research and analysis carried out using the SRS should act as a gold standard against which biases in algorithmic decision making can be benchmarked. This important recommendation requires that barriers to making use of the SRS within – and between – departments are minimised.

Government should consider a 'colour book' for the use of independent research evidence in policy. Nesta's Alliance for Useful Evidence has produced excellent [guidance and advice](#), which (similar to the Magenta Book for policy evaluation) could be adopted into a good practice statement for government.

The case study below (page 17), from Nesta's practice guide on 'Using research evidence', is a clear example of the importance of timely access to evidence for social and economic research in the public interest.

In addition, *[Repeated from answer to Q6:]* more could be done centrally to **support individual departments to simplify their decision making processes** so that, whether or not a data share is seen as 'top priority' by the data owning department(s), a decision can be made on whether it should be shared. At present, it is often the case that a department will only make a decision if it is seen as 'top priority', because the decision making process is arduous for all involved. This means that lower priority decisions (which may be 'top priority' for another department or organisation) are unable to progress.

A possible solution to this is for '**how to**' guides to be published – by the Cabinet Office, DCMS or ONS – to help data owners properly understand the legislation – particularly the 2017 Digital Economy Act – and make it clear what needs to be considered when sharing data, and what support is available. The ADR UK-funded ONS SRS, for example, is set up to support departments to enable access to anonymised data for approved researchers both within and

outside government, to drive evidence-based policy and, ultimately, better outcomes. In addition, **skills development is important**, and departments should ensure they have people in place to support activities related to data sharing and linkage, to support informed decisions being taken in a timely way.

Case study: “The tragedy of professional ignorance over cot death research” (Nesta, *Using Research Evidence*)¹

Cot death is a horror that can haunt new parents. Fortunately instances of cot death, or to use the medical terminology Sudden Infant Death Syndrome (SIDS), have gone down. But one of the biggest of tragedies of cot death is that if we had looked at the research instead of listening to the experts, many babies’ lives might have been saved in Europe, the US and Australasia.

Following the advice of health professionals such as the best-selling Dr Spock, a whole generation of parents laid their babies face-down in the cot believing that they were doing the right thing. But according to Dr Ruth Gilbert of University College London in [an article in the *British Medical Journal* in 2008](#), by 1970 there was significant evidence from clinical research that putting babies to sleep on their front increased the risk of cot death compared with putting babies on their backs. The ‘Back to Sleep’ public health campaign had a dramatic effect on sudden infant death, but was not launched until November 1991, and the safer sleeping position was not consistently recommended until 1995.

Dr Gilbert believes that the advice to put infants to sleep on their front for nearly half a century was “contrary to evidence available from 1970 that this was likely to be harmful”. A systematic review of preventable risk factors for SIDS from 1970 would have led to earlier recognition of the harm of sleeping on the front and might have helped prevent some of the 10,000 infant deaths in the UK and 50,000 in Europe, the US and Australasia.

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.

Comparable metadata standards across different fields of data/research would be very useful, as would **metadata being publicly available** to facilitate researchers developing their proposals and describing how they would want to analyse it. In addition, the publication of **user guides** describing how to analyse datasets should be published as standard.

The Data Standards Authority operates very much as an organisation focused on UK central government at present. Although aspirations are much wider, it would have to develop its approach considerably in order to function as a body which operates across the public sector and UK as a whole – reaching out to support external users of the data in addition to those within government – with much deeper collaboration across these interests.

¹ Nesta, [Using Research Evidence: A Practice Guide](#), Alliance for Useful Evidence

Mission four: Ensuring the security and resilience of the infrastructure on which data relies

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?

Security standards, such as those that are required to gain accreditation as a data processor under the 2017 Digital Economy Act (DEA), should be promoted much more widely. This would allow data sharing between data owning departments and with DEA accredited processors to happen more easily, and for data to be shared across accredited processors where this would promote further research use of the data.

In addition, data infrastructure service providers should be required to adhere to the **'Five Safes'** to ensure that privacy is protected and data misuse is prevented: **safe people, safe projects, safe settings, safe outputs and safe data**.

The operating model of ADR Wales partner the SAIL Databank has recently been recognised by academics in a [BMJ Opinion piece](#) as being one to be replicated for the secure sharing of anonymised data for research:

(To) reduce data release delays through increased capacity and more specialised data providers. Independent, accredited data providers should be created, with expert processing and disseminating capacity, knowledge of how data are used in research, and understanding of how best to prepare and deliver datasets to researchers (emulating the successful Secure Anonymised Information Linkage (SAIL) in Wales).²

SAIL's responsiveness is owed to a **simplified approvals process**, a **close and collaborative relationship with data providers** such as NHS Wales Informatics Service (NWIS), and the **core funding** it receives via Health and Care Research Wales, as well as from ADR UK, reducing cost barriers to de-identified data access. Further details about the SAIL model and its benefits can be found in its [recent profile](#) in the International Journal of Population Data Science.

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?

Assessing the robustness of security protocols when choosing data infrastructure services is a complex issue for service providers and clients. Additionally, fragmentation across clients adds

² F. Cavallaro et al., 2020, [Reducing barriers to data access for research in the public interest—lessons from covid-19](#), thebmjopinion, 6 July (accessed 02.12.2020)

cost and complexity to clients and reduces the ability of service providers to support a consistent standard. Establishing **consistent assurance levels** aligned to risk that are regularly updated would assist both clients and service providers. In addition, **promotion of consistent, well-publicised data standards**, such as those that are required to gain DEA data processor accreditation, is important for ensuring clients are well informed about the security protocols of data infrastructure services.

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?

With respect to government-held administrative data, **Government should invest in their own network of data services** (as it is doing, via investment such as ADR UK) and maintain these. We also suggest looking to internationally recognised standards and transparency in supply chain management for guidance on how these roles can be developed and implemented.

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.

Physical and technical security of data ('safe place') is only one of the **'Five Safes'**. Safe data, safe person, safe project and safe outputs should be considered of equal weight, because failures of information security are very often the result of misuse, human user error, and/or moving sensitive data out of a secure space and into an insecure space inappropriately. The problem with considering only one of the 'safes' is that the contribution of the others to overall risk mitigation is underappreciated. The result is excessive lockdown of data within an extreme setting for physical and technical security, which inhibits the use of the data.

The Digital Economy Act's Accredited Researcher status, along with Accredited Projects, shows that a diverse range of users can demonstrate compliance with a universal set of standards and conditions. An Accredited Researcher can be a civil servant in central government; a public sector worker in local government; a local director of public health; an academic; a charity worker; or a private sector employee. Similarly, a research project by any of these actors can be in the service of the public good. It is essential that these common conditions, which can be achieved by diverse actors, are recognised by data owners so that the data owner can benefit from the risk mitigations these standards provide. **Where these risk mitigations are in place, use of data can safely take place in suitable environments that can be outside of the locked-down servers of the data owner.** If these conditions and standards are not recognised and taken into account, access to data will only be to derived data products, or only for the civil servants of the department itself.

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.

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- **Strongly agree**

Yes, **the government should play a leading role in ensuring that data does not negatively contribute to carbon usage**, as it should for all policy areas. Climate change and sustainability are the most important longer-term issues faced by the world today, and all actions taken by government should account for this and avoid negatively contributing to carbon usage as far as possible. Data itself is a powerful tool for better understanding how government actions – including the collection, storage and use of data itself – can be done more sustainably. ADR UK has [several projects underway](#) making use of administrative data to better understand issues relating to climate and sustainability, with potential for much more to be done in this area.

Contact

Elizabeth Waind

Communications & Engagement Manager

ADR UK Strategic Hub

elizabeth.waind@ukri.org

www.adruk.org

[@ADR UK](#)

