

House of Commons Science and Technology Committee – Sharing data for public good: How can privacy be protected?

New inquiry: The right to privacy: Digital data

About ADR UK

[ADR UK](#) (Administrative Data Research UK) is a partnership of government and academic groups across all four UK nations (ADR England, ADR Northern Ireland, ADR Scotland and ADR Wales) and the Office for National Statistics (ONS). The partnership is coordinated by a UK-wide Strategic Hub within the Economic & Social Research Council (ESRC), which is part of UK Research and Innovation (UKRI).

ADR UK is creating linked research datasets from administrative sources, making these available to researchers through our network of trusted research environments (TREs). Administrative data is created when people interact with public services that keep records to carry out their day-to-day work. Although not originally created for research, this data has great potential to provide insights to help policymakers and others make better informed decisions. ONS is our major data infrastructure partner, with ADR UK directly funding the development and expansion of the ONS [Secure Research Service \(SRS\)](#). We also fund TREs in Northern Ireland, Wales and Scotland run by [NISRA](#), Swansea University (the [SAIL databank](#)) and [Research Data Scotland](#), respectively.

ADR UK started in 2018 as a pilot programme, where we tested different ways of working with data owners and researchers. As a result of the success of this pilot, in 2021 we had confirmation of £105m of long-term funding up to 2026. This decision was approved by ESRC, UKRI, BEIS and HMT.

Part of the success of the pilot was our collaboration with other UKRI investments that are also supporting opening up data for research, such as [Health Data Research UK \(HDR UK\)](#) and the [National Core Studies](#) programme. Through these collaborations, ADR UK has been able to open up access not just to administrative data for research, but also health data; for example, via the ONS SRS, the SAIL databank and Research Data Scotland. This is because ultimately, the design of a secure, robust TRE for administrative research works equally well for health data, which then opens up the potential for linking health and administrative data, as the ADR UK partners in Wales and Scotland have been doing for many years now.

As we move out of the critical phase into the recovery phase of the pandemic, ADR UK will be playing an increasingly pivotal role, as the need for insights based not just on health data, but education, benefits, income, homelessness, crime and justice data grows.

The ADR UK model for working with data owners and researchers is all about bringing government and academic groups together into collaborative partnerships. The aim is to deliver policy-relevant research that reinforces the feedback loop between those who have collaborated with us to open up access to data and the researchers commissioned to analyse it. Our model speeds up access to data by setting up partnerships between government data owners and external researchers to create the linked datasets, which are then hosted within

our TRE network and any external researcher can apply to access them for research that is in the public interest.

Since ADR UK began and the Digital Economy Act was fully enacted, both in 2018, the combination of having funded infrastructure and legislation in place means that research based on administrative data is now making a significant impact across a wide range of research sectors, increasing the bodies of knowledge in these areas, and informing policy practice. There has also been a steady rise in the number of research projects accredited by the UK Statistics Authority since 2018.

Year	Projects accredited
2019	23
2020	96
2021	127
Total	246

Fig 1. high-level metrics to illustrate the scale of acceleration in DEA-related approvals from 2018 to 2021.

Summary

Topic 1: Administrative data is an invaluable resource for public good. Sharing data benefits evidence-based policy making across the full range of policy areas including crime and justice, health and wellbeing, children and young people and housing and communities. ADR UK has already demonstrated the ability to safely and securely link and access government data across the UK to begin to provide these benefits. ADR UK has also commissioned and published a report from the Behavioural Insights Team (BIT) into the remaining barriers to data sharing within government. This identified two main types of barriers: resource limitations and judgements of risk, reward, and uncertainty.

Topic 2: There should be a more holistic approach to how health data is utilised for research in England, over and above the sharing of data for the purpose of supporting the health and care system and the operational use of data by the different parts of the health service. This would better facilitate wider public health research, for which you need to be able to link health data to other forms of data. This approach is already taken in other parts of the UK. Additionally, as we consider opening up more data for research, we should build on existing good practice such as the trusted research environments model which allows researchers to access data securely without the need for data to be released. ADR UK supports the vision for the health and social care system to have access to insights and analysis encompassing population-level information but also at the local-level. The Covid-19 pandemic has shone a light on the crucial role the research and analytical community has played in supporting local-level interventions. Coordinating with the devolved nations will also be critical to achieving a streamlined health data environment that serves the UK.

Topic 3: Arguably the biggest challenge for increased use of data is the need to reassure the public that their private information won't ever be made public. Across ADR UK, we do this by ensuring all our trusted research environments adhere to the "Five Safes", as evidenced by their accreditation under the Digital Economy Act (DEA, 2017). We also embed public engagement throughout our work, through community representative panels, consumer panels, public panels, steering committees, workshops, and by following our five public engagement principals.

Topic 4: Regarding appropriate safeguards, anonymisation may not always be appropriate (data may be less useful once anonymised), or possible (organisations need to have sufficient skills, resources, technology, and processes in place to undertake this work effectively). If de-identified information is accessed via a TRE that is accredited under a suitable process (for example under the DEA) and, as such, meets all the required security standards, then the de-identified data can be considered functionally anonymised, through a combination of the actions taken to create the de-identified dataset, and the environment in which the dataset is accessed. Finally, it is important that the public are engaged in data and privacy in a meaningful way. The public good that arises from using personal data for research should be made clear.

Topic 5: Existing governance arrangements in England for decision-making around the creation of linked administrative and health datasets to support research in the public good are not currently fit for purpose. Experienced researchers who understand all the processes can still take many years to work through them. Rather than NHS-D considering each application in isolation, ADR UK would like to collaborate with NHS-D and other related parties on the development of a framework to facilitate decisions being made around the linkage of health and administrative data. This would allow NHS-D to make decisions around such data linkage projects in ways that are already in place in both Wales and Scotland. The DEA has improved

the effectiveness of governance arrangements around the use of administrative data for research. The combination of having the funded infrastructure and legislation in place means that research based on administrative data is now making a significant contribution to the scientific evidence base across many research areas.

Topic 1: The potential benefits, including to research, to effectively use and share data between and across Government, other public bodies, research institutions and commercial organisations, and the existing barriers to such data sharing

Administrative data is an invaluable resource for public good. ADR UK is transforming the way researchers from across different sectors (academia, government, third sector organisations and research consultancies) access the UK's existing wealth of public sector data for the public good, so that policy can be based upon the best evidence possible. ADR UK's work spans a variety of UK and devolved government policy areas. The following five short case studies explain the benefits to research of being able to safely and securely link and access Government data (UK and devolved nations), where our work is fast progressing.

1. Crime and justice

One of ADR UK's flagship projects with huge potential for crime and justice policy impact is Data First: a major data linkage programme led by the Ministry of Justice (MoJ) which is linking existing data from across the justice system for the first time for research. The possibilities for improving justice system experiences – such as enabling people to access support quicker and more effectively and avoiding the escalation of civil justice problems into the criminal courts – are vast.

The £3 million ADR UK investment in Data First has already been making waves. It featured in the 2020 National Data Strategy and the UK Research and Development Roadmap, and a suite of linkable de-identified, research-ready datasets have already been deposited into the Office for National Statistics (ONS) Secure Research Service (SRS), so external researchers can apply to access them in an ethical and responsible way. The Data First project has been developed within the framework established under the Digital Economy Act (DEA) (2017), and further data deposits to the DEA-accredited SAIL databank (the ADR Wales trusted research environment) are planned.

We are also providing funding to support some of the [first users](#) of Data First data for research. Further funding opportunities related to policy-relevant crime and justice research will be announced as more datasets from Data First and other projects become available. This will include, for example, a [funding call](#) for applications to conduct research using a newly linked dataset consisting of the Ministry of Justice's (MoJ) Data First family court data and data from Cafcass (Children and Family Court Advisory and Support Service), via the SAIL databank. These Research Fellowships will enable researchers to improve our understanding of children and families' experiences of the family justice system. Examples of potential research include exploring the impact of long-running proceedings on children and families, the overlap of private and public law cases, and the characteristics of cases involving litigants in person (those in court without legal representation).

2. Children and young people

The past year has evidenced the importance of good education policy, and how damaging disruptions to education can be. Forming a more complete understanding of children and young people's lives, including how education provision affects later life, is vital for developing policy and services to properly support them and enable positive outcomes as adults.

ADR UK has several projects working to facilitate better evidence to inform education policy. One example is the Understanding Children's Lives and Outcomes programme in Scotland, which is led by ADR Scotland and involves linking pupil census data for Scotland with Scottish Government data on looked-after children, children's health, and the 2001/2011 national censuses. The linked data will give researchers a snapshot of the academic achievement, health, economic activity and wellbeing of children in Scotland, providing evidence for policymaking that aligns with the Scottish Government National Performance Framework.

Another example is from the ADR Scotland study on [growing up in kinship care](#). In Scotland, in the year 2019-20, 31% of children who could not live with their parents were living in formal kinship care, up from just 20% in 2010. As the number of young people living in formal kinship care increases, it is important that we better understand their experiences. The Growing Up in Kinship Care study aims to link administrative data from a range of sources to help us learn about the journeys, experiences, and outcomes of children and young people in kinship care.

3. Health and wellbeing

One of our pioneering projects with huge potential for both education and health policy impact is Education and Child Health Insights from Linked Data ([ECHILD](#)), which is linking children's health and education records for all children in England for the first time. This newly linked data will allow detailed analyses to be carried out on the trajectories of groups of children, making it possible to assess variation in provision and unmet need for children with diagnosed health conditions. Research using ECHILD will then help government and the providers of children's services to better understand their needs and to see which groups might be falling through the gaps. This will be particularly important in the context of understanding the impact of the Covid-19 pandemic on these children.

For both children and adults, maintaining good health and wellbeing is dependent on many factors, including the air we breathe. Ambient air pollution is the biggest environmental risk factor worldwide and a leading environmental cause of early death in the United Kingdom. Researchers from ADR Northern Ireland are exploring the [effect of air pollution on health and mortality](#), including relationships between pollution exposure and health conditions such as diabetes, dementia and Parkinson's disease. This will generate critical insights on the individual health costs of air pollution in Northern Ireland, and where changes to policy may be needed to mitigate the health risks.

4. Housing and communities

The provision of emergency accommodation for people sleeping rough as part of the government's Covid-19 response had the positive effect of reducing rough sleeping in 2020. However, this short-term crisis response may have less of an impact in the long-term, and good

quality data is essential for understanding how policy can work better to tackle rough sleeping, and homelessness more broadly, for good.

[Homelessness Data England](#) is a collaboration between ONS, the Department for Levelling Up, Housing & Communities (DHLUC) and ADR UK, aimed at creating a linked dataset about homelessness in England. The project involves linking together Homelessness Case Level Data Collection (H-CLIC) data – which records local authorities’ actions under the 2017 Homelessness Reduction Act (HRA) – from across England, and subsequently to other administrative datasets.

The HRA significantly reformed England’s homelessness legislation by placing duties on local authorities to intervene at earlier stages to prevent and reduce homelessness, and the H-CLIC data provides a unique opportunity to improve our understanding of the issue to in turn develop more effective policies to overcome it.

5. Climate & sustainability

Now that the UK has left the European Union, the government is tasked with rapidly developing new policies aimed at improving business prosperity, enhancing environmental stability and strengthening business and personal resilience outside of the EU – and agricultural policy is essential to this.

ADR UK is funding a project in partnership with government in all four UK nations to link de-identified data from across the country to create the first UK-wide data platform focused on agriculture. The [AD|ARC](#) (Administrative Data Agricultural Research Collection) project, led by researchers from ADR Wales, will enable researchers to uncover unique insights to better understand and support farmers, their households and communities. AD|ARC is also the first ADR UK project to truly involve all four nations of the UK, and the newly linked data will be hugely valuable for shedding light on ‘what works’ in agricultural policy that properly supports farming communities.

These five policy areas represent just a taste of where ADR UK is having an impact by enabling greater access to administrative data for research – and in turn putting the most comprehensive evidence possible at the fingertips of policymakers. Further details of many more data linkage research projects funded by ADR UK are available from the [ADR UK website](#).

Barriers to data sharing to support data linkage and access for research

In 2020, ADR UK commissioned the Behavioural Insights Team (BIT) to produce a report on applying behavioural insights to the challenge of linking and sharing government administrative data for research. The [published report](#) identifies remaining barriers in this area within government and suggests how behavioural insights can be applied to overcome them.

Using a combination of stakeholder interviews, desk research and a review of behavioural science literature, the report identifies two main types of barrier to data linking within government:

1. Resource limitations, including IT capacity, project management and staff time: participating in data linking projects is resource-intensive in terms of IT capacity, project

management and staff time. While these challenges are in part structural, there are also behavioural factors that can hamper effective project management.

2. Judgements of risk, reward and uncertainty, including misperceptions of the legal framework: organisations, their senior leadership teams and legal advisors need to make a series of judgements on whether and how to proceed with projects based on the risks and benefits they perceive. These teams must navigate legal risks, public perceptions and uncertainty around how the data will be used by external stakeholders (such as researchers).

Informed by behavioural science, the report offers a range of practical solutions to these issues, including the provision of open-source software solutions, synthetic data prototypes, advice on legal gateways, and the embedding of ADR UK-funded data engineering experts within government departments.

Activity underway to overcome barriers

ADR UK is working with its partners in government and the wider research community to see how the report's recommendations can be taken forward.

The use of synthetic data

In collaboration with BIT, ADR UK recently published a [report](#) on how public policy research could be accelerated with the use of synthetic datasets, by increasing the uptake of administrative data research while maintaining data security. The principal recommendation of the report is that ADR UK should encourage the use and sharing of low-fidelity synthetic data across government and with researchers. This will raise awareness about the different sources of administrative data now accessible and help researchers to develop their applications to access it.

DARE UK

To support longer-term decision-making, ADR UK is also working with HDR UK to lead the first phase of the UKRI-funded [DARE UK \(Data and Analytics Research Environments UK\)](#) programme. This programme which aims to design and deliver a national data research infrastructure that is joined-up, demonstrates trustworthiness and supports research at scale for public good. The programme was initiated to overcome many of the existing barriers to data sharing and use, and to develop a more coordinated and trustworthy data research infrastructure for the UK.

Topic 2: The extent to which data issues are appropriately addressed by the Government's National Data Strategy, its draft strategy, Data saves lives: reshaping health and social care with data, and its consultation, Data: a new direction

Consider the wider public health perspective when considering the use of health data for research

The wider public health perspective would benefit from a much more holistic approach to the consideration of how health data for England is utilised for research, over and above (to quote): *'enabling the proportionate sharing of data for the purpose of supporting the health and care system'* and the operational use of data by the different parts of the health service – e.g. *'putting patients at the heart of their health and care data, with easy access to their own healthcare records'* and *'giving health and care staff easier access to the right information to provide the best possible care through shared records and simplified information governance.'*

A more holistic approach would better facilitate wider public health research, for which you need to be able to link health data to other forms of data; for example, to bring together health and environmental data, which are critically important, particularly for vulnerable communities who are more impacted by poor environmental conditions. We have included some examples below for illustration. This more holistic approach as to how health data is utilised for research is already routinely taken in other parts of the UK:

ECHILD database: An example of how health data can be linked to education data to provide public benefits is the ECHILD database which links Hospital Episode Statistics to the National Pupil Database. This linked administrative data will allow accredited researchers to gain a detailed picture of chronic health conditions and other vulnerabilities faced by children in England and how these affect their educational outcomes. As well as providing insights into a multitude of issues surrounding children with special needs, it will also lead policy makers and health services to better understand the needs, and how to meet those needs, of vulnerable children during crises such as the Covid-19 pandemic.

Although this linked dataset has been securely accessed to support research by the research team who were originally funded to create the linked dataset, it has not yet been possible to open this access up for wider research use, although all parties (data owners, funded researchers and ADR UK) are committed to agreeing what the most appropriate governance model is to allow this to happen.

Shielding Patient List: Early in the pandemic, there was an urgent need within Welsh Government to understand the circumstances of those on the Shielding Patient List, given they were advised to severely restrict their activities to protect their health. ADR Wales could respond to this challenge immediately through linking health data to administrative datasets. For example, ADR Wales researchers estimated levels of access to outdoor space, given that in the early days of shielding they were advised not to leave their property, even for exercise.

ADR Wales researchers were also able to provide critical intelligence to education officials in local and central government on children living in shielding households and teachers who were shielding to help inform planning around reopening of schools in summer 2020. Further analysis by ADR Wales researchers using Covid-19 testing data linked to pupil and teacher level data found that the opening of schools in Wales between September and December 2020 was not associated with an increased subsequent risk of staff testing positive, although pupils were found to be at increased risk of testing positive, following cases appearing within their own year group. At a later stage in the pandemic, the focus turned to the vaccination programme and again in the context of reopening of schools, this time in autumn 2021, there was a need to understand vaccination rates of teachers. Once again, ADR Wales researchers were able to link health and administrative data to address this need.

Build on existing good practice when considering how to open up data to support wider research

We strongly recommend that efforts around the development and use of trusted research environments (TREs) build upon existing and evolving best practice and investments, rather than creating entirely new approaches. As stated in HM Government's Life Sciences Vision, TREs should be interoperable and highly secure, although it should be acknowledged that multiple TREs will be required across the landscape to reflect the diversity of data types, uses, and research and innovation needs. There should also be acknowledgement that effective management and curation of data for research typically follows areas of devolved responsibility across the UK. Centralising such data access could inadvertently result in less data being made accessible for research, rather than more.

As the scale and complexity of health and other data to be analysed by researchers increases, the TRE model allows researchers to access data in a secure way, without the requirement for data release. It is also worth noting the value to researchers of accessing de-identified health and administrative data, as exemplified via ESRC investments such as the UK Data Service, in addition to the TREs that are part of the ADR UK programme. Increasingly, this includes access to pre-linked, de-identified health and administrative datasets for research.

The UK Data Service and ADR UK TRE network have all successfully been accredited by the UK Statistics Authority under the Digital Economy Act (DEA, 2017). The DEA TRE accreditation model, where data is accessed remotely via secure systems, also mitigates the residual risk of re-identification of de-identified data via fuzzy matching to other datasets, since the controls in place within these environments means this is technically not possible unless previously approved by the data owners.

DARE UK

The DARE UK programme – which ADR UK is co-leading with HDR UK for the first phase - is focused on building a more coordinated national TRE ecosystem. The scope of DARE UK includes all research conducted by UKRI research councils that uses, or anticipates use of, sensitive data from different research disciplines and from across different sectors. DARE UK aims to develop aligned and streamlined approaches to build an ecosystem of open, secure, interoperable, federated TREs to deliver shared policies, standards, tools, and best practice from across health and non-health domains.

Fundamental to how all data about people are utilised for research is meaningful engagement with the public to explain upfront how their data will be used – not just how they access their own information. Again, there are tried and tested way of doing this already in place across the UK (see below for examples).

Facilitate insights and analysis at population but also local level

ADR UK supports the vision for the health and social care system to have access to insights and analysis encompassing population-level information but also at the local-level. As highlighted in the draft strategy, the Covid-19 pandemic has shone a light on the crucial role the research and analytical community has played in supporting local-level interventions. For example, in England, researchers from the Local Data Spaces programme, an ADR UK-funded collaboration between the Joint Biosecurity Centre, Office for National Statistics and the Department for Levelling up, Housing and Communities have identified two consistent core research priorities for local authorities: broader Covid-19 health impacts and inequalities; and economic vulnerability and recovery potential. This work has resulted in free Covid-19

analysis reports available to download for all local authorities in England. More generally, ADR UK's linked datasets, created from multiple sources across UK public services, have played a vital role in informing how research and innovation across sectors address the societal impacts of this pandemic.

Foster greater co-ordination and co-operation with the devolved nations

We note and support the strategy's commitment to work across central government and across the devolved administrations to improve appropriate data linkage to support people's health and wellbeing. As referenced above, while we acknowledge that England is the focus of the strategy, co-ordination with devolved nations for achieving a streamlined health data environment would benefit the whole UK in the long-term. Each devolved nation will have its own strategy, but the scale of the issues faced warrants some mention of how co-ordination and co-operation may be aspired to and achieved.

Topic 3: The ethics underpinning the use and sharing of individuals' data in health and care contexts

Ethical use of and access to data

Arguably the biggest challenge for increased use of data is the need to reassure the public that their private information won't ever be made public. Across ADR UK, we do this by ensuring all our trusted research environments adhere to the "Five Safes", as evidenced by their accreditation under the DEA (2017):

1. **Safe people:** researchers must demonstrate that they have the technical skills to use the data, either through academic qualifications or practical research experience.
2. **Safe projects:** researchers need to demonstrate that their proposal is an appropriate and ethical use of the data, that it will deliver clear public benefits and that they will publish their results to enable use, scrutiny and further research. These applications are reviewed by a panel of experts and, where appropriate, by an ethics committee.
3. **Safe settings:** researchers analyse data held in the trusted research environment remotely.
4. **Safe outputs:** once researchers have completed their analysis, outputs are checked before being released, to ensure data subjects cannot be identified from them.
5. **Safe data:** Before researchers access the data, it is de-identified by removing names, addresses and any other details that would directly identify the data subjects.

Understanding public attitudes to the use of administrative data for research

There are a number of 'unanswered questions' in relation to the UK public's attitudes towards the use of administrative data for research, which are in turn linked to the ethics underpinning the use and sharing of individuals' data, both in the context of health and administrative data. This is demonstrated in [ADR UK's review of previous literature](#) on the topic. Namely, these include:

- What do the public understand by the terms 'de-identified' and 'anonymous', and the differences between the two?
- What do the public understand by 'public interest', 'public benefit', or 'public good'? Does this differ among different demographic groups and communities?
- What are the public's views towards remote access to a secure research environment?
- What are the public's views towards a 'central store' of anonymised public sector data which can be accessed on an ongoing basis by approved researchers working on projects in the public interest?

The ADR UK Public Engagement Steering Group (PESG) will plan and implement public dialogue on these topics on a UK-wide level, via workshops, focus groups or other methods. We hope this will also be in collaboration with other relevant external organisations, such as the Office for Statistics Regulation (OSR). Another potential topic for wider work is exploration of how public views towards the use of data for research have changed as a result of the Covid-19 pandemic and its acceleration of the use of data for crucial insights to inform the pandemic response and highlight its impacts.

Meaningful public engagement

The ADR UK approach to embedding meaningful public engagement into everything we do is described in our [Public Engagement Strategy](#).

We define public engagement as a purposeful set of activities designed to promote an ongoing dialogue with the public about administrative data research, driven by active listening and responding. The purpose of these activities is twofold. Firstly, they help us demonstrate trustworthiness by listening and responding to the public's views about how administrative and health data should be used for research. This means that we aim to embed public engagement in every stage of our work, and to meet public expectations over and beyond the standard legal requirements for the secure use and sharing of data for research in the public interest. Secondly, public engagement activities help us maximise the public benefit of administrative data research; understanding the public's needs and interests means we can achieve benefits in the right places.

DARE UK public dialogue

In addition, key to the DARE UK programme is the development of a trustworthy data research infrastructure that maintains the confidence of the public. In this first phase of DARE UK, which runs from July 2021 to August 2022, a UK-wide public dialogue is being undertaken to explore public interests and concerns regarding how sensitive data should be managed and made accessible for research across the UK. A final public dialogue report is due to be published in Spring 2022.

Topic 4: The extent to which appropriate safeguards and privacy are applied in the usage and sharing of individuals' data

The role of privacy enhancing technologies

These play an important role in enabling responsible research and maintaining public trust. ADR UK only facilitates access to data through our DEA-accredited network of trusted research environments. These provide a secure environment where researchers can access and analyse data in the TRE environment only and are unable to download the data, with all research outputs being subject to an approval process and statistical disclosure controls to ensure that they do not contain personal data. TREs therefore aim to maximise data security and protect individuals' privacy whilst enabling researchers to access rich, linked data assets, and are an important tool to demonstrate trustworthy use of data. The ADR UK TRE network also enables remote research, including home working. This not only facilitates researchers working across institutional and geographical boundaries, but also promotes equality of access – wherever researchers are based across the UK.

Data anonymisation vs. de-identification

True anonymisation of data – so, not just the removal of personal identifiers, but the consideration of whether individuals can be indirectly identifiable when data are linked – can be technically challenging and complex. It also may not always be appropriate or possible, as it depends on knowing what other information could feasibly be accessed that would allow re-identification.

While there may be incentives for organisations to only allow access to data in an anonymised form (for example, the data will no longer fall within the scope of the GDPR), this may devalue the data to such an extent that it is no longer useful for research purposes. So, before anonymisation is contemplated, the utility of the resulting data for research purposes needs to be considered. Also, organisations need to have sufficient skills, resources, technology, and processes in place to undertake this work effectively.

Data de-identification, which refers to the likelihood of data subjects being re-identified directly from that data, is a much more useful concept when considering the research use of data. If de-identified information is accessed via a trusted research environment that is accredited under a suitable process, for example under the DEA, and as such meets all the required security standards, then the de-identified data can be considered functionally anonymised, through a combination of the actions taken to create the de-identified dataset, and the environment in which the dataset is accessed. As an example, the design of a DEA-accredited trusted research environment means researchers are not allowed to bring in other sources of data, then attempt to link these with the de-identified data being accessed within the trusted research environment, to re-identify data subjects. Therefore, the residual risk related to the creation of a de-identified dataset rather than an anonymised dataset is mitigated and the data retains its value as a research asset.

Involving the public in decisions around the use and sharing of data

It is important that the public are engaged in decisions around how data is used for research, and how it is shared, accessed and linked. The overall approach of the UK National Data Strategy would be strengthened by wider public engagement, participation, and communication. For example:

- The public good arising from research using personal data and data linkage should be made more transparent. The legal underpinning of data sharing for research or statistics is already relatively clear; it is the social contract with the public that needs to be earned and maintained, irrespective of what we can do legally.
- The National Data Strategy should explicitly cover functional anonymisation. This relates to data being stored and used for research and statistics according to the '[Five Safes](#)' (e.g., de-identified and in trusted research environments by qualified researchers) is accessed without explicit individual-level consent.

While we welcome efforts to provide clarity and reduce uncertainty in this complex area, the data protection regime needs to work for people as well as for business, economic growth, and innovation. This was highlighted in the Information Commissioner's Office response, to the UK National Data Strategy, which we endorse.

Synthetic data – an application of a privacy enhancing technology

We are aware that it is possible to create synthetic data that retains properties of real data, but which does not need to be hosted within a trusted research environment, so reducing the risks around attacks to access the real data. As such, synthetic data creation is considered to be one of many privacy enhancing technologies. We are aware it is technically possible to create a fully synthetic 'high fidelity' dataset that retains all of the statistical relationships of the original, but that theoretically can be released without restrictions, because it isn't 'real' data. However, from discussions ADR UK have had with data owners, we are not convinced this is a feasible solution to removing remaining barriers for research access to data. In particular, data owners are unlikely to agree to the construction and release of these datasets, because:

- All control of the dataset is lost
- All control of messaging related to research outputs based on the synthetic data would be out of their control
- The public wouldn't necessarily understand the difference between messaging based on a synthetic dataset vs. the real dataset

From ADR UK's perspective as a funder of both data linkage and research programmes, our concerns are also around the following issues:

- If these datasets were routinely created and released, data owners would feel they never needed to agree to allow access to the real data
- The research value be of analysis conducted on synthetic data, if it couldn't be easily validated against the real data

However, ADR UK is supportive of the construction of *low-fidelity* synthetic datasets to support training activities, raise awareness about the different sources of administrative data now accessible, and help researchers to develop their applications to access it. Our position is informed by the [report](#) we commissioned from the Behavioural Insights Team (BIT) on the uses of synthetic data in government. The report also discusses the different types, benefits, and challenges of creating and using synthetic data, making recommendations for ADR UK, government and stakeholders outside of government.

Topic 5: The effectiveness of existing governance arrangements, e.g., the Centre for Data Ethics and Innovation.

Governance arrangements for the use of health and administrative data

As ADR UK is a UK-wide programme, we have deep experience of working through the governance arrangements to support decisions around the use of administrative and health data for research purposes in England, Wales, Scotland and Northern Ireland. More information on how these work in practice for each of the devolved administrations of the UK can be provided on request. However, since the current inquiry is mainly focussed on health data, this submission focusses on our experience of the governance arrangements in place for England. Our view is that existing governance arrangements for decision making around the creation of linked administrative or health datasets to support research in the public good in England are not currently fit for purpose, despite a lot of good will across all parties. We illustrate this with a couple of case studies:

ECHILD study: ADR UK has been funding NHS-D through the [ECHILD programme](#) to create a linked research database of education and social care records and hospital contacts for all children in England born from 1995 until the present (about 15 million). The initial use is for an academic team led by Prof. Ruth Gilbert at UCL, in collaboration with teams at the London School of Hygiene and the Institute for Fiscal Studies. The intention is for this data to be made accessible to approved UK researchers via the ONS Secure Research Service. Work is ongoing to establish the permissions to open up access to the ECHILD database, but a legal basis has not yet been established. All parties (data owners, funded researchers, and ADR UK) are committed to establishing an appropriate governance model and are working hard to make this happen. To put this timeline into context, the initial applications for approval of the creation of the linked ECHILD database were started by Prof. Gilbert in 2014, five years before the ADR UK-funded project.

PICNIC study: The [PICNIC study](#) is led by Dr Pia Hardelid from UCL, and uses a dataset held in the ONS Secure Research Service linking ONS birth registration information to birth notification data, and HES for mothers and children from NHS-D. Pia's study expands this linkage further by linking in environmental and housing data; for example, 2011 Census data and air pollution and Energy Performance Certificate data. It has taken Pia over three years to get the approvals for these linkages in England. However, there is currently no clear governance process in place to facilitate this happening, without further researchers going back to the start of the application process that it has taken Pia over three years to navigate.

There are many health and administrative data linkages that would be strategically important to create, to support the management of the next phase of the Covid-19 pandemic and beyond. Rather than NHS-D considering each case in isolation (which we know takes many years), ADR UK would like to collaborate with NHS-D and other related parties on the development of a framework to facilitate decisions being made around the linkage of health and administrative data. If agreed, this would allow NHS-D to make decisions around such data linkage projects in ways that are already in place in both Wales and Scotland.

In contrast, the effectiveness of governance arrangements around the use of administrative data for research across the UK has been improved with the enactment of the Digital Economy Act (2017). This includes the governance processes that are the responsibility of the UK Statistics Authority to manage, which include ethics approvals along with people, project and data processor (TRE) accreditations, with support from ONS.