

National Data Strategy Open Call for Evidence

Response from Administrative Data Research UK (ADR UK) – July 2019

About ADR UK

[Administrative Data Research UK \(ADR UK\)](#) is a £44 million UK Research and Innovation (UKRI) investment drawn from the National Productivity Investment Fund (NPIF) via the [Economic and Social Research Council \(ESRC\)](#). It runs from July 2018 to July 2021.

Its objective is to transform the way administrative data are used for research, enabling valuable new insights into UK society by linking together data held by different public authorities. Research using administrative data can better inform policy decisions to improve people's lives. ADR UK facilitates safe and secure access to linked administrative data for accredited researchers whose findings give decision-makers the insights they need to address important policy questions.

ADR UK is made up of three national partnerships (ADR Scotland, ADR Wales, and ADR NI) and the Office for National Statistics (ONS), which ensures data provided by UK government bodies is accessed by researchers in a safe and secure form with minimal risk to data holders or the public. The partnership nature of ADR UK means that each country of the UK has a dedicated centre providing high-quality services for administrative data research.

The partnerships are coordinated by a UK-wide Strategic Hub, which also promotes the benefits of administrative data research to the public and the wider research community; engages with UK government to secure access to data for research; and manages a dedicated fund to support emerging priorities to create novel research-ready administrative datasets and associated research for public good.

To find out more, visit adruk.org (website launching end of July 2019) or <https://esrc.ukri.org/research/our-research/administrative-data-research-uk/>.

Key points

ADR UK appreciates the need for and welcomes the creation of a National Data Strategy. There are three core points that run through our response and which we would like to emphasise here. Namely:

- 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.
- 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.
- 3) There is a need for clearer structures for enforcing the **ethical and trustworthy** use of public data, with a focus upon **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.

People

Objective 1: To ensure that data are used in a way that people can trust

ADR UK is committed to the 2017 Digital Economy Act, and to maintaining the utmost levels of safety and security in its storage, use and sharing of public data for research purposes. We are also committed to using public data in a transparent way, to keeping people well-informed about how their data are used for research purposes, and about the findings of that research.

Research area: opportunities and barriers to trust

1.1. How can organisations (private, public or third sector) demonstrate trustworthiness in their use of data?

The ‘Five Safes’

The ‘Five Safes’ framework is central to ADR UK’s infrastructure design: safe people; safe projects; safe settings; safe outputs; and safe data.¹ Following this framework means that data are only used by trained and accredited professionals for research projects that deliver clear public benefits, in secure settings and with all outputs from data analysis carefully checked and confirmed as non-disclosive. Observing the Five Safes maximises the research potential of linked administrative data whilst maintaining the confidentiality of data subjects and ensuring that the data is only used for purposes in the public interest.

Public engagement

The report ‘[Trust, Transparency and Technology: Building Data Policies for the Public Good](#)’ by Policy Connect and the APPG on Data Analytics highlighted the importance of public engagement in the use of public sector data. The report advises that “rules made with little or no public engagement have led to avoidable errors which could contribute to public distrust in data use”, and says that the public should be engaged via a variety of methods including “open consultations, town-hall meetings, industry outreach, and other ways of directly engaging with members of the public and relevant stakeholders”.² ADR UK supports this recommendation, and strives to develop and maintain high quality public engagement as a central aspect of our work.

Part 5 of the 2017 [Digital Economy Act \(DEA\)](#) provides the legal basis for the disclosure of public sector data for research. ADR UK is therefore not reliant on consent as a basis for processing personal data, but nevertheless we have an ethical imperative to keep the public informed about the ways in which their data are being used.

ADR Wales has used an 18-member consumer panel to advise on the operation of their administrative data research centre. They found their panel to be a valuable addition to their work in health-related data linkage research, with the views of members providing a “positive outlook and a fresh, and sometimes unexpected, perspective on various issues.”³ Their reflection suggests that the incorporation of a consumer panel, as one method of public engagement, is valuable to work with data linkage research, as well as to researchers in general and consumers themselves. Members of

¹ Stokes, P. (2017), ‘The ‘Five Safes’ – Data Privacy at ONS’, *Office for National Statistics*, accessed 12/07/2019: <https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/>.

² Tindale, Jack, and Muirhead, Oona (2019), *Trust, Transparency and Tech: Building Ethical Data Policies for the Public Good*, Policy Connect, London, p.15. Accessible via: <https://www.policyconnect.org.uk/appgda/research/trust-transparency-and-technology-building-data-policies-public-good>.

³ Jones, Kerina H., McNerney, Cynthia L., and Ford, David V. (2014), ‘Involving consumers in the work of a data linkage research unit’, *International Journal of Consumer Studies*, 38, pp.45-51., p.45., accessible via: <https://cronfa.swan.ac.uk/Record/cronfa16840>.

ADR Wales' consumer panel also sit on the independent Information Governance Review Panel, which assesses all research uses of ADR Wales' data to ensure that the public voice is represented.

Publications and citations

Citations should include reference to the legal, ethical, and 'Five Safes' basis within which the research was carried out. Rather like the 'Red Tractor' food kite mark, this would allow peer reviewers, and users of the insights, to be confident about the trustworthiness of the use of the data. It would also put pressure on those publications that are opaque about the propriety of their data.

The Digital Economy Act and the standards and criteria it introduces should reassure the public about the use of their data, and where research has been produced within its framework this should be included in publications and citations. For this, and many other reasons, we suggest the DEA framework should be opted into, even if its legislative powers are not strictly needed. The DEA is a full response to the challenge of the GDPR and the 2018 Data Protection Act to ensure that when processing data under the *Public Task* condition, the legal basis for that processing is clear and comprehensive.

1.2. How easy is it for the public to find out about how information provided to or inferred about them by an organisation is being used?

The Policy Connect and APPG [report](#) (see 1.1 above) emphasises that public engagement is key to maintaining public trust. The report refers to the Department for Health's care.data programme, which was abandoned after "the means used by NHS England to communicate with the public and give them the ability to opt out was seriously flawed, which when exposed led to more than a million opt-outs, negating the value of the programme".⁴ Keeping the public informed about the ways in which their data are being used, as well as about the findings of research using their data, is therefore essential for maintaining public trust and the integrity of investments using public sector data, as well as being an ethical imperative.

It can be difficult and expensive for the public to benefit from the research that used their data because publication of results is too often behind a paywall. Publication of research in open access journals should be promoted, and Digital Object Identifiers used to locate the datasets used in the production of an article. Citations should also reference the deployment of the 'Five Safes'. ADR UK supports the establishment of new open access journals such as [Data & Policy](#), founded in June 2019 by the organisers of the annual [Data for Policy](#) conference and Cambridge University Press.

ADR UK strives to ensure the findings of research it enables are published and communicated in a form that is easily digestible and useable by government policymakers and wider society, as well as by academics. Often this will mean a 'dual publication' approach, for instance in an easily-accessible policy or research paper on a government website as well as in an academic journal, therefore improving public access to the findings of research using public data.

1.3. Are organisations (private, public or third sector) using personal data in ways that may damage trust?

There have been instances in which the use of personal data for operational purposes has led to trust being damaged. For example, a controversial Memorandum of Understanding (MoU) between the NHS and the Home Office was suspended in 2018 after concerns were raised by the Health and Social Care Committee. The MoU allowed the NHS to share patient data with the Home Office for immigration tracking purposes. The Select Committee said that it was not satisfied that the chair and

⁴ Tindale, Jack, and Muirhead, Oona (2019), *Trust, Transparency and Tech: Building Ethical Data Policies for the Public Good*, Policy Connect, London, p.28. Accessible via: <https://www.policyconnect.org.uk/appgda/research/trust-transparency-and-technology-building-data-policies-public-good>.

chief executive of NHS Digital had been “sufficiently robust in upholding the interests of patients, understanding the ethical principles underpinning confidentiality, or in maintaining the necessary degree of independence from Government”.⁵

An example of a commercial misuse of government-held personal data is the 2017 Google DeepMind trial, in which the Royal Free NHS Foundation Trust was deemed by the Information Commissioner’s Office (ICO) to have failed to comply with the Data Protection Act when it provided the personal patient details of around 1.6 million patients to Google DeepMind. The data share occurred as part of a trial to test an ‘alert, diagnosis and detection system for acute kidney injury’.⁶

When administrative data are used for research within the DEA framework and according to the Five Safes principles, all personal identifiers are removed before being shared with researchers. Data used for research rather than operational purposes therefore sees clear differences in terms of protecting the confidentiality of data subjects. De-identified data shared for research holds far less risk to the confidentiality of data subjects than still identifiable data shared for operational purposes.

In August 2017, the research team for the [Northern Ireland Cohort for the Longitudinal study of Ageing \(NICOLA\)](#), in partnership with the engagement team from the Administrative Data Research Centre Northern Ireland (ADRC NI), part of ADRC UK, convened a public panel drawn from participants of the NICOLA. The aim was to discuss with participants how data from the study – survey data – could potentially be linked with administrative data, enhancing what could be learned about the process and experience of ageing in Northern Ireland. Researchers and engagement specialists took panel participants through how their personal data were collected and used within the study as well as some initial findings, and then explored data linkage and what kinds of information and knowledge could be developed by linking datasets together. Participants were surprised to learn that such linking was not already used to improve research quality. Participants were perturbed about their data *not* being used in a more joined-up way. Although this is a biased sample, it adds to the evidence that the non-use of personal data for research is a harm that often outweighs concerns about how data are used.

Research area: concerns around trustworthiness

1.5. Do people know how information provided to, or inferred about them by, an organisation (private, public or third sector) is being used, stored and shared?

In September 2018 the Information Commissioner’s Office (ICO) [annual trust and confidence survey](#) found that 18% of respondents had a ‘good understanding’ of how their personal data are used by companies and organisations in the UK, compared to 10% in 2017. The majority of the 2018 respondents (55%) said they were ‘familiar with some aspects of how my personal data is used’, with 4% saying ‘I know nothing at all about how my personal data is used’. Furthermore, 23% of respondents said they agree with the statements: ‘It is easy to find out how my personal information is stored and used by businesses/organisations’; and ‘It is easy to find out whether my personal information is being made available to third parties’, compared to 21% and 20% respectively in 2017.⁷

These findings suggest that, although public understanding of and access to information about the ways in which their data are used, stored and shared rose slightly from 2017 to 2018, the large majority of the public still know little of how their personal information is used, stored and shared.

The Digital Economy Act requires that all data processors, all research projects, and all researchers that are accredited by the UK Statistics Authority must be published. This, when combined with

⁵ McKee, Selina (2018), ‘Gov’t suspends controversial NHS data sharing deal’, *PharmaTimes*, 10th May, <http://www.pharmatimes.com/news/govt-suspends-controversial-nhs-data-sharing-deal-1235020>.

⁶ Information Commissioner’s Office, ‘Royal Free - Google DeepMind trial failed to comply with data protection law’, *Information Commissioner’s Office*, accessed 12/07/2019: <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law/>

⁷ Information Commissioner’s Office (2018), *Information Rights Strategic Plan: Trust and Confidence*, <https://ico.org.uk/media/about-the-ico/documents/2259732/annual-track-2018.pdf>.

improved citations in open access to journals, will improve the availability of information about the use of personal data for research.

1.6. To what extent are people concerned about how data about them is used, stored and shared? Are some groups more concerned than others? Are there particular categories of data that raise more concerns than others?

The ICO's 2018 [annual trust and confidence survey](#) (see above) also found that one in three (34%) have high trust and confidence in companies and organisations storing and using their personal information. Respondents were also found to be significantly more likely to have high trust and confidence in the NHS, the police and national governmental departments and organisations than in private companies, including social media platforms and mobile, broadband and utility providers. 65% were found to have high trust and confidence in the NHS or their local GP, compared to only 15% for social messaging platforms.

Dialogue on Data 2014

In 2014 the Administrative Data Research Network (ADRN) – the predecessor to ADR UK – worked with Ipsos Mori to explore public views on using administrative data for research to produce the '[Dialogue on Data](#)'. The research, which involved 14 workshops with members of the public and sector experts, found that participants were familiar with the importance of data to modern society, but protecting their personal data was nevertheless very important to them. Participants expressed concern about their data being leaked, lost, shared or sold by organisations, feeling like they had little control over their personal data.⁸

Participants were also asked about the ADRN itself, with the most common concerns about the initiative being the possibility of re-identification of de-identified data; security of linked datasets; and whether the initiative represented a good use of public money. However, they did find the independence of the ADRN reassuring, as well as the security provisions in place and the potential for linked data to improve public service provision. There was, however, reassurance that there were no plans for a 'super' database under ADRN containing multiple linked data sources⁹.

ADR UK inherits these findings and will implement efficient and sustainable models for data linking where only the sub-sets of these de-identified datasets – the variables needed by the researcher to answer their research question – are made available for use by accredited researchers within the secure data centres. ADR UK also plans to conduct its own public engagement work to gauge public perceptions relating to this new model for more efficient use of linked administrative data for research.

[ARK Northern Ireland](#) to conduct research on [Public Attitudes to Data-sharing](#) in Northern Ireland as part of the [Northern Ireland Life and Times \(NILT\) Survey 2015](#). The survey included interviews with 1,202 respondents sampled via a systematic random sample of addresses taken from the Postcode Address File.

The survey had similar findings to the ICO, namely that high proportions of respondents trust their GP surgery (91%) and NHS (86%) to keep information or data secure and use it appropriately. Slightly smaller proportions trust government departments (73%) and academic researchers (72%) and trust in charities (51%) and commercial organisations (41%) is much lower. Overall, 42% said they had ever had particular concerns about how any of those organisations used the information they kept.¹⁰

The majority of respondents supported the concept of sharing of identified health data to improve services. Over 95% were in favour of sharing data within the health service by means of the electronic care record, and over two thirds of people were in favour of health information being shared to

⁸ Cameron, Daniel, et al. (2014), *Dialogue on Data: Exploring the public's views on using administrative data for research purposes*, Ipsos Mori: <https://www.ipsos.com/ipsos-mori/en-uk/dialogue-data>, p.3.

⁹ Ibid., pp.3-4.

¹⁰ Robinson, G., and Dolk, H. (2018), 'Public Attitudes to Data Sharing in Northern Ireland', *Access Research Knowledge Research Update*, 108, <https://www.ark.ac.uk/publications/updates/update108.pdf>.

improve access to services provided by other government departments, with information about mental health or children seen as more sensitive.

Regarding sharing data for research, 85% of people agreed that “if personal data can be made anonymous and a person’s right to privacy maintained, then the data should be used where there is a benefit to society”. Nearly one third of respondents insisted that “data should only be shared for research if there is explicit consent, even if this means you will have to abandon the research if there are difficulties contacting people”. It is worth noting that this attitude toward needing consent, held by a minority of respondents, is also associated with low trust in organisations including the NHS and government departments.¹¹

There was a high level of support for data protection measures, particularly de-identification, checks and penalties for researchers, and making sure research is of public benefit. Other measures also have high support, including data access in dedicated secure data centres, and making results public. The lower level of trust in commercial organisations to keep data secure and use it appropriately was reflected in lower support for data sharing with commercial organisations, and 50% believed the data protection safeguards implemented for academic researchers should be greater for commercial organisations. However, there was evidence that where there is great public benefit, there is more support for commercial access to data.¹²

Thus, public support for data sharing was found to sit on three pillars – trust in organisations, data protection measures, and public benefit. If any of these are reduced or taken away, public support falls, and conversely if energy is put into one of these domains but not in others, then it will not be enough to secure public support.

In addition to the above, recent work into public perceptions regarding the use of genomic data for research (as opposed to clinical care data) by researchers at Swansea University, a partner of ADR UK, has found that participants placed high value on genomic data for research. Their main concerns included the potential misuse of data; security; discrimination; and the purpose for which the data was being used.¹³

1.7. What commercial practices or behaviours have affected trust in the use of personal data? Have targeted advertising and ‘recommending’ affected trust?

In addition to the above, the ICO’s 2018 [survey](#) found that, in instances in which personal information is shared with third parties, or if an organisation is affected by a data breach, the impact upon trust and confidence in the organisation is far more negatively impacted when data is shared without prior permission or when the organisation does not inform the data subject about a breach. 23% of respondents said there would be a negative impact upon trust if an organisation shared personal information to third parties with the individual’s permission, compared to 85% when permission was not granted.

Furthermore, respondents were strongly in favour of a ban upon companies and organisations sharing personal information with third parties without permission, with 35% saying this was the main way to increase their trust and confidence in how their personal data are used. This strongly suggests that transparency has a huge impact upon the public’s trust in the way in which personal data are used and shared. Nevertheless, the response applies specifically to ‘companies and organisations’, and does not mention government; it is worth noting that sentiments towards data sharing across government may differ.

¹¹ Robinson, G., and Dolk, H. (2018), ‘Public Attitudes to Data Sharing in Northern Ireland’, *Access Research Knowledge Research Update*, 108, <https://www.ark.ac.uk/publications/updates/update108.pdf>.

¹² Ibid.

¹³ Jones K.H., Daniels H, Squires E.L. and Ford, D.V. (2019), ‘Public views on models for accessing genomic and health data for research: a mixed methods study’, *Journal of Medical Internet Research* (in press). doi:10.2196/14384; <http://dx.doi.org/10.2196/14384>.

The Open Data Institute's [2018 survey](#) investigating public attitudes towards data sharing in Britain found that consumers are prepared to make 'worthy trade-offs' of their personal data if it benefits themselves or others in society.¹⁴ 47% of respondents would share personal health data if it helped develop new medicines and treatments, whilst 37% (49% of 18-to-24-year-olds) would share personal data to help advance understanding of areas such as medicine and psychology. Age did make a difference, with 38% of 18-to-24-year-olds but only 15% of those over 55 saying they would share data about their spending habits if it would help save them money via things such as savings accounts, insurance policies and shopping discounts. Thus, amongst the younger generations, there is some support for targeted advertising and recommending by companies.

In general, it is fair to say that attitudes towards personal data security as a whole have shifted as a result of recent scandals such as Cambridge Analytica. Recent scandals with a connection to government, such as the previously mentioned NHS/Home Office and Google DeepMind data shares (see 1.3 above), are particularly relevant to ADR UK, as they are likely to have made the public more aware of the potential risks, both in terms of security and ethics, of the sharing of government-held personal data.

1.8. Have the General Data Protection Regulation (GDPR) and Data Protection Act 2018 made people more concerned about how personal data is managed? How has it influenced their behaviour?

As touched upon in the National Audit Office (NAO)'s recent report, '[Challenges in using data across government](#)', citizens' interest in and understanding of how their data are being used has been heightened by the GDPR.¹⁵ The authors stress that "well-publicised reports of misuse of data, such as those surrounding Facebook's sharing of data with private companies, have increased concerns about how information may be used, shared and exploited".¹⁶

As discussed above (1.3), recent misuse of data has not only happened in the private sector; there have been a number of cases in which data has been misused within government. These include the NHS/Home Office data share, as well as the Windrush scandal in which the Home Office's mismanagement of data – specifically its [destruction of the landing cards](#) of Windrush generation arrivals into the UK in the 1950s and 1960s – led to individuals being wrongly detained and deported from the country.¹⁷ The introduction of the GDPR was inevitable in the current climate of increasing digitisation; that same climate is enhancing citizens' concern over the safety of their personal data and the impacts that the misuse of data can have.

1.9. How far do existing protections, such as in the Data Protection Act, go in promoting transparency and trust? What, if anything, should the government do to further build trust?

Transparency is hugely important to government data initiatives, and ADR UK and other such research infrastructures provide services that include the necessary transparency on behalf of the users of these services, i.e. government departments (data owners) and researchers (data users).

¹⁴ Open Data Institute (2018), 'ODI survey reveals British consumer attitudes to sharing personal data', The Open Data Institute, 12th February: <https://theodi.org/article/odi-survey-reveals-british-consumer-attitudes-to-sharing-personal-data/>.

¹⁵ Davies, Gareth, Comptroller and Auditor General (2019), *Challenges in using data across government*, National Audit Office, p.38. Accessible via: <https://www.nao.org.uk/report/challenges-in-using-data-across-government/>.

¹⁶ Ibid.

¹⁷ Gentleman, A. (2018), 'Home Office destroyed Windrush landing cards, says ex-staffer', *The Guardian*, 17th April, accessed 12/07/19: <https://www.theguardian.com/uk-news/2018/apr/17/home-office-destroyed-windrush-landing-cards-says-ex-staffer>.

Without ADR UK and other such infrastructures, the transparency work of research teams would be more challenging, and would be a piecemeal response to a national issue.

In a similar way, trust in the use of data could be further built by the formation of 'data trusts'. A data trust is 'a legal structure that provides independent stewardship of data'.¹⁸ The Open Data Institute (ODI)'s recent report, '[Data trusts: Lessons from three pilots](#)', finds that "UK trust law is not appropriate", and "not-for-profit corporations are more suitable", stressing that organisations that hold data but aren't able to use and share it would benefit from data trusts.¹⁹ "Trustees of a data trust could be better placed to make decisions about who should have access to the data and why, and can centralise and simplify processes like handling data requests."²⁰ Thus, data trusts could provide the independence and expertise needed to ensure that data are used and shared in an appropriate and secure manner, increasing trust that data are being used ethically and responsibly.

International comparisons

1.10. Are there robust international comparison measures of trustworthiness in how data about individuals is used? Which are the most effective?

The efforts of the European Statistical System (ESS) to build trust in its use of personal data have been long, sustained, and effective. It combines a Code of Practice, a regulatory process, a legal basis, strong governance, and consistency in messaging. The 2012 '[Decision on Eurostat](#)' sets out the principles of "professional independence, impartiality, objectivity, reliability, statistical confidentiality and cost-effectiveness for the development, production and dissemination of statistics by Eurostat".²¹ Furthermore, to enhance accountability of the ESS, the [European Statistical Governance Advisory Board \(ESGAB\)](#), set up in 2008, prepares an annual report for the European Parliament and the Council on the ESS's implementation of the European Statistics Code of Practice. In addition, the [European Statistical Advisory Committee \(ESAC\)](#), also set up in 2008, advises on priority setting, taking account of user perspectives.²²

New Zealand's '[Trusted Data](#)' initiative, developed by the [Data Futures Partnership](#), promotes the trusted use of personal information, and its '[Data Dials](#)' are a good example of maintaining transparency and keeping citizens informed about the use of their data. The Data Futures Partnership was specifically founded to "help lead the development of a high-trust data use system creating value for all New Zealanders"²³. It should be considered an example of good practice in ensuring that people's data are used in a transparent and trustworthy way.

Government

Objective 5: To improve public services and government operations through the effective collection, sharing and use of data

ADR UK is committed to maximising the value of government data for research purposes, enabling the development of valuable new insights into UK society to better inform policy decisions and service provision across the country.

¹⁸ Open Data Institute (2019), *Data trusts: Lessons from three pilots: Summary report*, <https://theodi.org/article/odi-data-trusts-report/>.

¹⁹ Ibid.

²⁰ Ibid.

²¹ Eurostat and the European Statistical System (accessed 03/07/19): https://ec.europa.eu/eurostat/statistics-explained/index.php/Eurostat_and_the_European_Statistical_System#European_Statistical_System_.28ESS.29.

²² Ibid.

²³ Data Futures Partnership (accessed 04/07/2019): <http://datafutures.co.nz/>.

Research area: current use of data

5.1. How effectively are government and the wider public sector collecting, sharing, analysing and storing the data it holds? What does good practice look like? What does bad practice look like?

As there has until now existed no shared UK-wide infrastructure for government departments and authorities to share their data for research purposes, most data sharing initiatives for research have been done on an individual basis and involved a 'create and destroy' model. In these cases, data are acquired and processed for an individual research project, analysed, and then destroyed once the project is completed. This is unsustainable, as the sharing of data can be an extremely laborious and expensive process. Furthermore, government departments often lose out on academic expertise when data are only analysed in-house or shared directly only with other departments or authorities. These are clear examples of bad practice, costing significant amounts of time and money and failing to contribute to the development of a more sustainable practice.

Good practice for the sharing and analysis of government data would constitute a sustainable resource offering ongoing access to a host of linked, de-identified data for research. This is what ADR UK is working to develop, so that better quality research can be produced quicker and at a lower cost. ADR UK is working with the Office for National Statistics (ONS) to develop their existing [Secure Research Service \(SRS\)](#) – a facility for providing secure access to sensitive data for research – for this purpose. The SRS provides access for Approved Researchers working on defined and approved projects which serve the public good. ONS Statistical Disclosure Control (SDC) procedures mean that analysis results cannot disclose sensitive information, so the confidentiality of data subjects is protected.

The SRS already holds a variety of data, including: business survey data, such as Business Structure Database, the Annual Business Survey and Innovation Survey; earnings data, such as the Annual Survey of Hours and Earnings; Census sample microdata and Longitudinal Study; and social surveys, such as the Labour Force Survey and General Lifestyle Survey. ADR UK is working to develop the catalogue of data already in the SRS by adding newly linked administrative data from across government.

Similarly, the [Secure Anonymised Information Linkage \(SAIL\) databank](#), established in 2007 and now a partner of ADR Wales, is “a flagship for the secure storage and use of anonymised person-based data for research to improve health, well-being and services”.²⁴ The SAIL Databank receives funding from the Welsh Government’s Health and Care Research Wales and already holds billions of health-related records made available for research in anonymised form, and is working with ADR Wales to expand opportunities for linking administrative data for research.

The preceding investment to ADR UK, the Administrative Data Research Network (ADRN), conducted a review of the methods used and challenges faced by the initiative, resulting in a set of 18 recommendations for future administrative data research initiatives. These recommendations focused on data acquisition; approval processes; controls on access and disclosure; data and metadata; researcher support; and data reuse and retention.²⁵ The review’s key recommendation for future administrative data research initiatives was to cluster projects into themes and make data reusable within a safe setting. It was deemed that ADRN’s ‘create and destroy’ model, in which data linkages were made for individual research projects and then destroyed, was unsustainable. As a result, ADR UK’s model involves acquiring and linking data within specific ‘Strategic Impact Programmes’ (SIPs), or research themes, aligned to government Areas of Research Interest. The creation of these SIPs allows linked data to be held within a secure setting - such as the ONS Secure Research Service or

²⁴ SAIL Databank (accessed 05.07.2019): <https://saildatabank.com/about-us/overview/>.

²⁵ Jones KH, Heys S, Tingay K, Jackson P, and Dibben, C (2018), 'The Good, the Bad, the Clunky: Improving the Use of Administrative Data for Research', *International Journal of Population Data Science*, 4:1:03, accessible via: <https://ijpds.org/article/view/587/1029>.

the SAIL Databank in Wales - and re-used by other accredited researchers working to address research questions within the same general theme, rather than destroyed after single use. This is a far more efficient and sustainable model for sharing linked administrative data for research. The full report with all 18 recommendations can be read [here](#).

There are already examples of how linked administrative data curated by ADR UK has offered valuable insights to improve public service provision in the UK, such as:

Improving the lives of farmers in Northern Ireland

One of the most impactful pieces of [research](#) undertaken by the Administrative Data Research Centre Northern Ireland (ADRC-NI), part of ADR UK, involved both academic researchers and researchers from the Northern Ireland Department for Agriculture, the Environment and Rural Affairs (DAERA). One of DAERA's key goals was to improve the lives of rural dwellers and farming communities, with prior research indicating that farmers are at a higher risk of health issues due to long working hours, physical labour, isolation and accidents. Another key target was to improve expertise and education among farming populations; in particular, IT literacy.

In conjunction with the ADRC-NI Statistical Methodological Officer at Queen's University Belfast, this research project sought a better understanding of the health and educational issues within farming communities, linking these to DAERA policies and programmes. This allowed for comparison of the individual circumstances of farmers and farming families in order to better understand their livelihoods.²⁶

Utilising ADRC NI's ability to link data from the Northern Ireland Agricultural Census and the Population Census, researchers provided key evidence to inform government policy, establish robust baseline information, provide evidence of need, and aid in research targeting and outcome monitoring. This research has already saved DAERA £350,000 (the cost of conducting their own survey).

Tackling fuel poverty in Wales

[ADR Wales research](#) using linked administrative data has provided valuable insights into the connection between fuel poverty and health. In partnership with the Welsh Government, ADR Wales discovered that those who accessed the government's Warm Homes Nest scheme were less likely to seek help from the NHS and be admitted into hospital for cardiovascular and respiratory health issues. The 'NEST' measures appeared to have a protective effect, which was observed across all age groups.²⁷

This research suggested that fuel poverty schemes have a powerful impact beyond helping people heat their homes. Additionally, it led to an extension of funding for the scheme between 2018-21, as well as funding for the House Conditions Evidence Programme and the Housing Stock Analytical Resource.

Outside of ADR UK's remit, Welsh Government's [Well-being of Wales 2018 report](#), which measures current levels of wellbeing in Wales against the country's seven [Well-being Goals](#), is also a good example of the use of linked administrative and survey data for providing valuable insights into society and where policy improvements could improve lives. Key findings included, for example: 'Life expectancy has been increasing over the long term but in recent years those increases have shown signs of slowing and even stopped'; 'In the last year overall labour market performance in Wales has continued to be strong, with gaps in employment and inactivity rates between Wales and the UK remaining low in historical terms'; and 'Relative income poverty has remained steady for over a

²⁶ DAERA, *Farmer Equality Indicators: Report* (2018): <https://www.daera-ni.gov.uk/articles/farmer-equality-indicators-report>.

²⁷ Welsh Government (2017), *Fuel poverty data linking project: findings report no. 1: initial findings on the impact of the Warm Homes Nest Scheme*: <https://gov.wales/fuel-poverty-data-linking-project-findings-report-no-1-initial-findings-impact-warm-homes-nest>.

decade, and is highest amongst children.²⁸ These are important insights for Wales' ongoing ambitions to reach its seven Well-being Goals.

In order to maximise the potential of administrative data for research, shared approaches are needed across government. An example of good practice in this sense is [Health Data Research UK \(HDR UK\)](#)'s [UK Health Data Research Alliance](#). The purpose of the alliance is to 'facilitate partnership working across NHS organisations and other health data custodians, leading to an ethical, consistent approach to the use of data and public engagement'.²⁹ The alliance intends to coordinate the adoption of 'standardised tools, techniques, conventions and technologies to drive greater use of healthcare data for research and innovation in a trustworthy way'.³⁰ A standardised approach to administrative data research, to ensure consistency and a shared understanding of the methods and goals of the practice, is an example of good practice in data use in government and should be considered across other areas of data use.

5.2. What are the main barriers to more effective data use within government? Are there barriers in cases where government works with the private sector?

The 2017 [Digital Economy Act \(DEA\)](#) removes the legal barriers to data sharing for research purposes, and is the legislation which underpins the majority of the research planned by ADR UK. Chapter 5 of the DEA - 'Sharing data for research purposes' - says: 'Information held by a public authority in connection with the authority's functions may be disclosed to another person for the purposes of research which is being or is to be carried out', as long as certain conditions, such as the de-identification of personal data, are met.³¹ The barriers that remain are those in the customs, practices and precedents of departments and of research teams that were established before the 2017 Act.

However, there exists a lack of a shared understanding of the DEA and the assurance processes around it, and a clear governance structure is therefore needed to promote a better shared understanding of the legislation surrounding data use within government. As stressed by the National Audit Office (NAO) in their recent report, '[Challenges in using data across government](#)', "the Digital Economy Act has so far not given departments the reassurance they need to be confident about sharing data legally".³² They found that departments would welcome more support on "how to use the Act appropriately to support data-sharing".³³ The Department for Digital, Culture, Media & Sport is reportedly undertaking work to support people to use and better understand the Act.

In its report the NAO highlights three main barriers to the effective use of data within government: 1) Data is not always seen as a priority; 2) The quality of data is not well understood; and 3) There is a culture of tolerating and working around poor-quality data.³⁴ The report highlights that a lack of standards across government has led to inconsistencies in data, and suggests that government needs "a structured approach to investing in improving and using data", and "to develop the capability, leadership and culture to support sustained improvement in the quality of information available".³⁵ ADR UK supports these recommendations, which are of key relevance to our work linking government datasets for research in the public interest.

²⁸ Welsh Government (2018), *Well-being of Wales 2017-2018*, accessible via: <https://gov.wales/well-being-wales-2018>.

²⁹ Health Data Research UK, *UK Health Data Research Alliance*, accessed 12/07/2019: <https://www.hdr.uk.ac.uk/digital-innovation-hubs/uk-health-data-research-alliance/>.

³⁰ Ibid.

³¹ Digital Economy Act (2017): <http://www.legislation.gov.uk/ukpga/2017/30/contents/enacted>.

³² Davies, Gareth, Comptroller and Auditor General (2019), *Challenges in using data across government*, National Audit Office. Accessible via: <https://www.nao.org.uk/report/challenges-in-using-data-across-government/>.

³³ Ibid.

³⁴ Ibid.

³⁵ Ibid.

The report furthermore highlights that silo working in government can inhibit progress and make partaking in data sharing initiatives seem not to be worthwhile. “The Data Advisory Board found that return on investment for a department can often be difficult to justify in data projects because the benefits might be seen elsewhere in government”.³⁶ Government departments and public bodies therefore need a shared infrastructure for maximising the potential of administrative data, and to work together towards a shared goal.

For good reason, some of the main concerns for government departments regarding the sharing of administrative data relate to safety and security, which is a major barrier to the efforts of data sharing initiatives such as ADR UK. Departments have an obligation to protect the privacy of their service users and ensure that their data are not misused in any way. Data sharing initiatives must therefore demonstrate the utmost levels of safety and security in their handling of administrative data. As discussed above (1.1), ADR UK’s use of government administrative data abides by the ‘Five Safes’ to ensure that data are held securely and used only for purposes within the public interest. This reassures data owners that data will be kept safe.

5.3. Are there areas within data management, use and access where there is a skills gap in government? What additional data skills are needed in government?

A shortage of data scientists and analysts within government is a major barrier to more effective data use within government, and is a major reason why government data cannot be improved in terms of quality; departments simply lack capacity and resource. One of UK Research and Innovation (UKRI)’s four focus areas in its ‘[Foundations for excellent research and innovation](#)’ is on ‘leading talent: nurturing the pipeline of current and future talent’. UKRI recognises that “a highly-skilled and diverse workforce is crucial to delivering the Government target of a total of 2.4% of GDP to be invested by public and private sectors in research and development by 2027.”³⁷ In the remit of their new Future Leaders Fellowship, they aim to develop the next generation of innovation and research leaders, with particular reference to the need to create more highly valued roles for data scientists, “for the teams that are needed to tackle tough challenges.”³⁸ Such a focus should be adopted across government, to ensure that the skills and knowledge of the government workforce remain in tow with developments in data and technology.

Furthermore, there exists a general lack of awareness about what can be done legally and ethically with data in government, as mentioned under 5.2 above. Organisations are slow to adopt changes for fear of getting it wrong, and the repercussions that may result.

Research area: improving data use

5.6. What can government and the wider public sector do to improve its collection, sharing, analysis and storage of data?

To improve the sharing of data for research purposes, government needs a central, sustainable, secure resource which allows ongoing access to linked, de-identified data from across government. This is what ADR UK is working to develop, so that better quality research can be produced quicker and at a lower cost, whilst data are kept safe. Government departments should take advantage of existing investments like ADR UK, to reap the benefits of the work that is already being done to improve the sharing and analysis of data, and to support the development of an essential infrastructure for administrative data research.

³⁶ Davies, Gareth, Comptroller and Auditor General (2019), *Challenges in using data across government*, National Audit Office. Accessible via: <https://www.nao.org.uk/report/challenges-in-using-data-across-government/>.

³⁷ UK Research and Innovation, *Foundations for excellent research and innovation*, accessed 12/07/2019, <https://www.ukri.org/about-us/strategic-prospectus/foundations-for-excellent-research-and-innovation/>.

³⁸ Ibid.

We are, however, aware of the ultimate need to move towards a system in which live data can be funnelled into a place where researchers can gain immediate access. Administrative data is effectively out-of-date as soon as it is taken away from the department, and research findings can therefore only reflect the issue at the time the data was removed. This requires a shared infrastructure and central governance structure invested in over time, as called for by the National Audit Office (NAO) in their [recent report](#).³⁹

ADR UK supports the Institute for Government (IfG)'s recent call for '[Five things the UK government should publish](#)', particularly 4) More comprehensive performance data on public services. Public policy and the provision of public services should be based upon the best evidence possible, and high-quality data on public service performance is therefore essential and should be made available for accredited researchers to produce valuable insights into what does and doesn't work. The IfG gives several examples of public service performance data that should be available, including: better data on police activity; data on GP consultations; and more data on neighbourhood services.⁴⁰

The 2017-2020 [Government Transformation Strategy](#) already sets out plans for making better use of data by 2020, emphasising that 'data is a critical resource for enabling more efficient, effective government and public services that respond to citizens' needs'.⁴¹ The strategy states that 'we must earn and retain the trust of citizens and provide reassurance that personal and sensitive data is treated safely, securely and ethically within appropriate governance frameworks,' and makes it clear that 'we need to get better at sharing data across organisational boundaries in ways that citizens are comfortable with'.⁴² These statements remain relevant and important, particularly to the work of ADR UK in enabling better access to government data for researchers, and any National Data Strategy should complete and extend the already established goals of the Government Transformation Strategy.

5.7. How can the government create incentives for these improvements?

The benefits to government and society as a whole of sharing administrative data for research via a central infrastructure are manifold. Opening data up to broader expertise, from both academic institutions and other research organisations, creates opportunities for government's research questions to be answered, to better inform and influence policymaking and service provision. Fundamentally, better access to administrative data enables research that enhances understanding of how UK society works, and how its problems can be addressed, to improve the lives of people across the country.

Funding is also a key consideration. If initiatives like ADR UK are well funded in an ongoing fashion, this eases the financial burden of data sharing on individual departments and authorities. Furthermore, finding solutions to problems using an existing shared infrastructure rather than investing in new solutions would save public money. Failing to use the wealth of administrative data that already exists across government for research can also be considered a 'missed use' of data and is damaging to the UK's data economy. Meanwhile, policies are not based upon the best evidence possible.

³⁹ Davies, Gareth, Comptroller and Auditor General (2019), *Challenges in using data across government*, National Audit Office. Accessible via: <https://www.nao.org.uk/report/challenges-in-using-data-across-government/>.

⁴⁰ Freeguard, Gavin (2018), *Gaps in government data: Five things the UK Government should publish*, Institute for Government, pp.4-5., accessible via: <https://www.instituteforgovernment.org.uk/publications/gaps-government-data>.

⁴¹ Government Digital Service (2017), *Government Transformation Strategy*, accessed 15/07/2019, <https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020/government-transformation-strategy>

⁴² Ibid.

5.8. What best practice examples of data use in government can we learn from?

The NAO's recent '[Challenges in using data across government](#)' report highlights that departments may find it beneficial "to enhance the skills and understanding of executive board members to raise their awareness of the importance of good-quality data to their business".⁴³ The Ministry of Housing, Communities & Local Government has reportedly already begun providing training for leaders within local authorities to help support better use of digital services.

The report also highlights a recent move by the Department for Education to place the national pupil database in a secure environment (the ONS Secure Research Service), to increase both security and efficiency of use for accredited researchers.⁴⁴ The Department for Education are now working with ONS as part of ADR UK to link the All Education Dataset for England (AEDE) to the 2011 Census data. This linkage will allow data on educational attainment of children to be connected to their household circumstances, enabling more accurate measurement of the household structures that shape children's early and subsequent experiences.

5.10. What are the main opportunities for improving outward facing public services through effective data use?

Research using linked administrative data from across different areas of government has the potential to reveal valuable new insights into UK society, allowing policymakers to make better informed decisions that improve public service provision across the country. With a more comprehensive picture of life in the UK and a better understanding of how different areas of service provision interact with one another, policymakers can shape policy in ways which work better for all. This is ADR UK's ultimate goal.

Objective 6: To achieve alignment in government around data, with data shared and used cooperatively wherever appropriate

ADR UK is committed to improving data sharing for research, so that researchers are better-equipped to paint an accurate picture of UK society and policymaking can be based upon the best evidence possible.

Research area: Barriers to data sharing within government

6.1. When should public authorities open up access to data they hold with other departments? When should they not?

Departments should open up the data they hold when a public service may be delivered better, or better understood and evaluated, as a result; thus, when the opening up of data has clear potential to benefit the public good. It should be opened up when there is real potential to better inform policy, where an understanding is needed of issues that fall across or between departmental responsibility, and/or when the actions of one department may impact heavily on another. However, data should not be opened up to target *operational* enforcement activities such as was the case with the NHS/Home Office data share (see 1.3 above). It is also important to understand and be aware of unintended consequences of research findings using government data, such as upon certain marginalised groups.

⁴³ Davies, Gareth, Comptroller and Auditor General (2019), *Challenges in using data across government*, National Audit Office, p.43. Accessible via: <https://www.nao.org.uk/report/challenges-in-using-data-across-government/>.

⁴⁴ Ibid.

Departments should make sure to exercise proper caution when using data for some of the difficult, but necessary, purposes that it falls to government to perform. Opening up data would be wrong if that makes the purposes of holding those data in the first place more difficult to achieve. It is also important that time and effort is not invested in opening up access to data when the data are of such poor quality that analysis of it risks undermining trust and understanding (see 5.2 above). In cases such as these, money should rather be invested in creating better quality data capture systems.

All data shared with researchers via ADR UK must pass a 'public good' test to ensure that data are only being used for purposes in the public interest. Researchers wishing to use the data acquired and linked by ADR UK must first go through a stringent accreditation process, and their research project must be examined by an approvals panel to ensure it is ethical, delivers value and benefit for the public, and that it actually needs access to the data the researcher has asked for. Once they have received accreditation and approval, researchers must then access the data via a secure physical research facility (or an approved, secure connection to that facility) provided by one of the accredited ADR UK partners. Researcher activity at these facilities is closely monitored, and outputs checked, to ensure the data has not been misused in any way. It is essential that public data are held safely and securely and protected from misuse; in situations where this is assured, authorities should be confident that the benefits of sharing their data outweigh the associated risks.

6.2. When do people find that government has improved services through better use of data? In which areas would people most like to see data better used?

As discussed under section 1.3 above, a public panel conducted by the research team for the [Northern Ireland Cohort for the Longitudinal study of Ageing \(NICOLA\)](#) in partnership with the engagement team from the Administrative Data Research Centre Northern Ireland (ADRC NI), part of ADR UK, found that the participants were surprised to learn that there were several instances in which they would have expected government departments to already be linking data to answer policy or service delivery-related questions, when they weren't. Once the work was done to educate participants about data collection, linkage and use, they were more perturbed to find that their data wasn't being used in a more all-encompassing way, to answer questions and tailor services that might have a direct positive impact on their lives. This work suggests that rather than assuming people are cautious or afraid of their data being used, once they understand what the data can tell researchers, they are more frustrated that the data are not already being shared with and by people who can use it to improve their lives.

This is reflected in the findings of the 2014 [Dialogue on Data](#) (see 1.6 above), in which participants initially questioned why an initiative to enable administrative data research was necessary as they had assumed that government data was already routinely linked and shared across departments and services. However, they were also somewhat reassured that there were so many barriers to sharing currently in place, further emphasising the need to focus on increasing public trust in government's use of data.⁴⁵

6.3. What issues are there around government increasing access to the data it holds between departments and other parts of the public sector? Does the use of Privacy Enhancing Technologies sufficiently counter any such risks?

When data are shared with a third-party, it is never entirely risk-free. An almost inevitable result of increased access to data between departments and other parts of the public sector are situations in which either sensitive information is disclosed, or data are misused for purposes not in the public interest. However, risk and benefit must be weighed up, and ADR UK takes the view that the benefits to better sharing of data in secure and controlled environments far outweigh the potential risks.

⁴⁵ Cameron, Daniel, et al. (2014), *Dialogue on Data: Exploring the public's views on using administrative data for research purposes*, Ipsos Mori: <https://www.ipsos.com/ipsos-mori/en-uk/dialogue-data>, p.3.

When data sharing initiatives take special measures to ensure that data are kept as safe and secure as possible – as ADR UK does with the ‘Five Safes’ (see 1.1 above) – risk can be minimised and benefit maximised. However, where the need for these measures is not properly understood and acted upon, the level of risk is increased. There must therefore exist a stringent data governance structure which is well understood and abided by across government, as well as effective regulation that nevertheless does not unreasonably limit benefit.

6.4. What kinds of problems do barriers to data sharing create for government departments and public sector organisations?

Barriers to data sharing between government departments and authorities can lead to important policy decisions being made in isolation, with an incomplete understanding of how their policies affect other areas of life. A comprehensive understanding of UK society is not possible without data from across government being shared and linked. Just taking one area of UK society, criminal justice, it is not currently possible to link data on individual crimes (held by the Home Office) to data on individual perpetrators of these crimes (held by the Ministry of Justice), or know if it is the same for different people circling around the civil and criminal justice systems (data also held by the Ministry of Justice). ADR UK are facilitating a programme of work with HO and MoJ to bring these various sources of data together for research purposes.

It can be argued that better use of data within the health sector has led to better diagnosis and treatment (especially drug treatment) of disease. However, examples of sharing personal health data to support public health research, rather than clinical research, are limited in England – although this is not the case in other countries of the UK. This could explain why the social and environmental causes of disease often remain unknown or unaddressed. Rates of obesity (which is a major risk factor for a range of cancers and other life-limiting diseases), mental illness and other health issues in the population are increasing. It seems the use of personal health data is reserved for attempts to cure of disease rather than the promotion of wellbeing and prevention.

Nevertheless, Kerina Jones et al. in their 2017 paper, ‘The other side of the coin: Harm due to the non-use of health-related data’, highlight that there is ample indirect evidence that the ‘non-use’ of health data “is implicated in the deaths of hundreds of thousands of people and potentially £billions in financial burdens to societies”.⁴⁶ The authors stress that, although actual proven evidence of the harm of non-use of health data is hard to pin down as such, “it is nevertheless a real problem with widespread and serious, if largely unquantifiable, consequences”⁴⁷. The paper stresses that the most effective initiatives for addressing the harm of data non-use will be those that: “firstly, understand the pertinent sources, types and reasons for data non-use in a given domain in order to meet the challenges and create appropriate incentives and repercussions; and secondly, are cognisant of the multiple aspects to this complex issue in other domains to keep benefits and limitations in perspective, to move steadily towards socially responsible reuse of data becoming the norm to save lives and resources.”⁴⁸

As data from across different health authorities – much like throughout all other areas of government – exists in silos, a person’s interaction with one health service cannot be considered in the context of their interactions with others. In their paper, Jones et al. highlight a devastating case of a vulnerable young boy who died in 2011 following systematic abuse. “There were multiple visits and reports to the GP practice, health visitor, community paediatrics and emergency departments. Each instance was isolated as the data on presentations elsewhere were not available. Because of this, instances were not seen in context and the problems were not identified in time to save the life of the child”.⁴⁹ If healthcare data were joined up, tragedies such as this could be avoided.

⁴⁶ Jones, Kerina et al. (2017), ‘The other side of the coin: Harm due to the non-use of health-related data’, *International Journal of Medical Informatics*, 97, pp.43-51., p.43. <https://www.ncbi.nlm.nih.gov/pubmed/27919394>.

⁴⁷ Ibid.

⁴⁸ Ibid.

⁴⁹ Ibid., p.44.

6.5. What kinds of problems do barriers to data sharing create outside of government?

Researchers

For academics and other researchers, poor access to government-held administrative data can hinder the ability to paint an accurate and comprehensive picture of life in the UK, in the same way that it can for government. Survey data alone, for example, is limited to those who choose to take part, meaning some groups of people are underrepresented, and that research findings resulting from such data may not tell the full story. This can then trickle down into government decision-making, as decisions made on the basis of such findings may not benefit everyone. Administrative data, on the other hand, includes everyone who comes into contact with government and public services, and is therefore far more representative of life in the UK.

Furthermore, primary data collection can be expensive and time-consuming for researchers and when commissioned by organisations aiming to produce findings that have the potential to benefit the public good, as well as being a burden upon respondents. Access to existing administrative data can save time and money creating fresh data that is likely to be less representative of the issue being investigated. In addition, when access to administrative data is only possible on individual bases, the time spent acquiring data may cause research funding to run out before the data can be successfully acquired and analysed.

A sustainable resource which allows ongoing access to readily linked, de-identified data for research, like that being developed by ADR UK, will allow these barriers to be removed, and for better quality research to be produced quicker and at a lower cost.

Civil society groups and NGOs

A central tenet of the Administrative Data Research Centre Northern Ireland (ADRC NI), part of ADR UK, has always been public engagement and promoting public benefit through widening participation in data research, specifically to civil society groups. These organisations are often best-placed to understand and reflect issues facing different communities to government and policymakers and are often those making the most use of evidence developed by academic researchers. Supporting both their engagement in academic data research as well as the development of their own research using government data is therefore key to enhancing transparency, accountability, and participation when it comes to data. The argument can also be made that as government data are made up of data from individual members of the public, this is public data, and therefore publics – in this approach represented by civil society groups – should have a say in the way the data are used.

To facilitate this, the public engagement work of the ADRC NI from 2014-18 focused on creating pathways for civil society and NGOs to directly access linked administrative data for research purposes in similar ways to our accredited academic and government researchers. The proposals approved by the governance structures of the former ADRN, though never put in place, would have allowed researchers from NGOs to be named on projects in partnership with academics, and, eventually, to lead projects of their own. This would have been a massive step in the democratisation of data and the enhancement of government and data transparency.

However, reticence within government to data sharing, even between government departments and even after the advent of the Digital Economy Act, have prevented initiatives like this from being implemented. From a research perspective, this means that academic teams are missing out on analysis perspectives from NGOs and people with grassroots knowledge of the issue being studied. ADRC NI's experience has demonstrated the benefit that academic and governmental research gets from partnering with civil society, and the disservice done to knowledge generation and exchange when these partnerships are prevented from securing equal access to data. From a civil society perspective, it means that data are still treated as something that only certain groups of people can have access to, even when it is ostensibly public data. By not allowing wider access to data for research and innovation, these barriers and the structures and the culture that create them are hobbling potential positive impacts on knowledge, policy and public services.

ADRC NI's engagement work strongly suggests that de-identified government data should be accessible to both academics and NGOs on an equal basis.

6.6. Part 5 of the Digital Economy Act is designed to improve the use of publicly held information for specific purposes. How are organisations making use of the relevant powers, appropriately and effectively?

Part 5, Chapter 5 of the 2017 [Digital Economy Act \(DEA\)](#), 'Disclosure of information for research purposes', states that: "Information held by a public authority in connection with the authority's functions may be disclosed to another person for the purposes of research which is being or is to be carried out," subject to certain criteria being met.⁵⁰ ADR UK is making use of this power whilst meeting the necessary criteria, to make de-identified, linked administrative data available for research. By doing this, we are enabling valuable new insights into UK society, to better inform policymaking and ultimately improve lives.

Despite the very valuable, and long sought-for lawful authority to provide data for research now being available, few departments seem to be making concrete plans to take advantage of these provisions. An exception is the Department for Education. This was identified in the NAO report '[Challenges in using data across government](#)' (as mentioned under 5.8):

4.19 Some departments have previously struggled to understand how to share data effectively for research, but departments are starting to understand how best to support the needs of others. For example, the Department for Education previously shared around 500 different versions of the national pupil database with several hundred different organisations (for research, statistics and policy development). Its new chief data officer recognised the risk and placed the national pupil database in a secure environment with access limited to accredited parties.

4.20 The Digital Economy Act has so far not given departments the reassurance they need to be confident about sharing data legally. Our discussions with departments have shown us that they would welcome more support on how to use the Act appropriately to support data-sharing. The Department for Digital, Culture, Media & Sport is undertaking a series of roadshows to help support people to use the Act and understand where it might be appropriate. This will include case examples which can be helpful for users. It is also developing a data processing and sharing framework to help departments.⁵¹

International comparisons

6.7. Are there robust international comparison measure and case studies of governments' effective use of data within government and to deliver public services?

New Zealand's Integrated Data Infrastructure

New Zealand's [Integrated Data Infrastructure \(IDI\)](#), which holds de-identified microdata about the country's people and households, has enabled researchers to gain valuable insights into New Zealand's society and economy, to improve the outcomes of New Zealanders. Public engagement has been an important part of the IDI's work and key to its success, further emphasising the need for thorough public engagement in government data initiatives within the UK.

The data are about life events, like education, income, benefits, migration, justice, and health. It comes from government agencies, Stats NZ surveys, and non-government organisations (NGOs).

⁵⁰ Digital Economy Act (2017), Part 5, Chapter 5, 64. 'Disclosure of information for research purposes', accessed 12/07/2019: <http://www.legislation.gov.uk/ukpga/2017/30/section/35/enacted>.

⁵¹ Davies, Gareth, Comptroller and Auditor General (2019), *Challenges in using data across government*, National Audit Office, accessible via: <https://www.nao.org.uk/report/challenges-in-using-data-across-government/>.

The data are linked together, or integrated, to form the IDI. The IDI complements the [Longitudinal Business Database](#) (LBD), which holds linked microdata about businesses. The two databases are linked through tax data. The IDI uses the 'five safes framework' – safe people, safe projects, safe settings, safe data and safe output – to manage safe access to its data. Access to data is only granted if all of the five 'safes' are met.

The findings of research using this data can inform work on complex social and economic issues. These issues include topics like crime and vulnerable children. Hundreds of projects have already used linked administrative data from the IDI to provide valuable insights into New Zealand society – more information on the topics that have been covered so far can be found in the [Stats NZ Store House](#).

IDI study: Childhood forecasting of a small segment of the population with large economic burden

[The Dunedin Study](#) – an ongoing longitudinal study of the health, development and well-being of a sample of New Zealanders – found that a small proportion of the 1,000-member cohort accounted for most of the cohort's hospital stays, welfare benefits, prescription fills and criminal convictions.

A [2016 study](#) by A. Caspi et al., 'Childhood forecasting of a small segment of the population with large economic burden', used linked IDI data from five agency records including crime, social welfare, injury claims, prescription fills and hospital nights to extend the Dunedin Study's analysis by testing the extent of service use concentration using whole population data. Specifically, it tested the hypothesis that the cost of public health and social services is concentrated in a relatively small proportion of the New Zealand population who use multiple public sectors at a high frequency. It also tested whether this concentration varies by type of service (hospitalisations, pharmaceutical use, ACC claims, MSD benefits, criminal court conviction), sex, and age/birth year.⁵²

Contact

Elizabeth Waind
Communications Officer, ADR UK

elizabeth.waind@ukri.org
+44 7540 705148

Twitter: [@ADR_UK](#)

⁵² Caspi A. et al. (2016), 'Childhood forecasting of a small segment of the population with large economic burden', *Nature Human Behaviour*, 1:5, accessible via: https://moffittcaspi.trinity.duke.edu/sites/moffittcaspi.trinity.duke.edu/files/Caspi_NHB_Childhood_Forecasting_2016_0.pdf.