



Trust, Security and Public Interest: Striking the Balance

*A review of previous literature on public attitudes towards the
sharing and linking of administrative data for research*

May 2020

Elizabeth Waind
ADR UK (Administrative Data Research UK) Strategic Hub
Economic & Social Research Council



Executive summary

This literature review explores previous research into the UK public's attitudes towards the linking and use of administrative data for research, and the conditions under which it is felt it should and should not happen.

In recent years, a large amount of public consultation and attitudinal work has been conducted in relation to the collection and use of data, as well as specifically in relation to the linking and use of administrative data for research. This review of previous work finds that the public is broadly supportive of the use of administrative data for research, as long as three core conditions are met:

1. **Public interest** – any research using administrative data must demonstrate that it is in the public interest and has potential to lead to tangible benefits for society;
2. **Privacy and security** – data being linked and used for research must be de-identified, and protections must be in place to prevent it from being re-identified or misused;
3. **Trust and transparency** – trust in those holding and using data is paramount; and transparency around how data is held and used is essential.

None of these three conditions is sufficient in isolation; rather, the literature shows that the public's support for the use of administrative data for research is underpinned by a minimum standard of all three. The literature also shows, however, that in certain cases where the standard of one condition is very high – for example, public interest – this could mean that of another – for example, privacy and security – may, if necessary, be lower. An appropriate balance must be struck, and the proposed benefit must outweigh the potential risk. Nevertheless, certain minimum standards are expected in all situations to secure public support.

Three core conditions of public support for sharing administrative data for research



Public interest

Research using administrative data must be in the public interest and demonstrate real potential for public good.



Privacy and security

Data shared for research must be de-identified, with protections in place to prevent it from re-identification or misuse.



Trust and transparency

Trust in those holding and using data is paramount; and transparency around how data is held and used is essential.

This review shows that broad, conditional support for the use of administrative data in research has not only been found consistently, but has also been held over time, with data collection for the studies included spanning more than a decade from 2006-2018.

In light of this, it is now appropriate to move beyond widescale, general consultation on the use of administrative data for research and instead build upon existing knowledge by delving into specific areas of research. This could involve engaging with sub-sectors of society relevant to these areas, or with a cross-section of society on a particular aspect of the use of administrative data. This enables a greater focus on the issues important to the people whose lives may be directly affected by research conducted using their data, in a more specific context. **The purpose of such an approach would not be to consult on whether research using administrative data *should* be done – as has been the focus of previous literature – but rather to guide *how, why and when* it is done.** Nevertheless, it is important to continue to monitor and respond to any changes to public attitudes and adapt approaches if necessary.

This literature review sets out the findings of previous research and finishes by presenting the details of ADR UK's approach to enabling better use of administrative data for research – and to engaging the public with our work – in light of these.

Contents

1. Introduction	4
1.1. Aims and objectives	5
1.2. Definitions	5
2. Overview of existing literature	7
3. Existing public knowledge of administrative data research	8
4. Public interest	10
4.1. Data types	12
4.2. Demographic differences	13
5. Privacy and security	15
5.1. De-identification and anonymisation	15
5.2. Data access and security	16
5.3. Governance and regulation	18
6. Trust and transparency	19
6.1. Trust	19
6.2. Transparency	22
7. Striking the balance	24
8. Beyond transparency	26
9. The ADR UK approach	28
9.1. Public engagement with ADR UK	29
The Strategic Hub	30
ADR Northern Ireland	31
ADR Scotland	32
ADR Wales	33
10. References	34

1. Introduction

Administrative data is information created when people interact with public services such as schools, the NHS, the courts or the benefits system. This data is originally collected for operational purposes: to enable public services to carry out their day-to-day work, to monitor and improve their performance and to keep providing services in an effective way. For instance, the Department for Education collects data on educational attainment to identify schools and subjects where improvement is needed; and the NHS records details of admissions and appointments to monitor trends in hospital activity. Administrative data also includes basic information about people, such as notifications of births, deaths and marriages, the electoral register, and national censuses. Across the UK, administrative data is currently a largely untapped but information-rich resource for social and economic research. This wealth of data has the potential to provide valuable insights into our society and highlight where change is needed to improve policy and service provision.

The 2017 Digital Economy Act provides the legal framework for [ADR UK \(Administrative Data Research UK\)](#)'s work, enabling public authorities to provide administrative data to researchers where six conditions as set out under Section 64 of the Act – 'Disclosure of information research for purposes'¹ – are met. The six conditions are:

- i) *the information disclosed to the researcher does not identify an individual and any indirect identification is unlikely;*
- ii) *those processing the data for disclosure to researchers must minimise the risk of accidental disclosures and prevent any deliberate disclosures to others;*
- iii) *the disclosure to the researcher is by the data owner or another person authorised to do so;*
- iv) *the purpose of the research has been accredited;*
- v) *the processors of the data and the researchers are accredited;*
- vi) *and all parties have regard to the Code of Practice issued under this legislation.*

In addition to operating in line with this legal framework, it is essential that we operate ethically and openly, and in the knowledge that the public is supportive of the way in which their data is handled and used. Administrative data includes all those who interact with public services and therefore the majority of the population; that's what makes it so useful to research, and so valuable to a more thorough understanding of what does and doesn't work in public policy. If we are to use data about the public, this cannot be done without the public's support and, where possible, their input.

To maximise the impact of the research we enable, we must seek to better understand the public's interests and concerns in relation to the use of administrative data for research and shape our approach on the basis of it. We must also ensure the public remains engaged with our work as it progresses, via a meaningful and mutually beneficial dialogue.

¹ Digital Economy Act 2017, Section 64, '[Disclosure of information for research purposes](#)' (accessed 27/04/2020).

1.1. Aims and objectives

This literature review has two main aims:

1. To explore public attitudes, as found by previous research, towards the sharing, linking and use of administrative data for research, and the conditions under which it is perceived it should and should not happen;
2. To be a source of advice on approaches to public engagement, not only for ADR UK's work but also for other organisations and researchers working with administrative data.

To meet these aims, a review of previous public consultations and attitudinal studies on the topic has been completed, with a specific focus on previous work conducted in the UK. This includes academic and policy papers as well as existing reviews of previous research in the area. This is not a systematic review, nor a review of methodologies; rather it is a narrative review of the main themes identified across previous work.

This review will begin with an overview of existing literature and will then set out the general trends in previous study participants' spontaneous knowledge of and sentiment towards the use of administrative data for research. Next, it will introduce the three main conditions for public support for administrative data research identified across the literature: public interest, privacy and security, and trust and transparency. Finally, it will set out ADR UK's approach in light of the findings of this review.

1.2. Definitions

For the purpose of this review, the following key terms are defined:

Anonymised data, as defined by the Information Commissioner's Office (ICO) (2012, p.48), refers to "*data in a form that does not identify individuals and where identification through its combination with other data is not likely to take place*". For data to be anonymous, de-identification (see below) alone is therefore not enough; other safeguards such as those set out under the 'Five Safes' – Safe people, Safe projects, Safe settings, Safe outputs and Safe data (Office for National Statistics 2017)– provide the conditions under which identification is not likely to take place, and under which data can be considered anonymous.

De-identified data refers to data which has had all personal identifying elements such as names, addresses and identifying numbers removed, meaning individuals are therefore no longer directly identifiable. The Digital Economy Act *Research Code of Practice and Accreditation Criteria* states that: "*Data must be de-identified before they can be made available so that the data do not directly identify individuals and are not reasonably likely to lead to an individual's identity being ascertained (whether on its own or taken together with other information)*"².

² Digital Economy Act 2017, Part 5: Codes of Practice, '[Research Codes of Practice and Accreditation Criteria](#)', Part 1: Code of Practice (accessed 27/04/20).

It is important to note that both 'de-identified' and 'anonymised' data are referred to across the literature reviewed, with specific definitions not consistently provided in all cases. This does not detract from the fundamental findings of the studies included in this review; however, when the terms are used in the context of a previous study's findings, their precise definitions should be considered with some caution.

Public engagement, as defined by Gene Rowe and Lynn J. Frewer in their '*A typology of public engagement mechanisms*' (2005, pp.254-256), is a combination of three concepts – 'public communication', 'public consultation' and 'public participation'. These three concepts can in turn be understood as follows:

- In '**public communication**', information is conveyed to the public (e.g. by researchers or data handlers) in a one-way flow. There is no involvement from the public – their feedback is neither sought nor addressed.
- In '**public consultation**', opinions are sought from the public in relation to particular topics or initiatives. No formal dialogue is involved in this type of engagement – it is an exercise aimed at gathering information from the public regarding their opinions.
- '**Public participation**' refers to an exchange of information between members of the public and those leading on the initiative in question. This usually takes place in a group setting and involves representatives from both parties providing information to one another. In this form of engagement, the opinions of both the public and those seeking their views may be transformed as each party learns about the interests and concerns of the other.

Transparency, as defined by the ICO in the context of the General Data Protection Regulation (GDPR), "*is about being clear, open and honest with people from the start about who you are, and how and why you use their personal data.*"³ In relation to the use of administrative data for research (for which data is de-identified and therefore no longer personal), this means ensuring that clear and complete information about what data is being used and for which purpose is easily accessible to the public, for example on a dedicated website.

³ Information Commissioner's Office, Guide to the General Data Protection Regulation (GDPR), '[Principle \(a\): Lawfulness, fairness and transparency](#)' (accessed 17/04/20).

2. Overview of existing literature

Previous consultation and attitudinal work with members of the UK public in relation to the use of administrative data for research has in large part focused on health data. This is possibly due to a perceived greater value and sensitivity of this type of data (see Section 4.1). Nevertheless, there have also been a number of papers which have dealt with the use of administrative data more broadly, with some focusing on cross-sectoral data linkage (for example, Davidson et al. 2012 and 2013). Most previous work has focused more generally on public acceptability of data sharing and use, rather than on *how* data should be used in specific cases or areas of research for the benefit of society. However, it has explored the particular conditions under which data sharing is acceptable, therefore offering insight into acceptable approaches for data infrastructures more broadly.

The majority of papers included in this review have been conducted or commissioned by data infrastructures and social research institutions hoping to inform their own use of administrative data, rather than by academics for purely exploratory purposes. Most have involved physical engagement activities in the form of focus groups and workshops, whilst some have involved surveys or questionnaires, both online and in-person, as well as telephone interviews.

In total, **16 papers are included in this review, for which data was collected between the years 2006-2018**, therefore covering over a decade of recent consultation and attitudinal work. The studies were identified via an (unsystematic) online search, with only papers relevant to public attitudes towards the sharing and use of government-held administrative data for research being included in the review. A number of studies included cover attitudes to data use more generally, not just in relation to research, but are nevertheless relevant to the aims of this review. Literature that was not considered relevant and therefore not reviewed included papers focusing solely on commercial access to public sector data; papers concerned primarily with the linking of public sector data to private sector data; and papers focused more broadly on exploring the public's knowledge of, but not attitudes towards, the collection and storage (and not necessarily use) of data.

Included within the scope of the review is the 2014 'Dialogue on Data: Exploring the public's views on using administrative data for research purposes', a public consultation conducted by Ipsos MORI on behalf of the ESRC with the principal aim of examining public understanding of administrative data and attitudes towards linking government records for research. Specifically, the purpose of this study was to inform the work of the Administrative Data Research Network (ADRN) – the predecessor to ADR UK. It subsequently went on to advise much of ADR UK's foundational structure and approach.

3. Existing public knowledge of administrative data research

In general, previous consultation and attitudinal work has found that existing public knowledge of the sharing and use of administrative data for research is low, and that this can have an impact on levels of support for the practice:

- The 'Dialogue on Data' found that participants had very low initial awareness and understanding of social research, with some finding aspects of the subject matter *"complicated and difficult throughout the discussions"* (Cameron et al. 2014, p.2). None of the participants had heard of administrative data research prior to taking part in the dialogue (Ibid., p.22). Some questioned the value of social research at the beginning of the dialogue, with research findings being compared to *"common sense"* and concerns being raised as to whether social research leads to social value (Ibid., pp.2-3, 14). Most, however, attached some value to social research, and as participants' knowledge increased over the course of the consultation, support for using administrative data for research also largely increased (Ibid., p.18, 57).
- Participants of the Dialogue also commonly assumed that data was already linked and shared across government (Ibid., p.14). Despite this, there was a feeling that government does not currently make efficient use of administrative data and is therefore not acting in a joined-up way (Ibid., p.24). There was also confusion between the use of data for research as opposed to operational purposes (Ibid., p.15), a finding which was reflected in research undertaken by the Office for National Statistics (ONS) over the period 2009-2013 exploring public attitudes to the collection and use of data for statistics and research (2014, p.3). The ONS findings indicated that nearly half of the public assume that government *"already routinely links data about the population from multiple sources in a central data store"* (Ibid.).
- A systematic review of studies investigating public responses to the sharing and linking of health data for research by Mhairi Aitken et al. (2016b, p.5) found that participants of several studies were reported as being *"surprised that data are not already more widely used"*. A number of the studies reviewed by the authors also reported that participants considered not using data for research to be wasteful and against the public interest (Ibid.). Again, the authors found that, in general, support for administrative data research increased after study participants were informed of the potential benefits and the safeguards in place (Ibid.).
- Work exploring public attitudes towards the Scottish Health Informatics Programme conducted by Aitken et al. (2016a, p.719) found that a lack of existing knowledge of research using health data *"can lead to low understanding and lack of trust"*. A study by Sara Davidson et al. exploring the acceptability of cross-sectoral data linkage amongst the Scottish public (2012, p.iv) reported that when participants were reminded of the safeguards in place to protect individuals' privacy, many felt immediately more comfortable with the idea.

- In a study conducted by NatCen Social Research (2018, p.3) on behalf of UCL exploring public understanding and perceptions of data linking (particularly between health examination survey data and administrative records), respondents were found to have varying initial understanding of data linking, with some having a little knowledge and others having a greater existing understanding. Nevertheless, participants' understanding again changed substantially throughout the course of discussions, growing for many though not all (Ibid.).

These findings suggest that when the public has a better understanding of the value of research, they are more supportive of the use of administrative data for that purpose. This, alongside a general expectation that data is already linked and shared across government, demonstrates the need for greater transparency and more effective communication of the use of administrative data for research and its benefits – a need which will be discussed in greater detail in Section 6.2. These communications should use accessible, plain English and focus on what appears to be considered the most essential condition of research using administrative data: public interest.

When the public has a better understanding of the value of research, they are more supportive of the use of administrative data for that purpose.

4. Public interest

The most prominent trend identified in the literature reviewed is the importance of public interest (also referred to as 'public good', 'public benefit' or 'social value') as the primary driver of support for any use of administrative data. Without public interest at the centre of the work, the perception is that the risks associated with using this type of data could not be justifiably outweighed by the potential benefits.

- The 'Dialogue on Data' found that the public are largely happy with administrative data linking for research, with social value being one of several key conditions underpinning support (Cameron et al. 2014, p.57). Participants argued that "*social research should always be undertaken to deliver an outcome that is of social value*", and criticised research that seemed to have no tangible "*point*" (Ibid., p.19). The use of administrative data by government to improve public services was acceptable to most, with some seeing it as an "*act of citizenship*" to give data to government to help with decision making (Ibid., pp.21-22). In general, participants thought data ought to be shared within government as long as they aren't "*doing anything wrong*" with it and it is used for projects that will have social benefits, particularly improving government spending and key services like health and education (Ibid., p.24; 42).
- With a specific focus on patient data, a systematic review and ethical enquiry into public views on the use of this data for research by Jessica Stockdale et al. (2019, p.1) identified a similarly widespread willingness to share patient data for research for the "*common good*". Participants of the studies examined were found to be willing to share electronic health records (EHRs) for secondary purposes when there was a "*common*", "*greater*" or "*public good*", and when there was "*social responsibility*", "*altruistic attitudes*" and "*giving something back*" (Ibid., p.8). This rested upon the understanding that the benefits of medical research using this data could include the improvement of healthcare services, or of diagnosis and treatment of diseases (Ibid., pp.8-9). Similar to the findings of the 'Dialogue on Data', the review found that study participants had a sense of obligation or duty to help bring about these benefits (Ibid., p.9).
- Aitken et al., in a discrete choice experiment examining public preferences regarding the linking of health data for research (2018, p.5), found the most common preference regarding the purpose of data linking to be, by far, "*Research using linked information should only be used if it will have general public benefits*". 57% of respondents selected this as their main preference (Ibid.). This reflects the findings of Aitken et al.'s review of previous literature of the same focus (2016b, p.4), which highlighted that health research – and research in general – has broadly been found to be considered by the public as in the public interest. The authors found public benefit – as well as public trust, which will be covered in Section 5 – to be the factors that all other concerns or interests in research using public health data rested upon, either explicitly or implicitly (Ibid., p.19).
- For participants of a public consultation by Ipsos MORI on behalf of the Medical Research Council regarding the use of personal health information in medical research (2007, p.49), the advantages of medical research were considered to far outweigh the disadvantages. Only 6% said they feel the disadvantages outweigh the advantages, whereas 70% said they

feel the opposite (Ibid.). 60% of participants agreed that they have a responsibility as beneficiaries of medical research to allow their personal health information to be used for approved research projects, as long as their consent is given (Ibid., p.9).

- The 2015 Northern Ireland Life and Times (NILT) Survey, which included a series of questions relating to data sharing and privacy, found that 85% of respondents 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”* (Robinson et al. 2018, p.22).

In terms of what ‘public interest’ is conceived as, and what the specific benefits of data collection and use are perceived to be:

- Aitken et al.’s review (2016b, p.4) – which looked specifically at literature relating to the use of health data – found the benefits of data use to relate to the discovery of new cures and treatments, and the improvement of healthcare services in general. This was reflected in a study exploring attitudes of the Scottish public towards data sharing between the public, private and third sectors for research by Davidson et al. (2013, p.8), which found that participants tended to conceive public benefit in terms of improvement to public services or public health. It was considered to be on a collective rather than individual level, with the specific sub-sector of the population the data relates to being considered the group that should benefit the most from the research (Ibid., p.9-10).
- For participants of a citizens’ jury conducted by Mary P. Tully et al. exploring *“informed citizens”* attitudes towards different uses of health data (2019, p.11), public benefit in relation to the use of health data was seen as *“improving care and saving lives”*. *“Providing societal benefits through better public services”*, *“delivering improved outcomes for communities”*, and *“enabling research”* were cited as more specific benefits (Ibid.).
- A public consultation investigating attitudes towards the linking and use of personal data by the Wellcome Trust (2013, p.3) found that the benefits of data use more generally (for both personal and de-identified data) are considered to be: helping the government plan effectively; preventing and detecting crime; providing insights into population trends and statistics; tackling dishonestly such as fraud; making essential individual medical information available in an emergency; and, in terms of commercial uses, providing tailored marketing and convenience when shopping.

The public have been found to consider financial profit an unacceptable motive for the use of administrative data.

The public have been also been found to consider financial profit an unacceptable motive for the use of administrative data:

- For participants of Tully et al.’s citizens’ jury (2019, p.11), prioritising profit was considered unacceptable regardless of governance arrangements. The same sentiment was broadly reflected across the studies reviewed by Davidson et al. (2013, p.8). A study by Marion Oswald investigating attitudes to sharing personal data with the public sector (2014,

p.268) found that the majority of participants would be comfortable with their data being used to improve public services, but that only around a quarter were comfortable with it being used to make financial profit to *fund* public services.

Previous literature has also identified concerns that some research using administrative data could inadvertently work against the public interest and negatively impact society, by leading to findings that lead to particular communities or sub-sectors of the public being profiled and experiencing discrimination as a result:

- The 'Dialogue on Data' identified a feeling that local areas, for example, may be profiled, affecting the local population's ability to secure insurance or a mortgage (Cameron et al. 2014, p.34). Aitken et al.'s review of previous attitudinal work (2016b, p.20) meanwhile identified concerns that individuals or groups in society may be labelled as a result of research using linked data, and that policy based on the analysis of large datasets may be designed "*for the masses*", and not sufficiently take individual needs into account. Davidson et al. (2012, p.iv) found the public to be concerned about discrimination in multiple spheres, as a result of labels potentially being carried "*across sectoral boundaries*" when data is linked.

These findings show that public interest is paramount to public support for research using administrative data. The public conceives the benefits of using administrative data for research in terms of developing knowledge and improving our understanding of society – to improve public policy and services – rather than for purposes directly linked to financial profit.

It is worth bearing in mind, however, that previous literature has found no widely understood definition of 'public interest' amongst the public. It could mean different things for different individuals, and perhaps what matters more than defining the term is that the public perceives there are benefits of some sort. **Understanding what the benefits of administrative data research are considered to be by the communities the work aims to benefit therefore remains an important goal of public engagement.**

4.1. Data types

Despite the general trend towards support for research using administrative data in cases where the public interest can be sufficiently demonstrated, there are some differences in the perceived sensitivity of, and potential benefit from, using different types of data for research.

- During the 'Dialogue on Data', some participants expressed a feeling that some types of data were too sensitive and personal to be shared outside of the agency that collected it (Cameron et al. 2014, p.24). This included, for example records of domestic violence and data relating to HIV status, and was a feeling driven by concern about the potential consequences of the data "*getting into the wrong hands*" (Ibid.). Nevertheless, by the end of the Dialogue the researchers found that, overall, "*once there are strong researcher approval and security processes in place, [participants] were happy for data linking to go ahead using all types of administrative data*" (Ibid., p.40).

- Aitken et al.'s discrete choice experiment investigating public preferences regarding the linking of data for health research (2018, p.6) found that the type of data being linked was the single most influential factor shaping the preferences of participants. How profits are managed and shared was found to be the second most influential factor, with the purpose of the research coming out third (Ibid.). The findings also suggested that participants were more comfortable with different types of health data – for example, information from GP records and other NHS health records – being linked together than they were about health data being linked to other types of data, such as employment and benefits records (Ibid., pp.5, 7).
- Wellcome's attitudinal work (2013, p.11) found that many participants regarded *personal* health data (as opposed to de-identified health data) differently to other types of data. This primarily rested upon a perceived "*unquestionable benefit to people*" of experts having access to this type of information, especially in relation to illness.

These findings show that the type of data being linked is important to the public, and that personal – as opposed to de-identified, population-level – health data is considered to be particularly sensitive, as well as especially useful. This highlights the need for transparency around which data is being used and for what purpose, as will be discussed further in Section 6.2.

4.2. Demographic differences

Previous studies have also found some – albeit not dramatic – demographic differences in the level of support for administrative data research amongst the public.

- Aitken et al.'s discrete choice experiment (2018, p.6) found older age groups to be more likely to agree that "*research using linked information should not be allowed under any circumstances*", which was included as an option for each question. Those who selected this option at any point in the experiment were screened out; of those who were screened out, 42% were aged 55 and over, 34% were 35 to 54 and 24% were aged between 18 and 34, suggesting that older age groups are more concerned about the risks – or not as convinced of the benefits – of the use of linked data for research (Ibid.). This was reflected in the Wellcome Trust's research (2013, p.2), which found that younger people were more likely to be accepting of data collection and use than older people.
- The most notable demographic differences identified in Aitken et al.'s study (2018, p.8), however, were according to employment and working sector, with participants not in full-time employment being found to be more concerned with 'oversight arrangements' and the type of data being linked than those in full-time employment. Those working full-time were more concerned with the purpose of data linking, who the researchers were and how the profits arising from data linking would be managed (Ibid.).
- The Wellcome Trust (2013, p.2) reported that participants from socio-economic group C2DE (i.e. those in skilled, semi-skilled and unskilled manual job roles, and those on low or

no income⁴) felt more powerless to deal with the consequences of a data breach than those from socio-economic group ABC1 (i.e. those in managerial, administrative and professional, and supervisory and clerical job roles). In addition, participants from group ABC1 were found to be more likely than those from group C2DE to view health data as being potentially beneficial to society in the field of research, as well as in disease prevention, service planning and crime prevention (Ibid., p.4).

- The studies included in Stockdale et al.'s review (2019, p.11), however, represented conflicting findings in terms of differences of opinion across age groups and levels of education. For example, they found evidence of both younger and older age groups being in favour of sharing their data, as well as evidence of people with lower levels of education being both more and less likely to agree to sharing their data without consent (Ibid.).

These findings show that, in general, age and socio-economic background may have an effect on public support for the use of administrative data for research. Some areas of research may therefore be more acceptable than others to the communities whose lives they aim to benefit, depending on the group's demographic characteristics. It is thus important to involve the communities most relevant to specific areas of research in public engagement activities, so the views of those most affected by the work are sufficiently understood.

It is important to involve the communities most relevant to specific areas of research in public engagement activities.

⁴ Based upon the National Readership Survey Social Grade: <http://www.nrs.co.uk/nrs-print/lifestyle-and-classification-data/social-grade/>.

5. Privacy and security

Another key condition of public support for the sharing and use of administrative data identified in previous literature is safeguards to protect the privacy of data subjects and prevent data from being misused. Alongside the need for a tangible public interest, participants of previous studies have been found to be largely comfortable with data sharing and use as long as the necessary protections are in place to reduce to risk of reidentification and misuse. The main public concerns identified across the literature in relation to the privacy and security of data can be broken down into three main factors: de-identification; data access and security; and governance and regulation.

5.1. De-identification and anonymisation

The safeguard which stands out most prominently in the findings of previous research is de-identification, or anonymisation. De-identification appears to be the absolute minimum standard expected for the use of administrative data in research to be acceptable. It is important to note that both 'de-identified' and 'anonymised' data are referred to across the literature reviewed, though it is not always clear how 'anonymous data', as opposed to 'de-identified data', was understood by participants. When the terms are used in the context of a previous study's findings, their precise definitions should therefore be considered with some caution.

- In the 'Dialogue on Data', the de-identification of data was found to be one of three core factors in participants' support for administrative data linking (alongside "*the data is kept secure at all times*" and "*the data is linked for socially beneficial purposes*") (Cameron et al. 2014, p.4). Most participants – though not all – no longer considered de-identified data as personal, and therefore had no concerns around the use of such data, including being notified about its use (or asked for consent) (Ibid., p.33). This finding is reflected in work by Oswald (2014, p.265), ONS (2014, p.3), Davidson et al. (2013, p.8) and Aitken et al. (2016b, p.5), all of whom found that participants were significantly more comfortable with their data being collected, stored and used when anonymised.
- A concern raised by some participants of the 'Dialogue on Data' was whether de-identification was enough to stop people being identified. Though most generally understood that it is impossible to wholly guarantee data cannot be re-identified and were of the view the safeguards in place for the ADRN were enough to sufficiently limit the risk, some remained concerned about the de-identification process (Cameron et al. 2014, p.4; 30). They thought it may be possible to identify individuals if linked data, for example, included information that was unusual and might only apply to a small number of people (Ibid., p.30). Similar concerns around the potential for re-identification were found by Davidson et al. (2012), NatCen (2018) and Stockdale et al. (2019).
- For the 85% of respondents of the 2015 NILT Survey who agreed that data should be used where there is a benefit to society, this was under the condition that anonymisation and

Most participants of the 'Dialogue on Data' no longer considered de-identified data as personal.

the maintenance of personal privacy were assured (Robinson et al. 2018, p.22). Ipsos MORI's consultation (2007, p.9) found that 62% of respondents would be "*certain or more likely*" to provide their health information if there were assurances that it would be kept confidential. Meanwhile, work by the Royal Statistical Society (RSS) exploring public trust in data handling and attitudes towards data linking and privacy (2014, p.7) found that 55% of respondents supported data sharing when there was an assurance of anonymity, compared to 33% when safeguards were not guaranteed.

- In contrast to other studies, Oswald (2014, p.266) found that less than 40% of respondents were comfortable with the sharing of data, even when anonymised, though it should be noted that this was specific to medical and locational data. Oswald's work, which focused on sharing personal data with the public sector, also found that participants were less comfortable with their data being shared with other bodies than they were with it being collected, stored and used more generally by a single body, even when anonymised (Ibid.).
- In Wellcome's attitudinal work (2013, p.3), participants were found to have a strong sense of personal health data – as opposed to de-identified, population-level health data – as "*confidential, private and sensitive*", and there was a feeling that it should not be shared outside of "*secure, authorised bodies such as the NHS*". Population-level data, on the other hand, was regarded as anonymous, of benefit to all, and reassuring that it would be collected for the common good (Ibid., p.6). If data is anonymous and used and held within its respective research or health environment, participants had little or no issue (Ibid., p.6).

These findings show that the public distinguishes between personal data and de-identified (or anonymised), population-level data, with the sharing and use of the latter for research broadly being more acceptable than that of the former. There appear to be, however, some differences across the studies reviewed in terms of the proportion of respondents comfortable with their data

Most respondents were supportive of data sharing when de-identification or anonymisation was guaranteed.

being shared even when de-identified, though it is important to note that the methods of consultation and types of data sharing being explored differ across the studies, which may account for some of the differences in findings. Nevertheless, for most of the studies reviewed, most respondents were supportive of data sharing when de-identification or anonymisation was guaranteed.

5.2. Data access and security

Another key area of concern among study participants is around data falling into the 'wrong hands' – whether de-identified or not – and being misused to the detriment of data subjects and society. Data access and security have been identified as critical to addressing these concerns.

- The 'Dialogue on Data' found that participants were concerned about data being leaked, lost, or sold by organisations for profit (Cameron et al. 2014, p.3). In the context of ADRN, participants were initially concerned about who could gain access to data and whether it could be removed from the safe setting (Ibid., p.31). They were reassured on learning that

there were so many barriers to data sharing in place, and of the restrictions on access to data sourced by ADRN (Ibid., p.3).

- Participants of the Dialogue were strongly in favour of secure physical settings and concerned about remote access to a secure environment (Ibid., pp.3-4, 40). The authors stress, however, that the concept of remote access is difficult to explain and was *“perhaps not explained consistently across the workshops”* (Ibid., p.46). They found that *“the key idea that the data doesn’t leave the physical setting even when the researcher is working on it from a remote setting didn’t make intuitive sense to most”* (Ibid.). Participants did not fully understand how it would work and worried that, when data was being passed from one source to another over the internet, it might be at greater risk of being breached (Ibid.). The authors stress that *“further work would need to be done in understanding how best to explain the concept of secure remote access to the general public to allow the public to give a more informed opinion”* (Ibid.). Meanwhile, those who generally thought de-identified data is very low risk were more comfortable with remote access if the necessary protections were in place, such as limited passwords and the logging of all actions (Ibid., p.47). Some expressed that remote access to a safe setting would be necessary to avoid capacity issues at centres and make more efficient use of the money spent developing them (Ibid.).
- 78% of respondents of the 2015 NILT Survey said that researchers should have to go to a dedicated secure data centre to access data and conduct their analysis (Robinson et al. 2018, p.18). However, the option for researchers to access data via a secure remote connection to such a centre does not appear to have been included in the survey.
- Participants of the ‘Dialogue on Data’ felt reassured that there were no plans for a so-called ‘super database’ under ADRN, containing multiple linked data sources (Cameron et al. 2014, p.30). However, this appears to have been a spontaneous consideration of participants rather than something presented to them in detail, and the authors do not explain what such a database was understood to be. Participants were not given detail as to what such a database might look like in terms of security and data access restrictions, as this was not an approach proposed by ADRN. The response of one participant suggests that a ‘super database’ was conceived as a service offering open access to data, rather than access to only approved researchers in a secure physical facility – or via approved connection to one – with outputs checked: *“Everyone’s information is going to be centralised. How can they guarantee everyone’s motives? You always see on the news the concerns about security. We see the business with MI5 and MI6. Where are the guarantees?”* (Ibid.).
- Stockdale et al. (2019, p.9) found that participants were concerned that sharing their electronic health records (EHRs) may lead to them being leaked, lost or subject to unauthorised access and used to their disadvantage, whilst Wellcome (2013, p.3) found the same for the sharing and linking of personal data more generally.
- Amongst participants of Davidson et al.’s 2012 and 2013 consultations exploring attitudes towards cross-sectoral data sharing, these concerns were echoed, with participants of the 2012 study (p.iv) being concerned that data linking in itself would increase the likelihood of security breaches as hackers would be able to obtain a large amount of information *“in one hit”*.

- For participants of Ipsos MORI's consultation (2007, p.8), 13% of participants were anxious about a loss of control and personal health information "*falling into the wrong hands*". The RSS (2014, p.4) found that the addition of safeguards to protect data improves public support for data use and sharing from 33% to 51%.

These findings show that public concern for the security of data goes beyond de-identification alone, and that stringent data access restrictions are also expected to prevent data from being accessed by unauthorised persons and potentially misused. Access being limited to a physical safe setting in which researchers can be closely monitored appears to be of particular concern, with a feeling amongst study participants that de-identification alone may not be enough to protect data from re-identification and misuse.

5.3. Governance and regulation

In addition to the physical security of data and controls around access, participants of previous consultations and attitudinal work have also been found to value protections in the form of governance and ethical frameworks to regulate how data is used and shared.

- In their reviews of previous literature, both Aitken et al. (2016b, p.5) and Stockdale et al. (2019, p.11) identified an increase in public acceptance after study participants were informed about governance mechanisms. Stockdale et al. (2019, p.11) also found that previous studies have reported support for penalties for the misuse of data, with participants valuing laws to regulate access to and use of data, and for regulators and those accessing data to be held to high standards.
- Davidson et al.'s 2012 research (2012, p.iv) identified concerns about who would oversee the operation of data sharing frameworks and where overall accountability would lie if linked data were lost or stolen. Ipsos MORI's public consultation (2007, p.7) revealed a feeling that there was a need for an independent organisation to act as a "*buffer*" between researchers and the public to prevent misuse of information. Meanwhile, for participants of the 'Dialogue on Data', a common reassurance of the ADRN was that it would provide a systematic way to regulate administrative data linking (Cameron et al. 2014, p.3).
- However, in Aitken et al.'s exploration of attitudes towards the use of health data (2016a, p.720), participants expressed skepticism towards existing governance systems. They were concerned that committees of oversight bodies would by default operate in favour of sharing data for research (Ibid.). Davidson et al. (2013, p.10) identified strong support for the interests of the public being represented in oversight bodies. It was felt that this representation should happen in an indirect way, for example via a third party organisation that would regularly consult with the public and feed back to the oversight body (Ibid.).

Governance frameworks, in addition to the physical security of data, are clearly highly valued, but governance in and of itself is not enough. How that governance operates and who makes up oversight bodies is crucial to preventing bias and the misrepresentation of public interests.

Nevertheless, it is clear from previous literature that safeguards alone are not enough; it is essential those handling and using data are trusted, and that the safeguards in place are communicated effectively so the public is aware their data is in safe hands.

6. Trust and transparency

Alongside public interest and privacy and security, previous studies have found that trust and transparency are also key conditions for public support for the sharing and linking of administrative data for research. Individuals and institutions sharing, accessing and using data must be trusted to keep data secure and use it only for research in the public interest. Meanwhile, transparency is important for keeping the public informed about what their data is being used for and how it is kept safe and secure. Previous research shows that the specifics of projects using

If those handling and using data are transparent about how and why it is being used, they are likely to receive a greater level of public trust.

administrative data hugely affect the level of public support, and the public therefore expects a sufficient level of transparency to allow them to remain informed about how their data is being used.

Trust and transparency are closely linked: in their focus groups exploring members of the Scottish public's attitudes towards the Scottish Health Informatics Programme (SHIP), Aitken et al. (2016a, p.721) found that transparency played an important

role in the level of trust participants had in research institutions or data controllers, whether public or private. In the stakeholder workshops – for example with researchers and analysts – the authors also found that transparency and public engagement were expressed in terms of their positive impact upon trust (Ibid.). If those handling and using data are transparent about how and why it is being used, in an uncensored and accessible way, they are likely to receive a greater level of public trust.

6.1. Trust

Previous literature has found trust to be essential to securing public support for research using administrative data. Without trust, the public cannot be reassured that the use of their data is in their interest and the interests of society at large, or that their data is being held safely and securely and protected from misuse. Assurances of data security and good intentions mean little if the individuals or institutions using the data are not trusted to use data ethically and responsibly.

The studies reviewed found clear differences in the level of trust attributed to different types of institutions for various reasons, providing indications as to how an organisation or individual might be able to build trust.

- Participants of the 'Dialogue on Data' expressed dislike for the idea of commercial access to data for solely commercial benefit (Cameron et al. 2014, p.14; 42). They worried about government data getting into the hands of commercial companies due to low trust in government more generally (Ibid., p.21). Similar sentiments were identified amongst respondents of the Wellcome Trust's attitudinal work (2013, p.9), which found there was some cynicism in relation to the notion of the government linking data and fears about the government "*taking something away from people*". For participants of the 'Dialogue on Data' who were more trusting of government in general, however, there was a feeling that government use of data was benign and in the public interest (Cameron et al. 2014, p.23).

- Aitken et al. (2016a, p.720) found that whether the public trusts or mistrusts research is largely conditional and variable, and characterised by ambivalence. Although some groups, such as healthcare providers, were generally found to be trusted more than others, this did not mean that research conducted by these groups was automatically supported, or that research conducted by those who are typically less trusted was automatically opposed (Ibid.). Factors affecting the level of trust in an institution or individual to use health data were found to include: the data sharing processes in place; the level of transparency around the work; and the existence of accountability procedures (Ibid.).
- Aitken et al.'s literature review (2016b, p.17) found that previous studies relating to the sharing and use of health data have similarly indicated generally higher levels of trust in the public sector compared to the private sector, again due to a perception that the former has greater accountability and data protection mechanisms.
- Participants of Davidson et al.'s workshops with members of the Scottish public (2013, p.8) expressed a feeling that *who* was accessing data and for what purpose is of greater concern than the type of data being accessed. Participants demonstrated “*near universal acceptance*” of public bodies – including the Scottish government, the NHS, local authorities and the police – accessing anonymised data from other organisations for research (Ibid.). Again, this was a feeling apparently driven by a perception that public sector organisations are dedicated to delivering public benefits and have more stringent data protection measures in place than other types of organisation (Ibid.). This was also reflected in Stockdale et al.'s review of public views on the use of patient data for research (2019, p.10).
- For participants of NatCen's attitudinal work (2018, p.10), it was felt that the government collecting data in the form of the Census, and health data collected by the NHS, were important for planning for the future. On the other hand, it was felt that commercial companies would only want to access data for commercial gain (Ibid., p.11).
- The RSS (2014, p.1) identified what it coins a “*data trust deficit*”, in which the public trusts institutions to use data appropriately less than they trust them in more general terms. For example, 13% of respondents were found to have low trust in the NHS generally, whilst 17% have low trust in it using data appropriately (Ibid.).

Not all public bodies receive the same levels of trust, however, and some institutions are trusted more than others. In general, the NHS has been found to be associated with higher levels of public trust than most other public bodies:

Not all public bodies receive the same levels of trust, and some institutions are trusted more than others.

- For participants of the 'Dialogue on Data', the police were less trusted to keep data safe as they were perceived to be “*prone to abuse their authority*”, whilst participants of Oswald's 2014 survey were found to be more comfortable with the NHS handling their data than local or central government bodies (Cameron et al. 2014, p.22; Oswald 2014, p.266).
- The 2015 NILT Survey found that 91% of respondents trusted their GP surgery to keep information secure; 86% trusted the NHS more broadly to do so; and 73% trusted

government departments (Robinson et al. 2018, p.8). Of participants in Ipsos MORI's workshops (2007, p.46), 87% trusted GPs to have access to their personal health information and 59% trusted other healthcare professionals such as consultants or hospital doctors. The NHS in general was trusted by 51% (Ibid.).

- The RSS (2014, p.2) also identified greater levels of trust in the NHS, though they reported comparatively lower levels of trust in general than other studies: 36% of respondents trusted the NHS to use data appropriately; 25% trusted academic researchers and universities; and 13% trusted the government.
- Aitken et al. (2016a, p.717) suggests the greater general level of trust in healthcare providers compared to other public bodies may be due to individuals' existing relationships with primary healthcare providers, which are in some cases built up over many years, suggesting that having a familiar relationship with an individual is important for securing trust. For participants of Davidson et al.'s study (2012, p.iii), trust in the NHS was expressed in terms of a perception that health professionals generally serve to help the public and are expected to abide by a moral code of conduct, supposedly more so than other public bodies.

The public has not been found to be wholly opposed to commercial access to government-held administrative data, with many study participants identifying potential benefits to wider society of private companies having access to data in certain circumstances. In cases where these companies are granted access, the literature has identified a preference for greater controls than might be expected for public bodies accessing data.

Many study participants identified potential benefits to wider society of private companies having access to data in certain circumstances.

- In their focus groups, Aitken et al. (2016a, p.718) found that, while concerns certainly existed, the relevance of pharmaceutical company access or involvement in research was acknowledged. In cases where private companies were gaining access to administrative data, confidentiality – and especially anonymisation – were considered to be of particular importance (Ibid.).
- For participants of Tully et al.'s citizens' jury on appropriate uses of health data (2019, p.10), commercial use of data was not universally vetoed, and an assurance of public benefit was expressed as the main condition for access as with other data uses. Davidson et al. (2013, p.10) meanwhile identified an interest amongst study participants in controls to ensure that any profits from private companies using data were distributed in the public interest, for example by reinvesting in local communities.

In terms of trust in researchers:

- Participants of the 'Dialogue on Data' held the view that researchers who gain access to data should be "neutral", "unbiased" and "qualified", whilst researchers working for private companies, as well as private individuals, should not be granted access to linked data (Cameron et al. 2014, p.43). For some participants, a lack of familiarity with research was

associated with a basic lack of trust in researchers, whilst others with existing knowledge of how research works tended to have greater trust and fewer concerns in the use of data for research (Ibid., p.36-38).

- For participants of Aitken et al.'s discrete choice experiment (2018, p.5), respondents were found to feel most comfortable with university researchers, NHS staff or government researchers gaining access to data. Aitken et al.'s focus groups with members of the Scottish public (2016a, p.718) reported a preference for research conducted by academic as opposed to other types of researcher. For participants of the latter, an important factor influencing positive perceptions of academic researchers was found to be a view that they were more altruistic – or selfless – than other types of researchers, and less likely to be motivated by profit (Ibid.).

These findings suggest that, to develop and maintain trust, an individual or organisation must demonstrate a dedication to the public interest and possess the necessary mechanisms to maintain the privacy of data subjects and protect data from misuse. These appear to be the main drivers of greater trust in government and public bodies as opposed to private, profit-making organisations. However, the findings also show that trust is not straightforward, and that lower trust does not automatically mean that access to data is unacceptable; in some cases, lower trust may simply mean a need for greater data protections and assurances of public interest.

Nevertheless, fulfilling these expectations alone is not enough; effective communications to ensure that the public is aware of them, and feel appropriately informed about how their data is being used, are essential. This is where transparency comes in.

6.2. Transparency

Transparency around how administrative data is used and shared has been a key concern for participants of previous consultations and attitudinal work. Transparency has been associated with efficient public communications around how administrative data is held, used and shared.

- For participants of the 'Dialogue on Data', concern was expressed that the general public would not understand the process and purpose of administrative data research (Cameron et al. 2014, p.5). Nevertheless, several key messages were seen as important to communicate to the public in relation to the ADRN, including that: the data is anonymised; there are safeguards in place to prevent the loss of data; research using the data is socially beneficial; and the ADRN will make existing frameworks for sharing data for research more efficient (Ibid., p.51-52). Furthermore, whilst participants understood that their consent would not be asked for within the scope of the ADRN – for the same reasons as is the case for ADR UK under the terms of the 2017 Digital Economy Act – they also wanted to know that, in return for data about them being used, they would be able to access the findings of research conducted using it (Ibid., p.36).
- ONS (2014, p.4) also found that public acceptability of data use is improved by appropriate communications around what the data is being used for and why, whilst the RSS (2014, p.3) found that only 5% of participants in their research disagreed that they would “*really like to know what information government knows about me*” more generally.

- In Stockdale et al.'s review of previous literature (2019, p.11), the authors identified a feeling that better transparency around safeguarding processes, data sharing arrangements, research findings and data access regulations is needed. Study participants wanted a better understanding about the research being done using their health data and why, to which better public communications are key (Ibid.). Tully et al.'s citizens' juries (2019, p.11) suggested that the public benefit of health data use by commercial organisations must be made explicit, or the public will not find it acceptable.
- Aitken et al., in their focus groups with members of the Scottish public (2016a, pp.718-719), similarly found that participants felt it was important for researchers and data controllers to communicate the positive aspects of the use of data for research, with concerns about data linking stemming directly from a perceived lack of openness about how data is used and collected. Participants went so far as to suggest that a lack of openness may be a deliberate effort to "*withhold information from the public*" (Ibid., p.719). In Aitken et al.'s review of other literature (2016b, p.20), almost all of the studies reviewed reported that participants expressed a desire for more information and greater transparency about how and why health data is used for research, and the safeguards in place. The authors stress that this indicates that greater "*awareness raising*" is needed, in which information is given, but also in which public interests, concerns and uncertainties can be raised and addressed (Ibid.).
- Participants of NatCen's study (2018, p.20) wanted to see clear details of how, why and which research projects would be using their data, with a feeling that "*the more transparency the better*", to reassure individuals that the work being done was ethical and that data was being kept secure. Of the total participants – all of whom had taken part in the 2016 Health Survey for England – those who had previously disagreed to have their survey data linked to other forms of data recalled that the main reason for this was a lack of comprehensive information about the ways in which the data might be linked (Ibid., p.15). Ipsos MORI's 2007 work (p.9) similarly suggests that better information about the purposes of medical research might be the key factor in making the public more inclined to allow their personal health information to be used for research.

These findings suggest that transparency has a direct impact on the level of public support for data sharing initiatives, with a lack of transparency going so far as to cause individuals to decline to share their data in instances where consent is sought. Whether or not the highest levels of data

Transparency has a direct impact on the level of public support for data sharing initiatives.

security are in place, and a public interest exists, without the relevant information being publicly available such an initiative may experience a lack of public support.

In conclusion, trust and transparency together form another key condition of public support for research using administrative data, with transparency having a significant impact on the level of trust in individuals or institutions handling and using data. Nevertheless, trust and transparency

alone are not enough to ensure public support for data sharing initiatives, and must be maintained in balance with public interest and safeguards to protect the privacy and security of data.

7. Striking the balance

Previous consultation and attitudinal work shows that the public is broadly supportive of the sharing and linking of administrative data for research, so long as there are minimum standards around three key conditions: public interest, privacy and security, and trust and transparency. No study has identified any sole assurance that is enough to secure public support for research using administrative data; rather, all have highlighted that support cannot be guaranteed without a minimum standard of all of these conditions.

- In the 'Dialogue on Data', a perceived need for a tangible 'social value' did not sit in isolation as a condition of public support; "*data is de-identified*", "*data is kept secure*" and "*businesses are not able to access the data for profit*" were also identified as necessary (Cameron et al., p.57). The authors stress that participants were supportive of data linking if the main goal was to find out more about society or help government plan or carry out services, "*provided each project obtains methodological and ethical approval*" (Ibid., p.41).
- The 2015 NILT Survey found that public support for data sharing rests upon "*three pillars*": trust in organisations, data protection measures and public benefit (Robinson et al. 2018, p.25). If any of these pillars are reduced or removed, public support falls (Ibid.).
- Stockdale et al.'s review of previous literature (2019, p.1) found that, for the use of health data, while there was general willingness to share patient data for research in the public interest, "*this very rarely led to unqualified support*". Rather, it rested upon two key concerns about potential risks to privacy: data security and the motivations for using the data (Ibid.).

However, public support is not straightforward. No project using administrative data is the same, and the literature shows the specifics of any given project have an impact on public expectations

The potential benefits of research using administrative data must outweigh the risks.

of the measures needed to protect data. For example, some projects may require the assurance of greater safeguards than others if aspects of their approach are considered less robust, even once a minimum standard of each of the necessary conditions is achieved.

This may particularly be the case for projects with a less substantial public interest, or when those accessing data are less trusted. Ultimately, the potential benefits of using data must outweigh the risks to privacy and the possibility of misuse, and both a minimum standard and appropriate balance of the three key conditions – public interest, privacy and security, and trust and transparency – must be struck to achieve this.

- Participants of previous studies have expressed that they do not expect even the highest level of data protection to be entirely "*foolproof*" (Cameron et al. 2014; Davidson et al. 2012; Aitken et al. 2016b). There is an understanding that there is always some level of risk associated with the use of data – usually perceived in terms of the risk to the privacy of data subjects and the potential for misuse of data (see Section 5) – but the public is nevertheless comfortable so long as this is outweighed by the potential benefits.

- Aitken et al.'s review of previous research (2016b, p.14) found that study participants regarded breaches of security as *"always being possible"*, but that the risks were also regarded as *"tolerable"* when the purpose and potential benefits were sufficiently valued. Privacy concerns were balanced with recognition of the potential benefits of data sharing, with some study participants prioritising benefits over privacy (Ibid., p.15). A number of studies reviewed by the authors identified a feeling that the public do not want individual control over their data to hinder the benefits of research, but also wanted reassurance that public benefits and privacy would be prioritised over profits (Ibid., p.15; 17).
- The RSS (2014, p.4) identified more support overall for data sharing within government – with varying safeguards – when there are tangible benefits, compared to not sharing data at all due to privacy concerns. However, 35% of respondents disagreed with the statement: *"Once my data has been anonymised and there is no way I can be identified, I'm not really bothered how it is used"*, showing that a notable proportion continued to care about other aspects of how their data was used, even when anonymised (Ibid, p.3). ONS's work (2014, p.4) similarly found public views towards the use of data for research to differ according to who is using the data and for what purpose.
- Ipsos MORI's consultation (2007, p.7) identified a tension between the *"greater good"* and individual privacy. Although most participants saw the benefit of personal health information being used for research, they also had concerns about privacy (Ibid.). 69% of participants said they were *"likely"* to allow their personal health information to be used for research – suggesting further information about the specifics of the research may affect their decision – compared to only 14% who were *"certain"* to (Ibid., p.8).
- Oswald et al. (2014, p.270) found that there is no straightforward answer to whether the necessity of public services results in a readiness to *"share personal data and sacrifice a certain level of privacy"*. Rather, the *"benefits-versus-costs problem"* is significant, and the more tangible the benefit, the greater the comfort in data sharing (Ibid., pp.270-271).
- Aitken et al.'s review of previous studies (2016b, p.20) found that private sector access to data was not wholly opposed, and that the involvement of private companies was conditional upon whether the work was in the public interest. The authors found that in some studies there was a link between trust and the desired level of individual control over data, with lower trust being associated with greater favour for more stringent controls (Ibid., p.15). Aitken et al.'s focus groups (2016a, p.718) found that personal privacy in general, and a guarantee of anonymisation, were considered to be of greater importance when records were being accessed by private companies as opposed to public bodies.

These findings show that public support for research using administrative data is complex, and that whilst the public can appreciate the benefits of this type of work, there is also concern over privacy. The level of trust in researchers or data handlers in particular appears to have an impact upon the expected level of public benefit and data protection. Demonstrating that the potential benefit outweighs the risk when using administrative data for research is therefore paramount.

8. Beyond transparency

In addition to exploring the attitudes and sentiments of the public towards administrative data research, previous literature has given important indications as to the type of engagement the public expect to have when administrative data is used for research.

In general, public engagement has been considered essential to work which aims to make use of government-held administrative data. In a recent report with recommendations to the HM Courts and Tribunals Service (HMCTS) data strategy, Dr Natalie Byrom (2019, p.6), Director of Research at the Legal Education Foundation, stressed the importance of dedicating resources to engaging with a range of stakeholders, including the public, and testing the acceptability of different models. This recommendation is specific to the strategy concerning court data but is equally relevant to strategies setting out the management of other types of administrative data. A recent joint report from think tank Policy Connect and the All Party Parliamentary Group on Data Analytics recommends that *“the public should be engaged through a wide variety of methods”* when initiatives aim to make use of public data, including via open consultation, public events and industry outreach (Tindale and Muirhead 2019, p.15). The report stresses that *“rules made with little or no public engagement have led to avoidable errors which could contribute to a public distrust in data use”* (Ibid.).

- Emily Rempela et al. (2018, p.573), in light of the findings of a literature review of public engagement in new technologies, stress that transparency alone is not enough, and *“merely facilitates the potential for successful engagement to take place”*. The authors suggest that public engagement in government data science initiatives should either take a ‘normative aim’, giving the public a role in technological development in the form of public consultation or workshops, for example, or a ‘substantive aim’, which would go so far as to involve members of the public in developing and designing new technology (Ibid.). This participatory approach to public engagement, the authors say, is more effective at having a meaningful impact on development than a one-way communications-based approach and one-off events (Ibid.). The authors also suggest that identifying subsets of the public with whom it is more relevant to engage is important, rather than focusing on engagement with *“vast and discrete potential publics”* (Ibid., p.574).
- Rempela et al. also state that, while previous widescale consultations such as the ‘Dialogue on Data’ represent a step toward a better understanding of public views on aspects of data science, they *“do not equate nor reflect public influence”* (Ibid., p.575). They find that these previous studies are *“not sufficient for public engagement nor do they overcome trustworthy practice in other areas. In fact these consultations evidence a lack of any kind of decentred governance”* (Ibid.). They propose that public participation must move beyond privacy and consent, that it must give the public a role in deciding *“what government should use data science for”*, allowing the public to take part in decision-making processes *“around how data is created, collected and utilised for the good of society”* (Ibid., p.576).
- Aitken et al.’s focus groups (2016a, p.719) found preference for an *“open exchange of information and greater equity in the science-public relationship”*, in which public engagement is an indicator of the trustworthiness of data users rather than a way in which to build

trust. The authors stress that, although transparency may improve the trustworthiness of data sharing initiatives, it must involve “*open communication of uncensored information*” rather than a more selective transparency (Ibid., p.721). Furthermore, trustworthiness is more likely to be achieved if public engagement involves open dialogue in which public concerns can be responded to, instead of a one-way dissemination of information (Ibid.). Finally, the authors argue: “*Public engagement should not be aimed at ‘improving’ public trust in science, but rather at improving the trustworthiness of science*” (Ibid.).

- Tully et al. (2019, p.10) found that citizens’ juries as a method of public engagement may prompt a more informed judgement from the public about the use of data for research than, for example, surveys. The authors argue that citizens’ juries symbolically represent the community and are a way of providing both information and an opportunity to participate (Ibid.). Nevertheless, they also stress that this form of public engagement is costly and thus best suited to situations where there is a specific need to involve the public in decision making around complex matters (Ibid.).

A minority of studies have, however, identified preference amongst the public for less active involvement in decision-making and a greater focus on transparency, feedback and consultation:

- Davidson et al. (2013, p.10) found that, amongst members of the Scottish public, although there was unanimous agreement that public involvement in decision-making on data sharing is important, there was also a feeling that members of the public did not have the requisite knowledge and expertise to be able to take part in more specific decisions about the research that should be carried out using data.
- Participants of the ‘Dialogue on Data’ had mixed views, with some saying that the public should be involved in decision-making and others expressing that transparency was enough and that proactive engagement was not needed (Cameron et al. 2014, p.5). This was driven by a similar perception to that held by participants of Davidson et al.’s study, namely that the subject matter is too complicated to achieve a sufficient level of understanding amongst the general public to prevent them from worrying unduly about privacy and security (Ibid., p.51).

Nevertheless, these findings largely suggest that future public engagement should move beyond transparency and a one-way dissemination of information, and past widescale, general consultation on the use of administrative data for research. Rather, it should build upon existing knowledge about public attitudes to administrative data research as set out in previous literature and focus on gaining public input – via open and meaningful dialogue – in specific areas of research.

Future public engagement should build upon existing knowledge about public attitudes and focus on gaining input in specific areas of research.

In the context of ADR UK, the ‘Dialogue on Data’ has laid the groundwork for understanding the public’s views towards administrative data research, and our approach must therefore now shift to focus on gaining public input in relation to the specifics of the research being done. ADR UK’s approach will therefore seek the public’s input on specific areas of research, as outlined in detail below.

9. The ADR UK approach

Previous literature has shed light upon the public's attitudes towards the use of administrative data for research. This review shows that the public are broadly in favour of administrative data research, as long as certain conditions are met.

Ultimately, the proposed benefit must outweigh the potential risk, and this is dependent upon the specifics of any given project, including: the data being used; the questions being asked; the protections in place; and the institutions or individuals accessing the data. These attitudes have been shown to have been held over time, in work done both over a decade ago and more recently, with this review covering studies conducted between 2006-2018. Nevertheless, it is important to note that the nature of research is such that it is not always possible to know if it will ultimately prove beneficial. The findings are not known at the start, and all we can aim for is intended benefit.

ADR UK's structure and approach ensures that each of the three conditions underpinning public support for research using administrative data as identified in this review are at the core of everything we do:

1. **Public interest**

All research which hopes to use data curated by ADR UK must submit a research proposal that demonstrates robust methodology and public interest. 'Public interest' is assessed as per the definition set out in the *Research Code of Practice and Accreditation Criteria* under Part 5 of the 2017 Digital Economy Act (DEA), which was established on the basis of public consultation. According to the Code, 'research in the public interest' is:

"...research whose primary purpose is, for example, to: i) provide an evidence base for public policy decision-making; ii) provide an evidence base for public service delivery; iii) provide an evidence base for decisions which are likely to significantly benefit the economy, society or quality of life of people in the UK, UK nationals or people born in the UK now living abroad; iv) replicate, validate, challenge or review existing research and proposed research publications, including official statistics; v) significantly extend understanding of social or economic trends or events by improving knowledge or challenging widely accepted analyses; and/or, vi) improve the quality, coverage or presentation of existing research, including official or National Statistics."⁵

ADR UK-curated data is only made available to approved researchers for projects which can demonstrate public interest in line with this definition. This review has found no reason to challenge or seek change to the implementation of this definition in the context of administrative data research. However, as discussed above, there is no widely understood definition of 'public interest' amongst the public. Understanding what the benefits of administrative data research are considered to be by the communities the work aims to benefit therefore remains an important goal of public engagement.

2. **Privacy and security**

Improving lives is at the core of ADR UK's mission and the secure handling of data underpins this. Data shared with researchers via ADR UK is de-identified: when accessible

⁵ Digital Economy Act 2017 Part 5: Codes of Practice, '[Research Codes of Practice and Accreditation Criteria](#)', Principle 4: Public interest (accessed 17/04/20)

to researchers, it does not include any personal identifiers, with any elements that could be traced directly back to individuals – such as names and addresses – having been removed. Only the sub-sets of de-identified datasets – the variables needed by researchers to answer their specific research questions – are made available for use.

ADR UK partners have rigorous measures in place to ensure data cannot be accessed by any unauthorised persons, and operate according to the ‘Five Safes’ – a set of established safeguards to ensure data is kept safe and secure: Safe data, Safe person, Safe project, Safe place, and Safe output⁶. Once researchers have been vetted and approved, they must access data via a secure physical research facility – or a secure connection to one – provided by an ADR UK partner. Researcher activity is closely monitored and outputs are checked before being released to ensure data has not been misused in any way. The data made available to researchers by ADR UK is therefore anonymous as per the definition set out by the ICO (see Section 1.2), as these safeguards mitigate the risk of re-identification.

3. Trust and transparency

ADR UK is a publicly-funded, apolitical body. The academic research that ADR UK facilitates is driven by a desire to understand and improve society, rather than to advance specific agendas. Researchers wishing to use data curated by ADR UK must go through a rigorous approval process. They must be an approved researcher with the necessary qualifications and expertise to carry out the proposed research, and must have a research project with robust methodology that has been approved as being in the public interest.

We are committed to transparency and effective communication, and strive to ensure the public is kept well informed about our work via up-to-date digital communications, public events and more. We work hard to do this in a way that is understandable and accessible to a variety of audiences, using text, video, infographics and more.

9.1. Public engagement with ADR UK

In light of the findings of this review, and due to the volume of existing literature and the consistent findings it has had, ADR UK will move beyond widespread, general consultation about the uses of administrative data for research. We will now hone in on the more specific work underway to build upon – and not repeat – existing work. This will bring to life the research that has as of yet mostly been shown to the public in an abstract sense, and seek their engagement with specific research programmes. This will involve engaging with the demographics and communities to whom each of our projects is relevant and whose lives are impacted by the work – both directly and via relevant community representatives, for example charities and community groups. This is also in line with UK Research & Innovation’s ‘Vision for Public Engagement’ (2019), which sets out ambitions to “*engage under-represented communities and places with research and innovation*”, “*actively involve a wide range of people in their work*” and “*listen to public concerns and aspirations*”. We will continue to monitor any changes to public attitudes identified by other, broader attitudinal work, and adapt our approach where necessary.

⁶ Office for National Statistics (2017), [‘The ‘Five Safes’ – Data Privacy at ONS’](#) (accessed 17/04/20)

Each ADR UK partner has its own strategy and approach for public engagement. The approach of the Strategic Hub and each of the devolved partnerships, including ADR Northern Ireland, ADR Scotland and ADR Wales, is set out below.

The Strategic Hub



The [ADR UK Strategic Hub](#)'s public engagement approach involves an individual focus on each of the specific [Hub-led projects](#), to explore and address the interests and concerns of the sub-sectors of society whose lives may be impacted by the specific work underway. Engagement for each individual Strategic Hub-led data linking project will therefore involve a three-stage process, as follows:

1. Stage 1: Initial engagement with third-sector organisations

The first stage will involve initial engagement with voluntary, community and social enterprises (VCSEs), NGOs and other organisations able to represent the interests of the demographics relevant to each project, to raise awareness of the project and gather initial feedback on the work proposed. For example, for a project using children's data, this might involve engaging with charities working with or for children, which thus have the knowledge and expertise to offer input on the major issues faced by children today.

2. Stage 2: Formal engagement with third-sector organisations

It will then move on to formal engagement with organisations identified in Stage 1, for example in the format of workshops or roundtable discussions, to gather detailed input on the proposed research questions and design, and flush out any overlooked issues.

3. Stage 3: Direct engagement with members of the public

If deemed appropriate and practical, and as advised by discussions in Stage 2, direct engagement with members of the public to whom the research is particularly relevant will then be sought, for instance in the form of focus groups. We would work closely with the organisations engaged with in Stage 2 to enable meaningful engagement with the groups in question. The purpose of this final stage would be to seek the input of the people whose data will be used and whose lives may be affected to hear about the issues important to them and how they feel their data could best be used.

This work will be carried out not to consult on whether research using administrative data *should* be done – as was the primary focus of the papers included in this review, including the 'Dialogue on Data' – but rather to guide *how* it is done, in terms of research focus and questions explored, with those most relevant to the research at the centre.

It is essential that any work facilitated by ADR UK is in the public interest. This engagement work will also allow us to explore data subjects' views on what constitutes 'research in the public interest' in the context of each data linkage. Furthermore, for engagement to be meaningful, it is important that there is some mechanism for the findings to be fed into the research process, for example by helping to shape the research questions being explored, or other aspects of the research design. Working closely with researchers to embed the findings of public engagement into their research is therefore also key to this work.

ADR Northern Ireland



The governance structure of [ADR Northern Ireland \(ADR NI\)](#) – a partnership between the Administrative Data Research Centre Northern Ireland (ADRC NI) and the Northern Ireland Statistics & Research Agency (NISRA) – includes the establishment of a Public Engagement and Impact Committee (PEI Committee). The Committee, with membership from a variety of colleagues from across ADR NI, provides high-level input on issues relating to public engagement and impact for work underway across the partnership.

The public engagement strategy for ADRC NI involves a variety of approaches to engage with the public and civic society including via dialogue, workshops and symposia. During the ADRC NI investment period, ADRC NI were trailblazers in engaging with local Voluntary, Community and Social Enterprise (VCSE) groups as different sites of knowledge production and consumption. VCSE groups are often experts in their areas, both geographically and thematically. As such, they can enrich research endeavours with grassroots knowledge and practice and enhance impact by using research both as evidence in their own advocacy work, and as evidence in campaigning with decision makers. As part of ADR UK, ADRC NI continues to embed its previously developed relationships and models of working with VCSE groups to develop and steer its research projects.

ADRC NI also replicates – with some modifications – a Data Workshop Series which began during the ADRC NI investment. These workshops focus on: raising awareness among VCSE groups of the power and potential of data in their own work; how complex questions can be answered using data; raising public acceptability and demand for sharing data for research; and establishing and embedding positive working relationships with the VCSE sector.

Furthermore, by bringing key stakeholders onto steering committees for each project, ADRC NI maximises engagement with people and organisations with differing expertise and knowledge of the issues researchers are exploring; and gains buy-in and ownership from people and organisations who can turn research findings into policy and societal change.

This model of co-production has allowed researchers to work closely with community organisations and policymakers to develop and design impactful research and knowledge exchange events. These targeted events bringing together NGOs, service providers, policymakers and elected representatives, provide a useful forum for discussion of research and impact beyond the academic context, and have been particularly useful in Northern Ireland where there was no devolved government between January 2017 and January 2020.

ADR Scotland



[ADR Scotland](#) – a partnership between Scottish Government and the Scottish Centre for Administrative Data Research (SCADR) – has a two-pronged public engagement strategy:

1. **Building and maintaining trust:** this is done via the partnership’s Public Panel, as well as via broader engagement events such as public seminars on data sharing and linking, and at festivals to communicate the value and potential of administrative data and garner support and understanding of ADR Scotland’s work.
2. **Project-specific dialogue and stakeholder groups:** this involves engagement with policy and practice communities including third sector organisations and community groups that are able to represent the publics and communities relevant to each of ADR Scotland’s projects. Close working with these communities and building key stakeholder groups enables ADR Scotland to co-produce and shape research questions and maximise likely societal benefit.

ADR Scotland’s [Public Panel](#) forms a key element of the partnership’s public engagement approach. The Panel was first established under ADRN and has continued under ADR UK but with a new membership in 2019. The Panel, coordinated by SCADR, exists to:

- Help keep research and the use of administrative data in Scotland focused on the public good;
- Better understand the views and concerns the public hold about administrative data research;
- Advise on how best to engage with the public, make research accessible and effectively communicate findings;
- Discuss individual proposals for research;
- Become advocates for this type of research.

The Panel is made up of members of the public from across Scotland, from a range of different backgrounds and with lived experience of a range of issues that can be touched upon in ADR Scotland’s research.

ADR Scotland is also developing larger knowledge exchange symposium events around core areas, such as children and young people, to develop a greater public, policy and practitioner dialogue on administrative data research. These will explore how ADR Scotland can most sensitively and effectively deliver impactful research, and investigate where the most urgent priority areas are.

ADR Wales



[ADR Wales](#) – a partnership between Swansea University, Cardiff University, Welsh Government and the SAIL Databank – also has a long-standing [Consumer Panel](#) maintained by SAIL, which was originally established in 2011. The Panel allows ADR Wales to explore the public’s perspective on administrative data research, and provides opportunities for lay representatives to join research study teams. The Panel currently has 16 members, with ongoing recruitment.

Panel members are involved in all elements of the administrative data research process, from developing ideas and advising on bids and approval processes (via SAIL’s independent Information Governance Review Panel), to disseminating research findings.

The Panel’s role is manifold, and includes:

- Acting as advisors on issues in research
- Advising on how best to engage with the public
- Offering guidance on how to recruit people to study steering groups
- Providing views on data protection issues
- Discussing proposals for research
- Reviewing information designed for a lay audience
- Acting as advocates for data linking research.

Specifically, the panel is consulted on a range of topics relating to the use of administrative data for research, such as: research methodology; the acceptability of research proposals, and of research findings being used to target policy; how best to represent work in public facing materials; the acceptability of using particular types of data for research, and the research questions being asked using them; and how to build public engagement pathways into research projects, amongst other things, for a wide variety of projects investigating a range of issues.

ADR Wales coordinates a programme of engagement research with the wider public and particular topic or age groups, focusing on the use of various types of data and new developments. In addition, the ADR Wales Public Engagement team attend many events throughout the year to discuss the work of SAIL and the research conducted using the data. The events offer the opportunities to explain the work being done by researchers, gauge the public’s opinion on the research, and ignite collaboration opportunities. Stakeholder workshops are also held to with devolved and local government and third sector charities, to gain feedback on work already done and input on future work. The workshops provide an opportunity to tap into others’ expertise and receive valuable insight on projects using data curated by ADR Wales.

Previous literature has provided valuable insights into public attitudes towards the linking and use of administrative data for research and set out a clear route forwards for ADR UK’s own public engagement. You can find out more about ADR UK’s approach to enabling better access to linked administrative data for research, and keep up-to-date with our public engagement activities as they unfold, on the [ADR UK website](#).

10. References

Papers reviewed:

Aitken, M., et al. (2016a), [‘Moving from trust to trustworthiness: Experiences of public engagement in the Scottish Health Informatics Programme’](#), *Science and Public Policy*, 43:5, pp.713-723.

Aitken, M., et al. (2016b), [‘Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies’](#), *BMC Medical Ethics*, 17:73.

Aitken, M., et al. (2018), [‘Public Preferences regarding Data Linkage for Health Research: A Discrete Choice Experiment’](#), *International Journal of Population Data Science*, 3:11.

Cameron, D., et al. (2014), [‘Dialogue on Data: Exploring the public’s views on using administrative data for research purposes’](#), Ipsos MORI Social Research Institute.

Davidson, S., et al. (2012), [‘Public acceptability of cross sectoral data linkage: Deliberative research findings’](#), Scottish Government Social Research.

Davidson, S., et al. (2013), [‘Public Acceptability of Data Sharing Between the Public, Private and Third Sectors for Research Purposes’](#), Scottish Government Social Research.

Ipsos MORI (2007), [‘The Use of Personal Health Information in Medical Research: General Public Consultation’](#), Medical Research Council.

NatCen Social Research (2018), [‘Public Attitudes to Data Linkage: A report prepared for University College London by NatCen Social Research’](#).

Office for National Statistics (2014), [‘The Census and Future Provision of Population Statistics in England and Wales: Public attitudes to the use of personal data for official statistics’](#).

Oswald, M. (2014), [‘Share and share alike? An examination of trust, anonymisation and data sharing with particular reference to an exploratory research project investigating attitudes to sharing personal data with the public sector’](#), *SCRIPTed*, 11:3.

Rempela, E. S., et al. (2018), [‘Public engagement with UK government data science: Propositions from a literature review of public engagement on new technologies’](#), *Government Information Quarterly*, 35, pp.569-578.

Robinson, G., et al. (2018), [‘Public Attitudes to Data Sharing in Northern Ireland: Findings from the Northern Ireland Life and Times Survey 2015’](#), Ulster University.

Royal Statistical Society (2014), [‘Royal Statistical Society research on trust in data and attitudes toward data use/data sharing’](#).

Stockdale, J., et al. (2019), [‘“Giving something back”: A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland \(Revised\)’](#), *Wellcome Open Research*, 3:6.

Tully, M. P. et al. (2019), [‘Commercial use of health data – A public “trial” by citizens’ jury’](#), *Wiley: Learning Health Systems*.

Wellcome Trust (2013), [‘Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data’](#).

Other references:

Byrom, N. (2019), '[Digital Justice: HMCTS data strategy and delivering access to justice](#)', The Legal Education Foundation.

Information Commissioner's Office (2012), '[Anonymisation: Managing data protection risk code of practice](#)'.

Office for National Statistics (2017), '[The 'Five Safes' – Data Privacy at ONS](#)' (web page accessed 17/04/20)

Rowe, G., & Frewer, L. J. (2005), '[A typology of public engagement mechanisms](#)', *Science, Technology & Human Values*, 30:2.

Tindale, J., and Muirhead, O. (2019), '[Trust, Transparency & Tech: Building ethical data policies for the public good](#)', Policy Connect and the All Party Parliamentary Group on Data Analytics.

UK Research & Innovation (2019), '[UKRI Vision for Public Engagement](#)'.

About ADR UK

[ADR UK \(Administrative Data Research UK\)](#) is a partnership transforming the way social and economic researchers access the UK's existing wealth of administrative data, to enable better informed policy decisions that improve lives. By linking together data held by different parts of government and making it available to approved researchers, we are enabling vital research that has the potential to lead to more effective public services, in areas from improving education and healthcare to tackling crime. ADR UK does not collect any new data, and nor does it make personal data available for research; we link and enable secure access to existing, de-identified data for research in the public interest.

ADR UK is made up of three national partnerships – ADR Scotland, ADR Wales, and ADR Northern Ireland – and the Office for National Statistics (ONS). It is coordinated by a Strategic Hub based at the [Economic & Social Research Council \(ESRC\)](#), part of [UK Research & Innovation](#). ADR UK is initially a three-year investment from July 2018 to July 2021, supported by £44 million drawn from the National Productivity Investment Fund (NPIF) via the ESRC.

Contact

Elizabeth Waind
ADR UK Strategic Hub, Economic & Social Research Council
elizabeth.waind@ukri.org

www.adruk.org

 [@ADRUkdata](#)

 [@ADR UK](#)

 [ADR UK](#)

