

Public support for accessing and linking data about people from various sources:

Literature review

Dr Erika Kispeter

Warwick Institute for Employment Research

University of Warwick

June 2019

Contents

1.	Introduction.....	1
2.	Approach to the literature review	2
3.	Findings.....	3
3.1.	Conditions of public support for accessing and linking data about people from various sources	3
3.1.1.	What is understood by the phrase 'social and economic research' and about the need to access personal data to undertake such research?	3
3.1.2.	Conditions of public support for social and economic research.....	5
3.2.	Trust, trustworthiness, transparency and authenticity	14
3.3.	Exploring the acceptability of using linked data for research purposes	17
3.3.1.	What types of data are more or less sensitive when re-used for research purposes?.....	17
3.3.2.	Privacy and ethics	19
3.3.3.	Control and consent	19
3.3.4.	Data ownership	23
3.3.5.	Policies and safeguards.....	23
3.3.6.	What attitudes are held by different groups of the public?.....	25
3.4.	Public perceptions and understanding of public benefits.....	26
3.5.	What do people expect from public engagement?	28
4.	Discussion and Conclusions.....	29
5.	References.....	33

1. Introduction

The potential for data collected to be linked and used in research has led to growing academic and policy interest in the public acceptability of data linkage. There is also increasing recognition of the importance of ensuring a social licence for using linked data for research purposes. This report aims to review the literature which explores public attitudes to data and data science practices along the following broad research questions:

1. Under what conditions does the public support accessing and linking data about people from various sources? In particular the review addresses: What is understood by the phrase 'social and economic research' and about the need to access personal data to undertake such research? Under what conditions is such research seen to be in the public interest?
2. The second research question focuses specifically on the literature on trust, because 'public trust' has emerged as one of the key conditions of public support for data linkage. What is the current thinking on trust and the related concepts of trustworthiness, transparency and authenticity?
3. The third research question explores public acceptability for the use of (longitudinal) survey data and administrative data linked together in relation to: privacy and ethics, consent, data ownership, policies and safeguards and the public perceptions, focusing on different groups of the public where possible.
4. The fourth research question emerged in reflection to the early findings of this literature review: How do the publics make sense of the potential (public) benefits of social-economic research and data linkage in particular, and what do the publics expect of public engagement activities in the context of data sharing and linkage?¹

While the main focus of the literature review is on social-economic research, studies on public attitudes towards linking health data are also discussed.

The findings of the literature review are aimed to provide recommendations on how the ESRC can best proceed with the development of its data strategy and the delivery of

¹ Data-sharing involves information moving from one organisation or department to another, whereas data-linkage is defined as: "the bringing together from two or more different sources, data that relate to the same individual, family, place or event" (Cameron et al., 2014, p. 11).

research using administrative and linked data in a way that works positively and pro-actively with issues of public concern.

2. Approach to the literature review

The research parameters and the keyword search strategy (Table 1) were developed by the IER research team. This strategy was then used systematically across a selected number of electronic bibliographic databases: ABI/Inform Global (by ProQuest), Scopus, Web of Science and Science Direct. In addition, Google Scholar and Google were also searched. Finally, relevant websites² and journals³ were ‘hand searched’ for literature. The search was restricted to publications from 2009–2019 inclusive. In the search phase 675 pieces of literature were identified and after the initial screening phase 64 pieces of literature were added to the bibliographic software programme, Endnote. After considering the references in the selected literature and consulting the project Advisory Board, 128 pieces of literature are included in the Endnote library.⁴

Table 1 Keywords used in the search

Keyword 1	Keyword 2	Keyword 3	Keyword 4	Keyword 5
administrative	data	linkage	public	engagement
survey	(record)	link		consultation
longitudinal		sharing		attitudes
personal				opinion
				dialogue

² ESRC, The Royal Statistical Society (RSA), the Administrative Data Research Network, ONS, Understanding Patient Data, NHS Digital and Involve.

³ International Journal of Population Data Science, International Journal of Social Research Methodology and International Journal of Public Opinion Research.

⁴ Literature on different forms of public engagement is included in the Endnote library but not reviewed here.

3. Findings

In this section the findings of the literature review are presented around the three broad research questions listed earlier.

3.1. Conditions of public support for accessing and linking data about people from various sources

The evidence that is relevant to this question is discussed in four sub-sections, looking at the public understanding of social and economic research and the need to access personal data for such research; the conditions under which social and economic research is perceived to be in the public interest; the types of data that are perceived as more or less sensitive when re-used in research, and the different attitudes toward data linkage shared by different groups of the public. In each subsection the literature from the field of social and economic research is presented first and the health-related literature is presented second.

3.1.1. What is understood by the phrase 'social and economic research' and about the need to access personal data to undertake such research?

A study by Cameron and colleagues (2014)⁵ used public dialogue workshops to explore the public understanding of the use of administrative data and data linkage by the Administrative Data Research Network (ADRN). Participants had very low understanding of social research and those who held negative perceptions about social research felt that research findings were not more than 'common sense' knowledge. The report argues that scepticism about the value of social research was driven by the lack of understanding and that by the end of the dialogue workshop some participants had a more positive opinion on social research, because their knowledge had increased. However, many participants continued to question the purpose of social research on the second day of dialogue workshop, after several explanations. Closely related to this is the finding that participants often confused the research uses of data with operational uses (e.g. improving service provision). This finding is relevant if we consider that according to Davidson and colleagues (2013) the public view the primary aim of research as an improvement of public services.

⁵ The study was commissioned by the Economic and Social Research Council (ESRC) and the Office for National Statistics (ONS).

A number of studies (for example, Cameron et al., 2014 and Van Mil & Hopkins, 2015) found that although public dialogue participants often found it difficult to understand the concept and process of de-identification and data linkage, they were interested in the explanations. Cameron and colleagues (2014) highlighted that, perhaps unsurprisingly, participants were particularly interested in social research that was the most relevant to them personally.

There is evidence that health research is better known than social research: a study by the Wellcome Trust (2013)⁶ found that there was some awareness of specific examples of data sharing and linkage, for example the sharing of medical records within the NHS. However, participants found it difficult to envisage other forms of health data linkage and their knowledge about the collection and the use of health-related data had major gaps.

When discussing data linkage, participants were given a number of examples.⁷ These are described in some detail, to provide insights into how the public understood the need to link personal data from different sources for research purposes.

- The use of de-identified loyalty card data to inform public health programmes was broadly acceptable, but linking card data with personal patient health records was seen by some participants as an invasion of privacy ('the Government snooping on people') and exposing them to targeted messages (for example, sending warnings to people who buy unhealthy food).
- The example of linking anonymous household energy use to cold weather payments and hospital admissions was difficult to understand and, as the researchers argue, this led to the cynical interpretation that the Government would try to reduce payments.

⁶ The study's aim was to understand the general public's attitudes to different types of personal data and data linking. In particular, the study explored whether health and other types of data were viewed differently, and what the perceived risks and benefits of linking different kinds of data for research and other purposes were. The methodology included six focus groups and six telephone interviews with members of the public in England in spring 2013. A total of 50 people took part. The sample included people aged 18-70 from socio-economic groups ABC1 and C2DE. All respondents completed a self-completion questionnaire before their focus group or interview, which involved identifying benefits and disadvantages about personal data collection including health data.

⁷ Facebook user data being linked and used for targeted advertising; anonymous loyalty card data being used to influence public health programmes; loyalty card data being linked with personal patient health records; anonymous household energy use being linked to cold weather payments and hospital admissions; pre-term birth data being linked with educational achievement; medical research charity or pharmaceutical company linking mortality data with health records and statin use; routine Government administrative data linking education records, criminal records, DWP data and income data, e.g. from the *British Household Survey*.

- Linking pre-term birth data with children’s educational achievement later was seen by the minority as having potential benefits (if pre-term children indeed have additional learning needs), but the majority responded negatively and were concerned about the children being stigmatised, for example when applying to primary school)
- The example of linking education records, criminal records, DWP data and income data (for example from the British Household Survey) to administrative data was seen as not having a clear purpose. While some participants welcomed the idea that so called ‘benefit cheats’ might be identified, others felt a sense of ‘Big Brother’ society and were concerned that collecting data might disfavour people financially or incriminate them.

In the next section, evidence about the conditions under which research using linked data is perceived to be in the public interest is explored, including both the literature on socio-economic and health data.

3.1.2. Conditions of public support for social and economic research

There is a consensus in the literature that the public is ‘broadly’ supportive of data linkage and sharing both in social science and health research (Cameron et al., 2014, Davidson et al., 2013; Van Mil & Hopkins, 2015; NatCen, 2018; Aitken, 2016a; Wellcome Trust, 2013) if a number of important conditions are met.

A systematic review of the literature on public attitudes toward sharing and linking health and other personal data by Aitken and colleagues (2016a)⁸ identified the following conditions of public support for the use of data for research purposes:

- Research to bring *public benefits*: the research use of data is supported if it brings about (or has the potential to bring about) public benefits;
- Assurances that data will be used for *legitimate purposes*. However, there are different views on how, or by whom legitimacy is to be defined;

⁸ Aitken et al. (2016a) have conducted a systematic review and thematic synthesis of qualitative studies examining public attitudes towards the sharing or linkage of health data for research purposes. Twenty-five studies were included which were conducted between 1999 and 2013. The qualitative methods represented in the studies included focus groups, interviews, deliberative events, dialogue workshops and asynchronous online interviews.

- Assurance of individuals' *confidentiality*, which is largely associated with the anonymisation of data. There is a distinction between 'plain stats' and more detailed qualitative information, with the latter raising greater confidentiality and privacy issues.
- Assurances of *safeguards* to protect against misuse or abuse of data
- Assurances of *data security*
- Profit creation is accepted on the condition that *public benefits are prioritised over profits*.

Focusing more narrowly on attitudes to using social and economic data, a study by Cameron and colleagues (2014)⁹ identified two additional conditions: that *public benefits* of the research are *clearly articulated*, and that *private businesses are not able to access the data*. The 'red line' related to private businesses was relevant even if businesses used the data to deliver a public service. In other words, the study found that public benefit and the financial benefit of a private business were seen by the public as incompatible.

Another condition of support for data linkage was identified in a study¹⁰ by Van Mil & Hopkins (2015).

- The use of administrative data should "adequate, relevant but not excessive" (p. 14).¹¹

Turning to the literature on health-related research, two key conditions of public support emerged from a systematic literature review of public attitudes to the sharing and linking of health data (Stockdale and colleagues, 2018). Firstly, that the organisations involved in working with (patient) data has the right expertise and competencies to ensure data security, and secondly, the organisations has the 'right' motivations, i.e. those which serve the public, rather than private interests. It is argued that public attitudes to data-sharing can

⁹ This dialogue was conducted as part of the Public Attitudes to Science (PAS) 2014 project. It focused on the ESRC-funded Administrative Data Research Network (ADRN). The dialogue consisted of seven sets of workshops with members of the public and sector experts. These took place between in 2013 in England, Wales, Scotland and Northern Ireland. In each location, a group of 14-20 public participants attended an all-day workshop and reconvened two weeks later for a second day-long workshop. In total, 136 members of the public were involved (Cameron et al., 2014, p. 2).

¹⁰ A dialogue on the use and re-use of private sector data for social research (2015).

¹¹ The report is based on two rounds of public dialogue that involved 62 members of the public in three locations across England and Scotland. The aim of the public dialogue was to explore public views on access to and the use of data from private sector organisations for research purposes within the ESRC-funded Data Research Centres. The dialogues focused on the following types of data: High street consumption, financial data, transport and movement, online consumption, energy use, and communications, including social media.

be understood by using an *ethical framework*, which is widely accepted in biomedical research. The framework is based on the following four questions:

- Do the methods of data collection respect individual patient autonomy? (*Respect for Autonomy*)
- Could access to the data, or a particular use of the data, lead to individual harm? (*Nonmaleficence*)
- Are the objectives and the intended outputs of the study primarily concerned with contributing to the public good? Do they have clear scientific value? (*Beneficence*)
- Related to question iii, is any agreement between the NHS and organisations providing analytics (private or public) fair and just? Will it lead to an outcome which will provide benefit to the key stakeholders of the data, i.e. the patients? (*Justice*)

Focusing on the acceptability of private sector involvement in health-related data-intensive research, a study by Ipsos Mori found that dialogue workshops participants applied four key ‘tests’ when evaluating the value and the risks in the research case studies and the four tests (shown in Figure 1) were always applied in the order in which they are listed here (Ipsos MORI, 2016).¹² In other words, there is a hierarchical relationship between the conditions of public acceptability, with ‘*clear public benefit*’ being the most important.

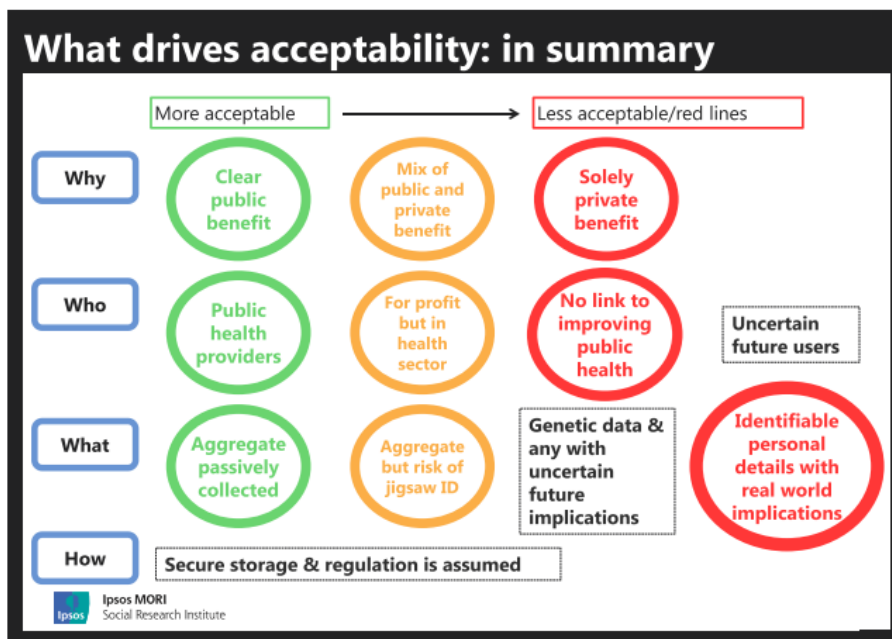
In the next section, the four tests are discussed in more detail. It is important to keep in mind that the focus of this study was on data sharing and linking with commercial companies, which means that the conditions of support may be stricter than those found in other studies.

- *Why – Does the activity’s outcome have a provable and sufficient public benefit?*

A clear benefit both to individuals and to wider society was seen as the only good rationale for breaking privacy and thus, the most important condition of acceptability.

Figure 1: Conditions of public acceptability of commercial access to health data (Source: Ipsos MORI, 2016a, p. 9).

¹² [Details about Ipsos Mori 2016 - One way mirror)



- *Who - Can the organisations doing this be trusted to have public interest at heart?*

Participants in the dialogue workshops were most supportive to academic researchers, charities, and public sector organisations accessing health data. Commercial access to health data was found to be acceptable if participants perceived provable and sufficient public benefits of the data access (test 1) and trusted the organisation accessing the data (test 2). Results from the survey element of the study show that majority supported commercial access to health data for research purposes without testing the organisation accessing the data (without applying test 2) if they felt there was a clear social benefit.

There were also a number of red lines: the public did not accept the sharing of health data with insurance and marketing companies and with third parties.

- *What - How anonymised or aggregated is the data?*

Participants in the workshops did not find aggregated data to pose a risk to the privacy of individuals, but thought that even aggregated data could lead to certain groups being discriminated against. There was also an important red line: any data sharing or linking that risked personal harm was seen as unacceptable. The study notes that participants found it hard to judge the likelihood of different types of data leading to harm.

- *How - Does the safeguarding, access and storage protocol reassure me that the data will be safe?*

It is argued in the study that the call for safeguards reflected participants' underlying concerns about the governance of data and their need for more discussion about regulating the outcomes from data sharing.

A study by Aitken and colleagues (2018a) aimed to better understand the relative importance of the key factors that shape public preferences for data linking and use in health research. To achieve this aim, the study combined a number of 'traditional' survey questions with a Discrete Choice Experiment (DCE).¹³ From the survey the following conditions for support emerged:

- *Potential profits* arising from research: Over half (62%) chose the answer "Any profit made from research carried out using linked information should be invested into public services" and only 8 per cent chose the option: "Any profit made from research carried out using linked information should be kept by those carrying out the research."
- *Public benefit*: Over half (57%) of respondents agreed with the statement "Research using linked information should only be used if it will have general public benefits."
- *The types of data linked together*: Almost half (48%) supported with the option "Information from your GP records being linked with information from your other NHS health records (e.g. hospital records)." According to the authors, this result suggests that most respondents prefer research to be conducted using only health data and are not comfortable with cross-sectoral data-linkage.
- *The researcher*: A third (31%) chose the answer that "only university researchers, NHS staff or government researchers" should have access to anonymised linked (health) data.

¹³ The DCE method makes it possible to understand choices between different fixed options around the sharing and linkage of personal information for research purposes and to establish the relative importance of different factors and considerations that constitute each option. The authors developed 240 different scenario pairs. 1,004 respondents completed an online survey in summer 2016. Limitations of the study methodology: The study was conducted online, thus respondents were confident users of IT and the internet and it is possible that people less confident using IT have different preferences. Finally, because the respondents signed up to take part in research, they may be more likely than average to be support the research use of data.

- *Overseeing and monitoring data linkage*: A third (35%) felt that “The process should be overseen by a [non-governmental] independent body” and another third (32%) chose “The process should be overseen by the relevant public service(s); for example, research that uses information from people’s health records should be overseen by the NHS”. Only 6 per cent chose “The process should be overseen by the organisations undertaking the research”.

In the ‘Discrete Choice Experiment’ (DCE) element of the study, the researchers developed scenarios, which were characterised by five attributes, and within each attribute there were four levels. To give an example, one attribute was ‘the researchers conducting the study’ and the levels ranged from the most restrictive “only university researchers” (level 1) to ‘university researchers, NHS staff, government researchers and commercial researchers such as market research organisations or pharmaceutical companies’ (level 4). Research participants were given pairs of scenarios and asked to choose the more preferable. The study notes that throughout the DCE respondents were reminded that the data linked and used for research was always anonymous and that individuals could not be identified.

The DCE study identified two main conditions for public support:

- the way profits from the research were managed and shared; and
- the types of data being linked.

In addition, respondents noted that the data being anonymised was an important consideration for them and in some cases this factor prevented them from completely rejecting data linkage. Aitken and colleagues (2018a) argue that this finding points to the importance of privacy, as a third key condition for supporting data linkage. It is also important to note that for a large group of respondents (461 people out of 1461) data linkage was unacceptable under any conditions. The questions about managing profits and the types of data being linked were the most likely to prompt respondents to reject data linkage – a finding that confirms the main results of the study.

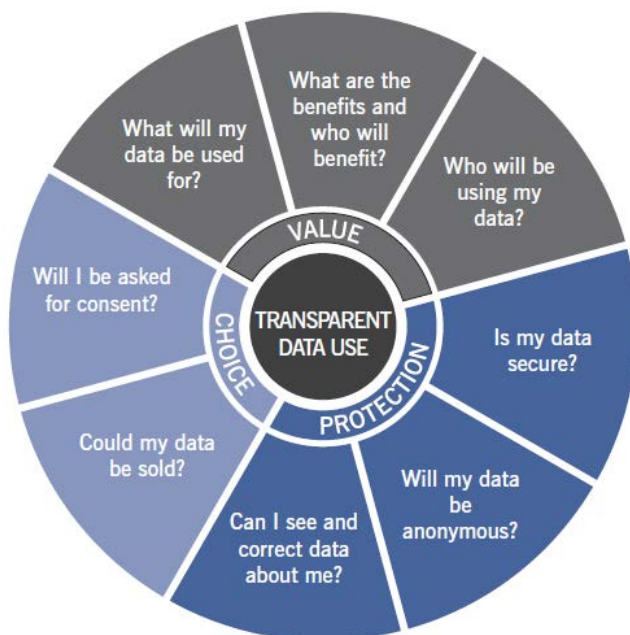
Rather than focusing on acceptability or preferences, the Trusted Data Partnership report (2017) discusses what information the people of New Zealand need to ‘have comfort’ in a proposed use of data. The eight main aspects of data use identified as important for New Zealanders are very similar to ‘conditions of acceptability’ discussed earlier:

- the specific *purpose and nature* of data use – what will my data be used for?

- the *benefits* – what are the benefits and who benefits (the individual, the whānau , the wider community, the government or a business?)
- *who* is collecting and using the data (whether the organisation itself is trusted, whether the information will be shared with third parties)
- the *security* measures (are there strict procedures around access and use?)
- Data *anonymity* (can an individual or whānau be identified?)
- The ability to see and correct the data and the ability to audit how it has been used
- the level of choice about the data and its use (ability to consent or withhold consent)
- whether the data could be sold.

What is new in the New Zealand study is that these aspects are grouped under three headings: *value*, *protection* and *choice*, and presented in the ‘Transparent data use wheel’, shown in Figure 2.

Figure 2 Wheel of transparent data use [Source: Data Futures Partnership, 2017]



Learning from the case of *care.data*

Looking at the issue of public support from a different perspective, but touching upon most of the key points mentioned earlier, Carter and colleagues (2015) used the negative public

response to the introduction of the care.data initiative as a case study¹⁴ to explore public support for using health data for research purposes. Drawing on the literature, the authors state that the public's cooperation with specific research projects depends on their perceptions that their participation in research is: voluntary and governed by the values of reciprocity, non-exploitation and service of the public good. Further, the study argues that co-operation with the researchers is secured through three principal mechanisms:

- the public's expectations about how research is conducted and regulated;
- their trust in the institutions and individuals who recruit them; and
- their beliefs in the wholesomeness and public value of the research endeavour – that health research is socially valuable and conducted in the service of the public good (Carter et al., 2015, p. 406).

The word belief is important here: Carter and colleagues argue that public support for research often depends more on 'a set of cues' about the social good and the safety of research participation, and on the institutional and professional credentials of researchers than it does on formal regulation of research or rational assessment of documents aimed at gaining consent (p. 406). Looking specifically at the case of care.data, Carter and colleagues argue that the main reason for the public concern was that data which was given in a medical context was seen to be 'unquestioningly repurposed' to another, seemingly unconnected context. This is argued to have broken the 'social licence' between the community of health experts and the public.

The study by Carter and colleagues (2015) defines social licence as: *The expectations of society regarding some activities (that are likely to cause public 'disquiet') may go beyond compliance with the requirements of formal regulation" (p. 404).*

The definition adopted by the Data Futures Partnership in New Zealand, where the concept of social licence is argued to have first emerged, focuses more narrowly on data: *"Social Licence is a term emerging from debates about data sharing [...] and rests on the assumption*

¹⁴ Care.data was an initiative to link GP and hospital data in England for the purpose of NHS commissioning, resource allocation, audit and research. There was little public consultation or information campaign and it was also revealed that NHS England had previously sold data to insurance companies. After members of the medical profession and the public expressed concern and some negative press coverage, the initiative was put on hold (This summary is based on Carter et al., 2015).

that when people trust that their data will be used as they have agreed, and accept that enough value will be created, they are likely to be more comfortable with its use” (Data Futures Partnership, 2016).

Sterckx and colleagues (2016) analysed the comments and questions that members of the public left on the dedicated website after the *care.data* initiative was suspended. They identified three factors that are relevant to this literature review:

- While the public wanted to further the common good, they did not want to be manipulated into doing it.
- The public wanted to be safeguarded against various potential risks and abuses.
- Finally, trust was an overarching issue.

A study by Hobbs and colleagues (2018) was also inspired by the story of *care.data*. A series of public engagement activities were conducted to identify public concerns about data-linkage and sharing. The analysis yielded two ‘global themes’: trust and mistrust and (breaches of) privacy and confidentiality.

In summary, the key conditions of support emerging from the literature:

- Public benefit (provable, sufficient, clearly articulated)/ legitimate purpose), including the issue of who will benefit from the data sharing/linkage?
- Privacy/confidentiality/the level of de-identification/anonymity
- Trust in the researcher/institution involved in data linkage/the use of data
- Data security/safeguarding data against misuse/data governance

The following conditions were also found to be important:

- Respect for autonomy
- Consent
- Managing profits/the involvement of the private sector
- The types of data being linked together

Trust has emerged as one of the key conditions of public support for data linkage. The next section briefly discusses the concepts of trust and trustworthiness and the related concepts of transparency and authenticity.

3.2. Trust, trustworthiness, transparency and authenticity

A survey by Ipsos MORI for the Royal Statistical Society (RSS) (RSS, 2014)¹⁵ explored public trust in institutions handling personal data, and attitudes toward data sharing and linkage by government bodies. It revealed a hierarchy of public trust in institutions and found that the public does not have a high level of trust in organisations to use their data appropriately:

- There was a 'trust in data deficit': Institutions were trusted less to use data collected from the public than they were trusted generally. Public health services saw a large deficit, but there were even larger gaps between general trust and trust with data for academics and charities. The trust in data deficit was: -10 percentage points for the NHS, -12 for GP surgeries and -23 for academic researchers and universities. The largest gap was for charities: -26 percentage points. The study suggested that the higher levels of trust in the public sector and in primary health providers in particular can be explained by the fact that participants are familiar with these institutions.
- The study also explored public opinions to the sharing of anonymised data outside of government organisations and found a clear hierarchy, with strongest support for data sharing with researchers, then charities, and lastly for profit companies. However, only 50 per cent of respondents supported data sharing with researchers in universities and 17 per cent opposed this.

The concept of trust is discussed in more detail in a paper by Aitken and colleagues (2016b).¹⁶ Acknowledging that there are clear patterns of public trust in different actors, the authors argue that these patterns do not straightforwardly translate into public support for research. For example, the focus group participants in the study accepted that the private sector (low level of public trust) may be involved in research using personal medical data, while, academic researchers (high level of trust) were not given unconditional support for accessing and using personal data. The key argument of the paper is that the public's relationship with science is too sophisticated to be characterised by the simple binary of trust and distrust: the public often adopts an ambivalent form of trust. The paper calls for a

¹⁵ Ipsos MORI polled adults aged 16-75, living in Great Britain using Ipsos MORI's online omnibus, a quota survey. Fieldwork for most questions consisted of 2,019 interviews, and one question of 1,000 interviews. Results are weighted by age, gender, region, social grade, working status and 'main shopper' status.

¹⁶ A study based on qualitative research that explored public attitudes to the Scottish Health Informatics Programme (SHIP).

shift of focus from the level of trust the public feels to the trustworthiness of research projects and researchers. This shift requires the research community to reflect on the institutional arrangements of science and explore on what bases public trust is founded and what it means for research and researchers to be trustworthy.

Stockdale and colleagues (2018) also adopted the trustworthiness framework when analysing the findings of their literature review and argued that the public evaluates the trustworthiness of research organisations by assessing their competence in data-handling and motivation for accessing the data.

Aitken and colleagues (2016b) argue that public engagement is an opportunity to make a research project and a research community (more) trustworthy, provided that the engagement exercise reflects the values, interests or concerns of the public. In other words, public engagement is both an indicator of the trustworthiness of the research and researchers and at the same time, a mechanism for ensuring this trustworthiness.

Transparency and authenticity

Aitken and colleagues (2016b) highlight that transparency is increasingly seen as a mechanism for addressing the lack of public trust in science and scientific institutions. The paper discusses three different understandings of transparency:

- *informational transparency*, openness about information on which decisions are based but also about the value of data-linkage,
- *participatory transparency*, enabling public participation in decision making and a more equal relationship between researchers and the public, and
- *accountability transparency*, which calls for openness about research governance and the accountability of decision makers.

It is argued that informational transparency, the open communication of uncensored information may be perceived by the public as an indicator of trustworthiness and researchers will be more likely to be perceived as trustworthy if transparency involves a dialogue with members of the public and opportunities for deliberation.

Aitken and colleagues (2016b) found that in many cases the concerns or reservations of the public stemmed from a perceived lack of openness about how (health) data was collected and used and how these processes are governed. Some participants described a sense of

unfairness and inequality: they argued that they were expected to allow increasing access to their data but that they were not expected to want to know for what purposes the data was used, how it was used and by whom. Participants wanted greater openness in relation to how data is currently used and how requests for data access are managed and emphasised that the information given to them should be 'accurate, impartial and uncensored'.

Aitken and colleagues (2016b) argue that transparency does not represent a solution to low levels of public trust: when the public is sceptical about participatory events 'transparency' may be perceived as insincere. This situation is described as a circular conundrum: transparency is necessary to build trust, but trust is required in order for the transparency to be recognised as adequate (Brown and Michael, 2002, quoted in Aitken et al., 2016b).

Aitken and colleagues argue that this circle cannot be broken by more transparency but rather by more 'authenticity'. Authenticity can be signalled through emotional engagement and "*demonstration of the pain or suffering endured through decision-making processes*" (p. 9). Aitken and colleagues (2016b) argue that researchers and decision-makers can develop authenticity by demonstrating that they have engaged with the public and attempted to incorporate and address different views to such an extent that it has caused them distress or suffering. Dialogues with the public are 'performances of authenticity': attempts to demonstrate to the public that research programmes or institutions are actively trying to address different, often conflicting viewpoints. It is argued that authenticity builds trust and confidence in the transparency of decision-making.

Cameron and colleagues (2014) found that there was little interest from the general public in becoming actively involved with the ADRN, rather, members of the public stated that transparency was sufficient, and it could be ensured through some form of public representation in the decision-making process around the ADRN.

Responding to the care.data failings, the 'Understanding Patient Data initiative' by the *Wellcome Trust* aims to develop a shared language between researchers and the public to describe health data and the benefits of research. Carol Porteous (ADRN Scotland, 2017) argues that the initiative signals to the public that the research community is trying to be more co-operative and calls for similar initiatives in data intensive social sciences.

3.3. Exploring the acceptability of using linked data for research purposes

This section of the literature review further explores the public acceptability of using data from different sources linked together along these key issues: privacy and ethics, consent and control, data ownership and policies and safeguards. Some of the themes discussed have already been discussed in section 3.1 when the focus was on the conditions of public support for data sharing and data linking. This section looks more closely at how the public thinks research using linked data should be done, while the previous section looked at whether such research should be done at all.

3.3.1. What types of data are more or less sensitive when re-used for research purposes?

Cameron et al. (2014) found that in contrast with their low awareness of social research, dialogue participants were familiar with the importance of data to modern society. They were quickly able to give numerous examples of providing their personal data to businesses and government and saw this as an “unavoidable aspect of modern life” (p. 3).

A number of studies have found that health and care data are seen by the public to be different than other forms of personal data (Davidson et al., 2013; Wellcome Trust, 2013). Davidson and colleagues (2013) argue that health data is seen as special, because of its potential to be used for treating or preventing illness was seen as an ‘unquestionable’ public benefit. In the deliberative exercise, reported in the study, postcode data was seen as sensitive because of its potential to be disclosive and because it could lead to people from certain postcode areas being stigmatised (for example, living in an area of high deprivation, or high level of crime). As for demographic and socio-economic data, very few participants expressed concerns about sharing data on an individual’s sex, age, ethnicity, disability and maternity. Data on sexual orientation was seen as more sensitive, again, because of the fear of potential stigma and harm. Data on religion on the other hand was seen as less contentious. Finally, banking and other financial information was seen as very private and there was concern that sharing such data could lead to financial crime.

The literature review by Aitken and colleagues (2016a) found that particularly sensitive types of data were: mental and sexual health, sexuality and religion.

Research on how participants of longitudinal surveys consent to linkage to different types of administrative records also provides insights into what types of data are perceived as more or less sensitive when re-used for research purposes. A study by Mostafa (2014) analyses consent rates to linking survey data in the *Millenium Cohort Study*¹⁷ (MCS) to health and economic data records. Consent rates were the lowest for linking parents' economic records to the survey (81%), suggesting that parents' economic data was seen as more private than their health records (87%). The consent rates for linking the children's education and health records were much higher and were almost at the same level (94 and 93% respectively).

In their analysis of young people's consent behaviour in the PEARL Study (Project to Enhance ALSPAC through Record Linkage)¹⁸ Boyd and colleagues (2015) found that the consent rate for linking school records was the highest (93%), followed by the consent rate for health records (92%), for criminal conviction and caution records (90%), with the rate for financial records at the lowest level (85%). Finally, a study by Al Baghal (2016), analysed the consent rates to linking administrative data and the *Understanding Society* study, and also found that consent rates for linking health records were lower than those for education records for both the children and the parents.

These findings suggest that education records are perceived as the least sensitive (indicated by the highest rate of consent to data linkage), irrespective of the focus of the longitudinal study: ALSPAC is traditionally biomedical, while the MCS and US are more social science. The finding that the focus of the study does not impact on consent rates, contradicts the 'shifting the goal posts' argument, which was put forward by analyses of the care.data case, mentioned earlier¹⁹.

A *Eurobarometer* survey (Eurobarometer 359, European Commission, 2011a, quoted in Castell et al., 2014) found national differences in what type of data is considered more or less sensitive. British respondents were stricter than the EU average about what data they considered to be personal, with financial and health data considered particularly private.

¹⁷ [Short description of the Millenium Cohort Study]

¹⁸ PEARL is a randomized controlled trial within the Avon Longitudinal Study of Parents and Children (ALSPAC). Cohort members were invited to re-enroll at age 18 and consent to linkage to their health and administrative records (education, benefits and earnings, and criminal convictions and cautions records).

¹⁹ This argument was made by Andrew Boyd.

3.3.2. Privacy and ethics

In addition to what has already been discussed about the trade-off between the (potential) public benefit and the risks of data linking, Cameron and colleagues (2014) found that in discussions about the remit of the Administrative Data Research Network (ADRN), dialogue participants rejected the following ideas to protect the privacy of individuals: creating large databases with many variables and data from a large number of public sector sources; the linking of some passively collected administrative data, in particular geo-location data; and linking administrative data with business data.

A study by Van Mil & Hopkins (2015) found that the issue of ethics was key to the public acceptance of linking private sector data for the purpose of social research and that dialogue participants found the expertise and reputation of social researchers (in this study, researchers working at the ADRN) reassuring.

3.3.3. Control and consent

The literature review by Aitken and colleagues (2016a) highlights that individual control over how data is used (perceived autonomy), is very important to the public. However, in the literature there is no clear consensus about what the need for 'control' implies and how researchers should respond to it. In the reviewed literature control is largely discussed in relation to consent: the public feels that individual consent is particularly important when the research involves named or identifying data; qualitative information; the use of genetic data or where a commercial entity is involved in research. However, in some studies consent is seen merely seen as an act of courtesy.

This finding is confirmed by Audrey and colleagues (2016)²⁰ who find that the young people in their study tended to think about consent as opt-in consent, despite having a range of consent procedures explained to them. The authors argue that it is difficult to change the perceptions of what consent means. Another finding is that anonymisation is not always seen as sufficient to eliminate the need for consent. Even when research uses anonymized data, the study finds that some research participants felt it was polite that their permission

²⁰ [Description of the study]

should be sought for secondary use of data about them. Finally, for some participants consent meant a way of being informed about the secondary use of their data.

Much of the literature reviewed by Aitken and colleagues (2016a) links public attitudes to consent with the level of public trust in institutions involved in the research. The key argument is that if the participants trusted the organisations that handled their data, they were less likely to demand strong forms of control (individual consent). When this trust is lacking, participants want to have greater control over their personal data. This finding is confirmed by two large European projects, PriSMs²¹ and SurPRISE²², which also conclude that if people have trust in the institution, they tend to be comfortable about decisions taken on their behalf. However, when research participants were convinced of the value of research and the potential for public benefits, individual consent was seen as non-essential (Aitken et al., 2016a).

A study by NatCen (2018) which involves respondents to the Health Survey of England (2016) found that participants could not remember whether they had given consent to data linking or not, and they could not remember the reasons for their decision either. Thinking about the reasons for consenting to data linkage, respondents suggested a number of reasons: perceived societal benefits (especially in the health sector), assurance that the linked data could not be used to identify them and trust in the organisation running the survey (in this case, NatCen). Reasons suggested for not agreeing to data linkage included the lack of detail about exactly how the data would be used and the indefinite timeframe for which it could be used. A key finding of this study is that participants held mixed views on whether consent should be sought on an on-going basis for each use of the data.

A study by Ipsos Mori finds that attitudes to consent changed during the deliberative workshops (as did perceptions of acceptability) and some participants moved away from the initially preferred opt-in consent models. The study suggests that because the participants felt more knowledgeable about the processes and safeguards, their trust in the decision-making process around data collection and sharing improved, and they became more relaxed about consent (Ipsos Mori, 2016a, p. 13).

²¹ <http://prismsproject.eu/> cited in Green et al. (2015)

²² <http://surprise-project.eu/research/> cited in Green et al. (2015)

Robinson and colleagues (2018) explored the public attitude to consent by asking study participants if, and to what extent they felt consent to data linking was necessary. This question was asked after examples involving data linking were discussed in detail. Respondents were divided on this issue: 30 per cent felt that consent was vital and the research could not proceed without it, a third felt that researchers should try to get consent for data linking where possible, and 30 per cent felt that consent was not necessary as long as the data was de-identified (Robinson et al., 2018, Table 11, p. 14). This study also explored the relationship between participants' attitude to consent and the level of trust they had in different organisations to keep data secure. Attitude to consent was measured by using the following two responses: anonymisation is 'not always necessary' and it is 'always necessary'. The study identified strong relationships: those who felt that research should not be conducted without consent were more likely to state that they did not trust the NHS (20% vs. 9%), the government (35% vs. 20%), their GP practice (11% vs. 5%), commercial organisations (63% vs. 54%) and academics (29% vs. 19%).

Community-level control as an alternative of individual consent

Stakeholders in the study about the Scottish Health Informatics Programme (SHIP), conducted by Aitken and colleagues (2016b) acknowledged that as individuals they had little control over the governance of data-sharing processes. Thus, it was an important consideration for them who was in control of these processes. They felt that members of the public should have a role in overseeing [data sharing] processes and that involving lay representatives in decision making would ensure better accountability and protection of public interests.

Different forms of community level control over data linking, such as a 'citizens' governance panel' are currently being discussed for the Great North Care Record Cities (Mulrine et al., 2018).

In the conclusion of their paper, Audrey and colleagues (2016) argue that public opinions about consent in relation to data linkage are complex and diverse, and it would be challenging to offer solutions that are acceptable to the majority of the public. They call for "pragmatic, imaginative and flexible approaches" to consent if "research using data linkage is to successfully realise its potential for public good without undermining public trust in the research process" (p. 12).

3.3.4. Data ownership

A number of studies (Bartlett, 2012; Cameron et al., 2014; Van Mil & Hopkins, 2015) find that dialogue participants believe de-identified personal data remains, or should remain theirs, because it is data about them. In a survey conducted for the Royal Statistical Society (RSS, 2014) a third of respondents (35%) said that they continued to care about how their data is used, even in its de-identified form. In contrast, a study by Castell and colleagues (2014) found that when dialogue participants were given an explanation about the de-identification of a dataset, they often felt that the remaining data did not belong to them anymore, in other words, “what was left could have come from anyone” (p. 136). When discussing a scenario where names, addresses, National Insurance numbers and other familiar numbers were removed, many participants felt that they were not part of the dataset any longer (p. 136).

Studies find that data ownership is a complicated concept for most dialogue participants to grasp. They also raised (legitimate) questions about the ownership of linked data sets, as the following quote demonstrates:

They [the Data Centres] are going to have data coming in, they're going to be processing it, so although it originally came from a source, the new information that's been collected, is that owned by them or does [ownership] still go back to the original [data controller]? Van Mil & Hopkins (2015, p. 30)

3.3.5. Policies and safeguards

A number of studies highlight that there was a low level of public awareness about current research practices, including safeguarding and regulations. Van Mil & Hopkins (2015) found that that the public doesn't know who holds data about them. The survey commissioned by The Royal Statistical Society (RSS, 2014) about trust in institutions and attitudes toward data sharing found significant knowledge gaps: many participants thought more data was shared than is commonly the case, for example, a third thought that health records were sold to private companies. Discussing the use of health data by private sector organisations, a study by Ipsos MORI found that there was not an awareness of the current regulatory framework around the use of anonymised health data and dialogue participants knew very little about how datasets were used and managed (Ipsos MORI, 2016a, p. 4).

The study by Van Mil & Hopkins (2015), mentioned earlier, identified the following expectations towards acceptable data infrastructure:

- the infrastructure should operate independently of both government biases and of the interests of commercial organisations;
- safeguards are needed, including clear vetting procedures for researchers and penalties for data breaches;
- impartiality and expertise should lie at the heart of decision making about data acquisitions.

Robinson and colleagues (2018) found that vetting researchers was seen as the most important safeguard, but penalties for data security breaches were also supported by more than 90 per cent of survey respondents. Discussing the necessary safeguards if research using personal information was carried out by commercial organisations, half of respondents felt that there should be even more safeguards applied. However, as discussed earlier, when participants were confident that there were (potential) public benefits to research carried out by commercial companies, they were more supportive of the research idea than the research team had expected.

Turning to the question of how long researchers should be allowed to have access to linked data Cameron and colleagues (2014) reported that dialogue participants held one of two views: linked datasets should be stored so that researchers could have access after the initial research project is complete, or they felt that the destruction of data after project completion was reassuring and therefore desirable. A compromise suggestion was archiving linked data with researchers reapplying for permission to use it again.

A survey conducted for the Royal Statistical Society (RSS, 2014) found that:

- The respondents supported transparency, as most agreed with the statement that they “would really like to know what information government knows about me” (only 5% disagreed).
- However, a small minority, of 5 per cent said that they had ever asked a government department, public service or private company what information they held about them, suggesting that the public was unlikely to drive action on transparency.

- The study found more support for the government preventing misuse of personal data rather than a demand for personal control: “It’s too difficult for me to keep control of all my personal data – it should be the government’s job to prevent anyone misusing it.” 40% agreed with this, 25% disagreed.

3.3.6. What attitudes are held by different groups of the public?

Cameron and colleagues (2014) found that particularly older participants and those with non-technical jobs often had little or no awareness of social research or evidence-based policy making as existing practices.

A study by the Wellcome Trust (2013) highlighted how different groups of the public had different concerns about privacy when thinking about the research use of data. Study participants from lower socio-economic classes found it less acceptable than average that individuals would be targeted with specific health-related messages as an outcome of sharing data from their supermarket loyalty card data.

Looking at health data, Aitken and colleagues (2018a) identifies differences between different groups of respondents in their Discrete Choice Experiment discussed earlier. There are slight differences by age, gender and the presence of long term health conditions, but the greatest differences between respondents are found when comparing respondents according to employment status (employed or non-employed). In addition, respondents whose work involves handling or managing data, or are familiar with data security issues because of their work (‘key sector workers, in the study) are found to be more concerned than average with oversight arrangements (16%, compared with 8% for non-key sector respondents) and the purpose of data linkage (22%, compared to 17% for non-key sector respondents). Participants whose work is not related to data security are found to be more concerned with who the researchers are (19%, compared to 12% for key sector workers) and what information was being linked (30%, compared to 25% for key sector workers). However, there was little difference between the two groups in relation to the issue of the management and sharing of profits.

Finally, the respondents who chose the answer ‘Research using linked information should not be allowed under any circumstances’ were more likely to be in older age groups: 42% were aged 55 and over, 34% aged between 35 and 54 and, 24% aged between 18 and 34.

3.4. Public perceptions and understanding of public benefits

Focusing on how the 'public benefit' is conceptualised in the case of the ADRN initiative, Cameron and colleagues (2014) found that dialogue participants did not have a clear definition of what socially beneficial research was. However, they found most of the case studies presented during the workshops (and described above) met that threshold. In general, research that has the potential for clear policy or practice impact was thought to be more socially beneficial than blue skies or theory-led academic research. Analysing data from the same public dialogues that Cameron and colleagues discussed, Pope (2014) reports that 'better, more efficient and cost-effective use of administrative data' (p.32) was also seen as a form of public benefit.

The benefits have to be 'tangible': Castell and colleagues (2014) found that public support for individuals' data being used was strongest when there were tangible public benefits: "88% support the use of people's data to help develop treatment for cancer, 73% support data being used to improve the scheduling of transport services and 70% support data use to prevent crimes" (p. 137). In another part of the study, participants highlighted medicine, education and social care as good areas in which to put personal data to use – Castell and colleagues (2014) explain this by pointing out that in these areas the implications of research are fairly clear.

In the literature on health data, the study by the Wellcome Trust (2013) found that there was public support for data linkage that had the potential to improve social and mental health services by enabling a more joined up, efficient administrative service.

Analysing survey responses, Davidson and colleagues (2013) found that participants conceptualised but did not necessarily define primary and secondary societal benefits. Primary benefits were seen outcomes that arose directly from research and included new and improved products and services – for example, drugs to treat diabetes. In contrast, secondary benefits were broader and longer-term, such as better public health. Davidson and colleagues emphasised that there was a view that secondary benefits would save the government money in the long run (pp. 78-79). The concept of 'secondary beneficiaries' is also relevant here: participants identified two groups who might become beneficiaries of research, namely: other countries, including developing countries where research funding

and infrastructure is limited, and future generations, who could benefit from research conducted now (Davidson et al., 2013, pp. 80-81).

Scott et al. (2018)²³ and Aitken et al. (2018b) argued that there was not an accepted definition of 'public benefit' and it can mean very different things to different people in different contexts. A recent study by Aitken and colleagues (2018b) explored public understandings of public benefit in health research along two questions: Who is the public and how should members of the public benefit? Discussing the first question 'Who is the public', two key considerations emerged: scale and need. Thinking about scale, dialogue participants argued that the public should be considered as inclusively as possible (i.e., benefit the maximum number of people). Thinking about need, dialogue participants acknowledged that research can have greater or lesser potential impact for different groups. This led to recognition of the value of health research being targeted at smaller, vulnerable groups within the public who would potentially benefit the most. In this sense public benefits were conceptualised as benefits to particular smaller groups within the public. Participants added that in the long-term this would be of benefit to everyone in society.

Considering the findings about how members of the public should benefit, Aitken and colleagues (2018b) identified three categories:

- Direct results: cures, increased medicalization, improving lives, with a focus on health improvement and enhanced lifestyle; preventative measures.
- Indirect results or 'knock-on effects': small numbers of vulnerable people benefiting from health research.
- Dialogue participants suggested that increasing scientific knowledge is in and of itself a benefit and that that the public could benefit from greater engagement with the scientific community. Participants considered it essential that policy makers, governments and/or the health service act on research findings in order to realise the potential public benefits. Participants focused especially on assurances that any

²³ The study is based on the findings from workshops run by Involve, Understanding Patient Data and the Carnegie UK Trust. The workshops took place in six local authority areas to examine how the risks and public benefits associated with data sharing are recognised, quantified and evaluated by stakeholders. The study's main aim was to understand the data trade-offs made every day by local government and civil society actors, and the reasons for these decisions.

benefits would be realized and lives would be improved, while the nature of public benefit was a less important consideration.

When participants discussed the (potential) public benefits of concrete research projects, rather than thinking about public benefit in the abstract, three key discussion points emerged (unprompted): who would benefit from the project; which project would have the greatest impact; and to what extent workshop participants expected that the potential benefits would be realised. Looking at 'the greatest impact' in more detail, participants considered two factors: which project would benefit the most people and which would have the biggest benefit for those in greatest need.

Analyzing the public's views, Aitken and colleagues identified three groups who should benefit from data linkage: *individuals*, *specific groups* and to *society* more widely. Another important finding is that participants considered it very important that the mechanisms of research impact should be in place and that this is demonstrated. Finally, the study found that no one spoke of benefits in terms of economic benefit.

3.5. What do people expect from public engagement?

Castell and colleagues (2014) found that the public wanted scientists to talk more about their work, particularly about the social and ethical implications. People wanted a two-way communication and overwhelmingly thought that scientists, regulators and government should engage in public dialogue about science, with three in ten respondents saying they wanted to have more of a say on science issues. However, the majority thought that 'experts' and not the public should advise the Government on science issues. People did not think the Government was doing enough to consult the public on science – the study suggested that this opinion was based on widespread public cynicism about consultation events.

Green and colleagues (2015) argued that the disparity between public support for specific research ideas and public concerns about data sharing and linking in general can be explained by the phenomenon that people tend to look more favourably on the things they have some personal experience of and tend to rely on media reports when they do not have such experience.

Aitken and colleagues (2016a) also found that there was public appetite for more information about how data are used in research and the safeguards that are in place to protect against misuse, abuse or harms. They also found that studies which used deliberative methods and gave participants information about research practices led to greater support and acceptance, or less concern about research uses of data. Most studies included in the review reported that participants wanted more information about research. The study by Van Mil & Hopkins (2015) found that participants were also keen to see education programmes for children and their parents on big data and the work of the Data Centres, which indicates a desire for broader information campaigns and awareness raising activities, beyond the remit of the Data Centres.

While participants of the ADRN dialogue workshops acknowledged that their attitudes to data linking had changed as a result of the dialogue (Cameron et al., 2014), some felt that there was no need for communications or public engagement, because, in their own words: *'It's just too complicated'*, *'It might panic people unnecessarily'* and *'Telling everyone would be costly.'* Others felt that not informing the public about the work of the ADRN would lead to suspicion and that some members of the public would be interested in data linking and would understand the process. This group of participants recommended that the research community should be as transparent with the public as possible, and information about de-identification, data security measures and public benefits should be in focus. Somewhat surprisingly, this group of dialogue participants also felt that the amount of 'active communications' with the public should be kept to a minimum.

It is highlighted in the literature that public engagement activities can be lengthy, and it is difficult to devise successful awareness raising methods that convey accurate and impartial information. Dialogue members also acknowledged that there was hard work involved in the dialogues, as the following quote demonstrates: *"It's taken the best part of 8 hours to get (data linking) across to us, who are actively interested. It'll be more difficult for a member of the public"* (participant in Cardiff, quoted in Cuthill, 2014).

4. Discussion and Conclusions

This literature review has explored the growing body of literature on public attitudes to data and data science practices. While the main focus of the literature review was on social-

economic research, studies on public attitudes towards linking health data were also discussed.

The first research question asked under what conditions the public supports accessing and linking data about people from various sources. In particular the review addresses: What is understood by the phrase 'social and economic research' and about the need to access personal data to undertake such research? Under what conditions is such research seen to be in the public interest? What types of data are more or less sensitive when re-used for research purposes?

The review identified the following key conditions of support: there must be a provable, sufficient, clearly articulated and legitimate purpose to the research, which is seen to benefit the public. The question of which specific groups of the public will benefit is also relevant here. The second key condition is that issues around privacy and confidentiality and the level of de-identification (often referred to as 'anonymity') are addressed and clearly explained to the public. Public trust in the researchers and research institution who are involved in data linkage and the use of data is the third key condition. Finally, issues around data security, in particular safeguarding data against misuse and data governance were identified as a key condition of public support. Slightly less important, however, still relevant conditions were: respect for the autonomy of the public, consent, how the profits emerging from the research were managed and whether the private sector was involved and finally, what types of data were linked together.

Trust has emerged as one of the key conditions of public support for data linkage, and in Section 3.2 of this report the concepts of trust and trustworthiness and the related concepts of transparency and authenticity were explored and discussed. The relationship between trust, transparency and authenticity was argued to be complex: transparency is necessary to build trust, but trust is required in order for the transparency to be recognised as adequate (Brown and Michael, 2002, quoted in Aitken et al., 2016b). Aitken and colleagues (2016b) argued that this circle cannot be broken by more transparency but rather by more 'authenticity'. In a finding that is very relevant to this literature review, Aitken and colleagues argued that researchers and decision-makers can develop authenticity by demonstrating that they have engaged with the public and attempted to incorporate and address different views to such an extent that it has caused them distress or suffering.

The third research question explored public acceptability for the use of survey data, with a particular focus on longitudinal surveys, and administrative data linked together.

This section further explored the public acceptability of using data from different sources linked together along the key themes of: privacy and ethics, consent and control, data ownership and policies and safeguards. Taking the issues discussed in the previous section further, this section was focused closely on *how* the public thinks research using linked data should be carried out.

Looking at the different types of data, it was found that data related to mental and sexual health, sexuality and religion were particularly sensitive. Individual control over how data was used ('perceived autonomy') was highlighted to be a key issue, however, there was no clear consensus about what the need for 'control' implies and how researchers should respond to it. Debates about control and consent were linked to the issues of trust, in particular, in the public's trust in the ability of research organisations to keep their data safe.

Audrey and colleagues (2016) argued that would be challenging to offer solutions that are acceptable to the majority of the 'general public' and they called for pragmatic, imaginative and flexible approaches to consent to data linkage, while emphasising that protecting the public trust in the research process was essential.

The final issue identified in this section of the literature review was that of community-level control: it was argued to be an alternative to individual consent. Mulrine and colleagues (2018) were experimenting with community-level controls in their work on the Great North Care Record Cities – the outcomes of this work will be important future additions to the findings presented in this literature review.

In the final section, the literature was explored to answer two related questions: how the publics make sense of the potential (public) benefits of social-economic research and data linkage in particular and what the publics expect of public engagement activities in the context of data sharing and linkage. Unsurprisingly, there was not an accepted definition of 'public benefit' and the question 'Who is the public' was also asked in the literature.

Studies based on public engagement found that the publics argued that the definition of the public should be as inclusive as possible (i.e., research and data linkage should benefit the maximum number of people). This finding was not in contradiction to the argument

emerging from the health literature, where public benefits were also conceptualised as benefits to particular smaller groups within the public in the short term.

Another important finding was that participants in public engagement suggested that increasing scientific knowledge was in and of itself a benefit and that the public could benefit from greater engagement with the scientific community. Participants considered it essential that policy makers, governments and the health service act on research findings in order to realise the potential public benefits. Participants focused especially on assurances that any benefits would be realized and lives would be improved, while the nature of public benefit was a less important consideration.

Finally, there was consensus in the literature that the public want more 'two-way' communication about data research, particularly about the social and ethical implications and the safeguards that are in place to protect against misuse, abuse or harms. However, it was also found that the public wanted 'experts' to advise the Government on science issues. They also found that studies which used deliberative methods and gave participants information about research practices led to greater support and acceptance or at a minimum, less concern about the use of data for research purposes.

5. References

- ADRN Scotland (2017) *ADRN 2017 Session 2A Public engagement and value of administrative data*. Retrieved from: <https://www.youtube.com/watch?v=yFTJpqDgNjI>.
- Aitken, M., Jorre, J. d. S., Pagliari, C., Jepson, R., & Cunningham-Burley, S. (2016a). 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., Cunningham-Burley, S., & Pagliari, C. (2016b). Moving from Trust to Trustworthiness: Experiences of public engagement in the Scottish Health Informatics Programme. *Science and Public Policy*.
- Aitken, M., McAteer, G., Davidson, S., Frostick, C., & Cunningham-Burley, S. (2018a). Public Preferences regarding Data Linkage for Health Research: A Discrete Choice Experiment. *International Journal of Population Data Science*, *3*.
- Aitken, M., Porteous, C., Creamer, E., & Cunningham-Burley, S. (2018b). Who benefits and how? Public expectations of public benefits from data-intensive health research. *Big Data & Society*, *2018*, 1-12. doi:10.1177/2053951718816724.
- Al Baghal, T. (2016). Obtaining data linkage consent for children: factors influencing outcomes and potential biases. *International Journal of Social Research Methodology*, *19*(6), 623–643.
- Audrey, S., Brown, L., Campbell, R., Boyd, A., & Macleod, J. (2016). Young people's views about consenting to data linkage: findings from the PEARL qualitative study. *BMC Medical Research Methodology*, *16*(34).
- Boyd, A., Tilling, K., Cornish, R., Davies, A., Humphries, K., & Macleod, J. (2015). Professionally designed information materials and telephone reminders improved consent response rates: evidence from an RCT nested within a cohort study. *Journal of Clinical Epidemiology* *68*(68), 877-887. doi:http://dx.doi.org/10.1016/j.jclinepi.2015.03.014.
- Cameron, D., Pope, S., & Clemence, M. (2014). *Exploring the public's views on using administrative data for research purposes*. Ipsos Mori. Retrieved from: <https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/>.
- Carter, P., Laurie, G. T., & Dixon-Woods, M. (2015). The social licence for research: why care.data ran into trouble. *Journal of Medical Ethics* (41), 404–409.
- Cuthill, V. (2014). *Why the dialogues? Reflections from the ESRC*. Retrieved from: [xxx]
- Data Futures Partnership. (2016). *Exploring Social Licence*. Data Futures Partnership. Retrieved from: <http://datafutures.co.nz/our-work-2/talking-to-new-zealanders/>.
- Data Futures Partnership (2017) The path to social licence. Data Futures Partnership. Retrieved from: <https://trusteddata.co.nz/wp-content/uploads/2017/08/Background->

[Trusted-Data.pdf](#).

Davidson, S., McLean, C., Treanor, S., Aitken, M., Cunningham-Burley, S., Laurie, G., . . . Sethi, N. (2013). *Public Acceptability of Data Sharing Between the Public, Private and Third Sectors for Research Purposes*.

Green, E., Ritchie, F., Webber, D., Mytton, J., Deave, T., Montgomery, A., . . . Chowdhury, S. (2015). *Enabling Data Linkage to Maximise the Value of Public Health Research Data. Final report to The Wellcome Trust*. Retrieved from <https://wellcome.ac.uk/sites/default/files/enabling-data-linkage-to-maximise-value-of-public-health-research-data-phrdf-mar15.pdf>

Hobbs, J., Tully, M., & Ali, I. (2018). Exploring public attitudes towards data sharing and linkage in health and social care. *International Journal of Population Data Science*, 3(3). doi: <http://dx.doi.org/10.23889/ijpds.v3i4.932>.

Ipsos MORI. (2016). *The One-Way Mirror: Public attitudes to commercial access to health data*. Retrieved from <https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf>.

Mostafa, T. (2014). *Variation within Households in Consent to Link Survey Data to Administrative Records: Evidence from the UK Millennium Cohort Study*.

Mulrine, S., Murtagh, M., & Minion, J. (2018). *Great north care record - public engagement report*. [link]

NatCen. (2018). *Public Attitudes to Data Linkage: A report prepared for University College London by NatCen Social Research*. Retrieved from <https://www.bl.uk/collection-items/public-attitudes-to-data-linkage-natcen-social-research>.

OECD. (2016). *Research Ethics and New Forms of Data for Social and Economic Research*. Retrieved from: <https://www.oecd-ilibrary.org/docserver/5jln7vnpxs32-en.pdf?expires=1552478809&id=id&accname=guest&checksum=593B209A3786219B885B0AFE0DBF77FA>.

Pope, S. (2014). *Public views on using government administrative data for research: public dialogue method and findings*. Ipsos Mori. Retrieved from: <https://esrc.ukri.org/files/public-engagement/public-dialogues/public-views-on-using-government-administrative-data-for-research-public-dialogue-method-and-findings-ipsos-mori/>.

Van Mil, A. & Hopkins, H. (2015). *Big Data: Public views on the use of private sector data for social research: A Findings Report for the Economic and Social Research Council*.

Robinson, G., Dolk, H., Dowds, L., Given, J. E., Kane, F., & Nelson, E. (2018). *Public Attitudes to Data Sharing in Northern Ireland: Findings from the 2015 Northern Ireland Life and Times survey. Technical report*.

Royal Statistical Society (RSS). (2014). *Public attitudes to the use and sharing of their data. Research for the Royal Statistical Society by Ipsos MORI*. Retrieved from

<https://www.statslife.org.uk/news/1672-new-rss-research-finds-data-trust-deficit-with-lessons-for-policymakers>.

Scott, K., Burall, S., Perrin, N., Shelton, P., White, D., Irvine, G., & Grant, A. (2018). *Data for public benefit: Balancing the risks and benefits of data sharing*. Retrieved from https://understandingpatientdata.org.uk/sites/default/files/2018-04/Data%20for%20public%20good_0.pdf.

Stockdale, J., Cassell, J., & Ford, E. (2018). "Giving something back": A systematic review and ethical enquiry of public opinions on the use of patient data for research in the United Kingdom and the Republic of Ireland. *Wellcome Open Research*, 3(6).

Wellcome Trust. (2013). *Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data*. Retrieved from https://wellcome.ac.uk/sites/default/files/wtp053205_0.pdf.