DIALOGUE ON DATA

Exploring the public's views
on using administrative data
for research purposes

Daniel Cameron
Sarah Pope
Michael Clemence
Ipsos MORI Social Research Institute

Office for
National Statistics

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Table of contents

Executive summary

This report describes the findings from a public dialogue on administrative data commissioned by the Economic and Social Research Council (ESRC) and the Office for National Statistics (ONS). The overall objectives were to explore public understanding and views of administrative data and data linking.[1] The dialogue focused on two uses of administrative data, one that is currently being established and one that may go ahead in future:

- The new ESRC-funded Administrative Data Research Network (ADRN) that was set up in late 2013.[2]
- The potential use of administrative data linking as one of the options for conducting the 2021 census (alongside an annual survey).[3]

The findings and recommendations from this project will help inform the ESRC and ONS as they develop administrative data linking, and particularly their strategies around public engagement and communications, as well as the Administrative Data Research Network (ADRN) and Administrative Data Service (ADS) as they develop their public engagement and communications strategies. This report focusses on the findings from the dialogue that relate to the ADRN.

The dialogue consisted of seven sets of workshops (14 in all) with members of the public and sector experts. These took place between the 19th of October and the 9th of November 2013 in England (London, King's Lynn, and Manchester), Wales (Cardiff, Wrexham), Scotland (Stirling) and Northern Ireland (Belfast). In each location, a group of 14-20 public participants attended an all-day workshop on a Saturday, and reconvened two weeks later for a second day-long workshop. At least two experts attended each of the workshops. In all, 136 members of the public and 20 experts attended the dialogue workshops.

The value social research and data

Participants generally had very low initial awareness and understanding of social research. While knowledge increased over the course of the dialogue, some found aspects of the subject matter complicated and difficult throughout the discussions. At the beginning of the dialogues, low awareness of the uses of social research drove scepticism about its value. This scepticism was challenged by reading case studies and hearing from working social scientists about the numerous uses and impacts of social science research.

However, negative perceptions of social research as a whole were sometimes a driver of views of administrative data linking and the ADRN. Participants who held these negative perceptions compared research findings to *"common sense"* and questioned why they need to be evidenced. Later in the dialogues, when participants had learned more about the aims and methods of social research, they tended to be more positive about its value. Even so, blue skies or theory-led academic

---

[1] Administrative data is the data about individuals that is collected for operational purposes in the delivery of government services, but that also has the potential to be used in numerous social research projects.

[2] The ADRN consists of four Administrative Data Research Centres (ADRCs) that will act as national hubs for researcher access to administrative data, and an overarching service that will help these centres to function, known as the a Administrative Data Service (ADS).
3 The government has asked the ONS to explore alternatives to a full paper census in 2021. They are due to recommend their preferred option in spring 2014. The other front-running option is a full online census.

research continued to be valued far less than research perceived as socially beneficial (i.e. research that has the potential for clear policy or practice impact).

Participants did not have a clear or shared definition of 'socially beneficial'. However, while deliberating across the workshops, and when speaking to working social scientists, participants were generally very open as to what counted as social value, and thought that most of the real life examples presented across the workshops met that threshold. However their initial views suggest that the social value of research is not well understood by the public.

By contrast with their low awareness of social research, 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.

Nonetheless, keeping their personal data secure was very important to them, and they worried about their data being leaked, lost, shared or sold by organisations that hold it. Participants also felt that they had little control over their personal data. These general concerns about data security more widely drove particular security fears with relation to administrative data linking.

Participant views on the Administrative Data Research Network

Participants initially questioned why the ADRN was necessary, as they had assumed that government administrative data is already linked and shared across departments and services. Indeed they were somewhat reassured by the fact that there are so many barriers to sharing currently in place.

As they were shown information about research projects that have used linked administrative data, asked questions of experts, and heard the opinions of other participants, their opinions of the ADRN plans changed, sometimes moving between negative and positive perceptions several times. The main issues that influenced their views are outlined below.

The most common concerns about the ADRN initiative were:

- De-identification as a concept, and the possibility of re-identifying individuals in particular datasets
- De-identification in practice, including the details of the process and the destruction of personal data
- Security and who can gain access to linked datasets
- Cost and whether the ADRN initiative represent a good use of public money

However, there was also a range of features of the ADRN initiative that participants found reassuring, including:

- The independence of the Administrative Data Service (ADS) and Administrative Data Research Centres(ADRC), and the fact that they are located in universities
- The potential for linked data being used to allocate funding effectively, and improve public services
- The potential for more efficient use of data and therefore cost savings for government

- Restrictions on who can access linked data
- The ADRN providing a systematic way to regulate administrative data linking
- No plans for a 'super' database containing multiple linked data sources

Those participants who were familiar with social research and how it is used tended to trust researchers' motivations more, and therefore had fewer concerns about the ADRN initiative. Others who started the day with low awareness of social research (and therefore low trust) found discussing the issues and speaking to experts interesting and reassuring.

By the end of the second day of dialogue, most participants had concluded that they supported administrative data linkage via the ADRN, if three main conditions were met:

- The data is fully de-identified as per the process described in the workshops[4]
- The data is kept secure at all times
- The data is linked for socially beneficial purposes

While the other concerns and reassurance points presented were important, these issues were the key to broad understanding and support of the ADRN. In many cases, participants needed extensive information and discussions with experts and researchers in order to be satisfied that these conditions would be met by the ADRN.

Principles and rules for Administrative Data Research Centres

Participants were generally content with the current set of suggested principles and rules around accessing linked administrative data through the ADRN. A crucial consideration for participants was creating safeguards to ensure that all data is secure. Even where they were convinced that the linked administrative data would be fully de-identified – and not all were – this usually meant that they were strongly in favour of secure physical settings and against remote access. Those who would allow remote access were those who thought that de-identified data would not be of interest to anyone but a social scientist working on the particular question it was linked for.

Many participants deferred to experts in terms of the specific rules or procedures within the ADRCs. Once they were satisfied that the process would be rigorous, and that approval is granted by independent, qualified people, then they were happy for project by project decisions to be made on access, what data can be linked and how long the linked data can be held.

The main exception to this willingness to defer to experts was in relation to the involvement of businesses. Participants argued for strict and specific rules because of concerns about profits being made from linking personal data. As this issue was not the focus of this dialogue, more research would be needed to fully understand public attitudes towards allowing businesses to link administrative data, or allowing business data to be linked to government data.

---

[4] Participants generally understood that it is impossible to guarantee that individuals cannot be re-identified within linked datasets, but most were of the opinion that the processes and safeguards in place within the ADRCs limit this risk sufficiently, taking into account the type of data that will be linked.

Information and engagement

Based on their experience of the dialogue, many participants were sceptical of the value of informing the general public about the ADRN initiative. Often they were worried that the general public would not understand such a complex topic through simple messages, and thus would be become worried about data security and privacy when there is not necessarily a reason to be. However, they thought that if the ADS and ADRCs intend to keep the public informed about their work, then the key messages to stress should be:

- How the data is being used to benefit society
- The de-identified nature of the data
- The fact that data linking is currently possible, and the ADRN makes the process simpler and more efficient
- The fact that commercial companies cannot gain access to the data

Some participants thought that members of the public with an interest in the topic should be involved in decision making within the ADRCs. For many, it would be enough for ADRCs to be transparent and make information available to those who seek it out, rather than proactively engaging or informing the public.

Recommendations

The key policy and communications recommendations following this dialogue are summarised below.

Policy:

- The ADRN and ADRCs should consider requesting that researchers who want to access linked data submit concrete plans for publishing their work, and disseminating it to relevant policy makers.
- High levels of public concern around security should be reflected in ADRCs' and ADRN operational plans. For example, any move towards virtual access to this data would need to be clearly demonstrated to be as secure as access through a secure physical setting.
- Bearing in mind that satisfaction with the ADRN initiative seemed to be linked to confidence in the de-identification process, work should be undertaken to develop the best way to explain de-identification in a way that people readily understand. The visual technique used in the dialogues was useful in helping participants conceptualise the process, which suggests that a video or animation may be a useful way to communicate around this issue.
- Further public dialogue or research would be needed for any expansion of the ADRNs remit. Specifically, further research should be done to understand the public's views on allowing businesses to access linked administrative data.

Communications and engagement:

- The ADRN, ADRCs and social researchers should bear in mind the low familiarity and awareness of the general public when communicating their work, and remember that they may always need to explain, at the broadest level, why their work is necessary.

- Transparency is key, while high-profile public information campaigns are unnecessary. Ensuring public representation in the decision-making process could be important in proving the transparency of the ADRC and ADRN.

- General public communications about the ADRN and the ADRCs should focus on the societal benefits of social research through developing strong case studies, showing how the projects they have enabled have led to policy change or service improvement.

Chapter 1: Introduction and methodology

Ipsos MORI was commissioned by the Economic and Social Research Council (ESRC) and the Office for National Statistics (ONS) to undertake public dialogue to explore public understanding and opinion of administrative data and data linking.

This dialogue is a sub-set of the wider Public Attitudes to Science (PAS) 2014 project being undertaken by Ipsos MORI for the UK Government Department of Business, Innovation and Skills (BIS) and the Economic and Social Research Council (ESRC). The main element of this project was a random probability survey with 1,749 UK adults aged 16+ and a booster survey of 315 16-24 year-olds. One of the topic areas for this survey was public opinion and attitudes towards 'big data'. This is an emerging area in science and policy, yet public knowledge and opinion on it is not well known.[5] The PAS survey questions asked about on numerous potential uses of big data, including operational and research purposes.

The PAS findings show that, on balance, the public oppose personal data being used for commercial gain. While a majority seem to be relatively unconcerned about the use of their records in 'big data' analysis[6], there is strong opposition to some of the specific ways in which *private* companies might operationalise this data. For example, 62% of people oppose websites using people's online browsing histories to create personalised adverts for products that people are more likely to be interested in.

By contrast, they largely support the use of personal data in contexts where there is a tangible public benefit, such as in medicine, transport and policing. Most people (56%) support combining the data held by multiple government departments and using them to better tailor public services to individuals. This question does not however tell us whether people support this use of government administrative data for research, for operational purposes, or for both.

In addition to collecting nationally representative quantitative data on the topic, a need was recognised for more in depth, qualitative research about public attitudes towards the potential uses of government 'big data' i.e. the large administrative datasets that are held by government departments and agencies. Both ONS, who are considering the application of big data to population statistics, and ESRC, who are setting up a national big data network for administrative data, have specific policy and communications challenges relating to public views of big data that were best met by a dialogue approach.

The dialogue consisted of a series of reconvened workshops across the United Kingdom. Its principal aim was to examine public understanding of administrative data, and to uncover attitudes towards the linking of government records for the purposes of research. It represents a starting point in what will be an on-going public engagement strategy by Administrative Data Research Network and the ESRC to ensure that the public are kept engaged with and informed about their work with administrative data.

---

[5] The PAS survey includes questions about people's attitudes to 'data' across private and public organisations. The topline results are included in the methodological appendix.

[6] 61% agree with the statement "I don't mind how data collected about me is used, as long as it's anonymised and can't be linked back to me"

1.1 Background

ONS and the ESRC (on behalf of the Administrative Data Research Network (ADRN)), share a common goal of exploring in more detail public perceptions of administrative data and data linking. Part of the remit of each of the centres that make up the newly-formed ADRN is to build public trust in the data sharing process and the work carried out through the network.

Administrative Data Research Network

Administrative data is the data about individuals that is collected for operational purposes in the delivery of government services, but that also has the potential to be used in numerous social research projects. Although the technology now exists to link and effectively analyse this data, it has tended to be underutilised and under-shared in the past for a number of reasons[7]:

- Legal issues around the protection of individual's information when using and storing data
- The length of time that it takes to create agreements to share administrative data between sources
- The lack of a joined up system to connect different sources of administrative data

As part of the Government's £64million investment in big data technology, the Economic and Social Research Council (ESRC) has funded four Administrative Data Research Centres (ADRCs) in the United Kingdom to act as national hubs for linking and anonymising administrative data. They are also funding the Administrative Data Service (ADS), an overarching service that will help these centres to function.

The ADRCs were established in October 2013 at four universities around the UK – Southampton, Swansea, Queens University Belfast, and Edinburgh – and the ADS has also been set up at the University of Essex. The purpose of these centres is to facilitate research based on linked, routinely collected administrative data, by:

- Providing state of the art facilities for access to de-identified administrative data by accredited researchers.
- commissioning and creating new linked administrative data resources for a growing research agenda
- conducting original research using linked administrative data
- engaging in training, capacity building and public engagement

At the time of the dialogues the Administrative Data Research Network and the centres that make it up are still in their early stages; they have recruited staff but have yet to commence operations. The ADRCs will be funded initially for five years. The findings and recommendations from this project will help inform strategy and operations of the ADRCs, and help inform their ideas around public engagement and communications.

---

[7] A greater exploration of the issues that have prevented greater use of administrative data can be found in the Administrative Data Taskforce's report "*Improving Access for Research and Policy*", available here: http://www.esrc.ac.uk/_images/ADT-Improving-Access-for-Research-and-Policy_tcm8-24462.pdf

1.2 Steering group

This dialogue was overseen by a steering group of experts who have been involved at all stages of the project design and execution. They have met three times through the course of the project: initially to set objectives and scope for the project; the second to input on and sign off material for the dialogues; and the third time to discuss the findings of this report.

Members of the steering group also attended many of the events to participate in the discussions.

The steering group members were:

- **David Walker (chair)**, Editor at Public Leaders Network
- **Melanie Knetsch (secretary)**, Deputy Head of Communications, ESRC
- **Genevieve Groom**, Senior Research Officer, ONS
- **Liesbet van Zoonen**, Professor at Loughborough University
- **Mary Hickman**, Independent research consultant and member of ESRC's Methods and Infrastructure Committee
- **Sarah Cunningham-Burley**, Professor at the University of Edinburgh
- **Vanessa Cuthill**, Team Head , ESRC
- **David Carr**, Policy Advisor at Wellcome Trust
- **Jane Naylor**, Methodology, Office for National Statistics
- **Daniel Start**, Design and Engagement Specialist, Sciencewise-ERC
- **Kerry Seelhoff**, PAS 2014 project manager, Department for Business, Innovation and Skills
- **Maria Sigala**, Senior Policy Manager, ESRC

1.3 Objectives

The aims of this research project are to examine public views around the use of government administrative data in in social research. It also seeks to investigate what sort of procedures should be set in place, and the concepts and language that need to be used to reassure the public about the safety and security of their public records in the data linking process.[8]

These overall aims were broken down into four objectives:

- **To better understand the cultural barriers to linking administrative data**, and work out how the process by which the data is linked affects attitudes.
  Specifically, to explore how the following aspects of the linkage process affect views:
  - o Why - The end use of data (administrative/statistical/operational) and people's ability to understand the difference between different potential uses
  - o What – The type of data being linked; with a particular focus on more personal data (those fields mentioned under the 2010 Equalities Act – marital status, income, sexuality, etc.) and 'cross-sector' data linking (e.g. matching records from public and private sector providers).

---

[8] Findings about the language and concepts that people do and do not understand have been embedded throughout the report and conclusions.

- o Who – The identity/qualifications/sector/motivations of the people carrying out the linking
- o Where – The impact of the environment in which the linkage is to take place
- o When – The length of time over which data about an individual is to be held, or period of time over which it is acceptable to draw administrative data
- **To begin the process of creating a language around government administrative data and data linking that is meaningful and accessible to the public.**
  - o What sort of assurances are required, and what sort of language is necessary, to reassure the public on data linking and to earn their trust?
- **To test the public perceptions of the rules that ESRC ADRCs will be subject to and to provide the ADRCs with data on public attitudes and appetite for engagement** in order to help frame their strategies in this area.
  - o How much do people want/need to know in order to be reassured and engaged? How should any arrangements be future-proofed?
  - o Are there demographic differences in the level of interest in, opposition to, and engagement with data linking?
  - o Who needs to be prioritised for engagement, why and how?

1.4 Methodology[9]

Approach and overall methodology

Six sets of reconvened six-hour public dialogue workshops, each involving 16 to 20 participants, were conducted in London, Manchester, King's Lynn, Cardiff, Wrexham, Stirling and Belfast between 6 October and 9 November 2013. A total of 129 participants attended the seven reconvened workshops.

A qualitative, deliberative approach was considered the best way to allow participants to explore this topic, from both a personal and a citizen perspective. Qualitative methods allow participants the freedom to express the issues that are salient to them and develop their views in the light of discussion and debate. A reconvened approach allowed participants enough time to digest the complex information that they received on the first day, and reflect on the topic outside of the dialogue environment.

Further dialogue was introduced through the attendance of 'experts' at each of the events. These experts described their work, answered participants' questions, and engaged in debate about the ADRN and the proposed changes to the ONS. Their specific role at each event is described in more detail below.

Participants were recruited on-street. Quotas for gender, age, socio-economic group and ethnicity were set to ensure participation of individuals from a range of backgrounds reflective of the areas they came from and the broad diversity of the UK population. Soft quotas based on trust in researchers, views on data linking and use of data related technology, again to ensure a broad range of attitudes within each of the workshops. The exceptions were King's Lynn where all participants

---

[9] Please see the Dialogue on Data: Methodological Annexe (published separately) for further methodological details.

were aged under 35, and Cardiff, where all participants were aged over 55. These age-specific workshops were included as researchers hypothesised that age would have a strong influence on views on this topic.

A final two and a half hour reconvened meeting was conducted on 20 November 2013 in London, the aim of which was to involve participants in the analysis of the findings from all three workshops, and allow a chance for further dialogue with representatives from the ESRC, ADS and ONS. A small number of participants from the London, King's Lynn and Cardiff workshops were invited to attend this workshop, which ultimately involved 8 participants. Participants were chosen for their active engagement at the initial workshop and to ensure a range of views on the issues were reflected.

Materials and data collection

The workshops were deliberative and dialogic in nature, so that participants were given information about social research and the ADRN and access to expert views about the issues covered.

During the workshops, participants were encouraged to think about data, social research, data linking and rules that should be in place for data linking. They were given time and structured exercises to support exploration of the topic and expression of views and underlying values.

Ipsos MORI worked with the ESRC and the ONS to design materials for the workshops that would facilitate discussions, and ensure that participants were getting the right amount of information at the right level of detail at different points in the workshops. Facilitators followed a discussion guide throughout the two days to ensure that the same topics were covered in all locations. All materials were reviewed by the steering group at an early stage, and signed off after several iterations by ESRC and ONS. They are included in the Annexe 2.

The experts at the workshops were invited to engage with participants, answer questions and join in discussions as and when they wanted to, except for during initial spontaneous discussions about different topics. Activities that participants engaged with throughout the two dialogue workshops included:

- **Spontaneous group discussion:** This allowed for discussion of top of mind views about data, data linking and social research.
- **Presentations:** These introduced participants to the key information they needed to discuss the ADRN.
- **Case study considerations:** Several examples of research projects where linked administrative data has been helpful were used to prompt discussion of the uses, benefits and drawbacks of this type of research.
- **Data linking 'journey' visualisation:** Physical prompt cards and boxes were used to help participants visualise the physical 'journey' that data makes between the person from whom it is collected and the researcher examining linked datasets.
- **'Ask a researcher' session:** After lunch at the first workshop, participants were given a chance to hear about researchers work and ask any questions they wanted about data linking, or research more widely.
- **Design task:** Participants were given a pro-forma and asked to design the ideal rules for the ADRCs.

At several points participants were encouraged to note, record and contribute their views independently before discussing them as a group to ensure individual views were captured.

A note on interpretation of qualitative data

Qualitative research approaches (including public dialogue workshops) are used to shed light on *why* people hold particular views, rather than *how many* people hold those views. It is used to explore the contours of peoples' views, and what factors seem to shape or underlie them. Qualitative research allows us to explore the diversity of peoples' views and recognise that views may not be resolute.

The results are intended to be *illustrative* rather than statistically reliable and, as such, do not permit statements to be made about the extent to which something is happening. Given the qualitative nature of the data collected from the dialogue, this report aims to provide detailed and exploratory findings that give *insight* into the perceptions, thoughts and feelings of people, rather than statistical evidence from a representative sample.

It is not always possible in qualitative research to provide a precise or useful indication of the prevalence of a certain view, due to the relatively small number of participants generally involved (as compared with the larger respondent bases involved with quantitative studies). So, the views of *proportions* of the qualitative group should not be extrapolated to the population at large. Sometimes, ideas can be mentioned a number of times in a discussion, and yet hide the true drivers of thoughts or behaviours; or a minority view can, in analysis, turn out to express an important emergent view or trend. The value of qualitative work is to identify the issues which bear future investigation. Therefore we use different analysis techniques to identify how important an idea is.

In reporting we state the *strength* of feeling about a particular point rather than the number of people who have expressed that thought. Having said this, is it sometimes useful to note which ideas were discussed most by participants, so we also favour phrases such as "a few" or "a limited number" to reflect views which were mentioned infrequently and "many" or "most" when views are more frequently expressed. Where views apply only to a subset of participants, e.g. participants in King's Lynn, we have highlighted this in the text, as this may indicate differences by region or age. Any proportions used in our reporting (e.g. a 'couple' or 'handful' of participants), should always be considered indicative, rather than exact.

Another consideration in the interpretation of qualitative data is the role of *perceptions*. Different views on an issue make up a considerable proportion of the evidence presented in this study. It is therefore important to bear in mind that although these perceptions may not always be factually accurate, they represent the truth for those who hold these views.

Verbatim comments have been included in this report to illustrate and highlight key points, i.e. those views either shared by a large number of participants or reflecting the strong views of a smaller subset. Where verbatim quotes are used, they have been anonymised and attributed by location, e.g. London.

1.5 Report Structure

The remainder of the report is divided into the following chapters:

Chapter 2: The value of data and social research

This chapter outlines participants' broad views on data and social research. These views were addressed directly in the first half of the first dialogue workshop, and were discussed in more detail throughout the rest of the workshops. The perceptions outlined here are important for understanding the more specific findings in the rest of the report, as general attitudes towards data and use of data in research underpinned attitudes towards the ADRN.

Chapter 3: Views of Administrative Data Research Network initiative

This chapter describes participants' views about current plans for the Administrative Data Research Network. It outlines spontaneous views towards the ADRN, before going on to examine the main areas where participants required further clarification, issues of continued concern, and points that reassured participants across the two dialogue workshops.

Chapter 4: Proposed principles for Administrative Data Research Centres

This chapter describes participants' post-deliberation views on the principles and specific rules that should govern access to linked data. These rules represent the result of two workshops spent digesting information, dialogue with experts and debates among participants.

Chapter 5: Information and Engagement

This chapter examines the information that participants found to be most useful in explaining the ADRN initiative. It also covers particular terms, concepts or messages that concerned or reassured people about the role of ADRCs, and examines participant views on the most appropriate level of engagement and information for the wider public. Finally, it outlines preferences for how the ADRCs, ADS and ESRC should engage with the public.

Chapter 6: Conclusions and recommendations

This chapter outlines the most important findings from the dialogue process, and the lessons that the ADRCs, ADS, ESRC and ONS can draw from them.

Chapter 2: The value of data and social research

This chapter outlines participants' broad views on data and social research. These views were addressed directly in the first half of the first dialogue workshop, and were discussed in more detail throughout the rest of the workshops. The perceptions outlined here are important for understanding the more specific findings in the rest of the report, as general attitudes towards data and use of data in research underpinned attitudes towards the ADRN initiative.

Key points

- Few participants placed a strong value on social research in and of itself. While they wanted the benefits of societal understanding, their lack of familiarity with the process and methods of social research resulted in concerns about wasted effort.
- This lack of intrinsic support for social research affected how participants reacted to the rest of the information they received throughout the day, and the questions they asked. The key concern that recurred throughout the day was whether social research actually leads to social value.
- Personal data security was very important to participants, and this framed much of the discussion. They were particularly concerned about identity theft, and personal data being sold on to other organisations.
- Overall, participants trusted government's intentions more than commercial companies on data security and protection from data misuse (such as selling information on to third parties) or fraud.
- Participants commonly assumed that governmental administrative data is already linked and shared across departments, and supported this for operational uses.
- Views of ownership of data and data security drove many of the more detailed views on the ADRN outlined below.

2.2 The value of social research

Few participants placed a strong value on social research in and of itself. While they wanted the benefits of societal understanding, their lack of familiarity with the process and methods of social research resulted in concerns about wasted effort, and their cynicism about government use of data made them worried about misuse of research findings. Several of the expert attendees at the workshops commented that this was the key lesson for them from the dialogue, above and beyond the detailed views of the ADRN.

This lack of intrinsic support for social research affected how participants reacted to the rest of the information they received throughout the day, and the questions they asked. The key concern that continued to recur was whether social research actually leads to social value. Where it does not, or cannot be proved to, participants tended to see social research as, at best, of questionable value and, at worst, as *"pointless"*. This has many implications for how the ADS, ADRCs and the ESRC should communicate about the work of the ADRN.

Awareness of social research

One of the key themes across the workshops was participants' very low awareness and understanding of social research. Particularly among certain groups (including older participants and those with non-technical jobs) there was little or no awareness of social research or evidence-based policy making as existing practices.

While this was to be expected at the start of the first workshop, many participants continued to question the purpose of social research (and thus devalue it as an enterprise) across both days of dialogue, even after several definitions, recaps and explanations from working social researchers. As discussed earlier in this chapter, research uses of data were often confused with operational uses.

Participants were given four case studies of projects where administrative data has been linked for research purposes; in order to facilitate discussion about the reasons researchers might want to use the ADRN. These were:

- **Offending, Employment and Benefits project:** Links administrative data between the Ministry of Justice, the Department for Work and Pensions (DWP) and Her Majesty's Revenue and Customs (HMRC) in order to better understand what happens to offenders once they leave prison.
- **Index of Multiple Deprivation:** Links data from most government departments, including health records from the NHS, unemployment and tax data from HMRC and benefits information from DWP to provide a measure of "deprivation" for very small areas.
- **The National Pupil Database (NPD):** Links data on pupils from multiple different sources including examination results, attendance records and school level data into a single database containing detailed information on every child in school in England.
- **'Digitising Scotland':** This project aims to computerise and link all birth, marriage and death certificates in Scotland from 1850 to 1974 to enable researchers to produce anonymised family trees across this time. Unlike the other case studies used, this project has not yet been completed.

Despite the low levels of awareness and understanding of social research, participants usually found these examples very interesting. They also enjoyed hearing about the research presented by the experts. This was especially true for research in the area of education, medicine and social care, where they could understand the reasons that research is carried out and see the direct uses of the findings to change policies.

> *"[The education case study is] useful – can look at pupils under achieving in certain areas, so you can see where teachers are not doing their job properly."*
>
> *King's Lynn*

However, where the implications were less obvious or immediate, participants found it hard to understand the value of social research. Much depended on how the examples were explained. For example, in Stirling, where the researcher who works on the 'Digitising Scotland' project was present in the workshop, participants were much more enthusiastic about that particular case study. In other workshops the applications and impact of this research were less immediately obvious to participants.

<u>Understanding of social research</u>

Across all workshops participants' lack of familiarity with the aims and methods of social research meant that they made many incorrect assumptions, the most important of which are outlined below. Some participants had a much stronger understanding of social research, particularly by the end of the dialogue. As such, the issues described below were not raised by *all* participants. Instead, they are intended to show the variety of assumptions that other members of the public might make when considering their views on the ADRN and its work.

1) The **extent of what is not known** about society, and the extent of what is not properly evidenced about society. Perhaps because we live in such a data-driven culture, participants said that they did not understand the "point" of some of the case studies i.e. the reasons that research might be needed on a given topic. They assumed that the findings were already known, either through common sense, or simple observation.

> *"I can understand [the Offending, Employment and Benefits project], but it's obvious."*
>
> *Belfast*

> *"Why do we need [the National Pupil Databases] if we have Ofsted? If you can get it already, why do they need a database?"*
>
> *King's Lynn*

In some cases this was related to an assumption that government already links all of its own data. For example, when discussing the data linking journey in the afternoon, many did not see why researchers might need to link tax and benefits data for their research.

> *"Is there not stuff out there already? Can't they just read up on it?"*
>
> *Cardiff*

2) The **process by which research happens**. Throughout the dialogue workshops, participants assumed that the government commissions all social research directly. Some also thought that the government controls the findings. For example, in one location participants were surprised when the ADRC expert explained that researchers usually come up with their research questions themselves:

> *"Researchers have a choice then? They don't get told what to do?"*
>
> *Belfast*

This meant that the link between academic researchers and administrative data was often unclear, with participants presuming that only government researchers would want to use this data (because they would be the ones who could use it directly for public good). Often, this was because participants saw little difference between research and operational uses of administrative data. They were also surprised by length of time the research process can take, which meant that arguments in favour of the ADRC that were linked to greater efficiency were not particularly powerful.

3) The **aggregation of data**. Social researchers who access linked administrative data almost always work with *aggregated* anonymised quantitative data. However, participants often assumed that researchers need to access data at an individual level. They assumed this because they intuitively

think about data as data about individuals, rather than groups. This assumption caused concerns around de-identification for some, as outlined in the next chapter. Experts at the workshops often needed to explain repeatedly and in detail why they are only interested in looking at groups, and not individuals, when conducting social research.

4) **How understanding is generated** from data. While the general principle of creating statistics from aggregated information was understood by most participants, specific case studies often required more information. The example of family tree research in Scotland was questioned because participants were unclear how historical data could be useful for present understanding:

> *"I'm interested on a personal level but it's not relevant to the future, the past won't change. It's irrelevant"*
>
> *King's Lynn*

Lack of familiarity with the long-term nature of how social understanding is generated meant that some participants were dismissive of exploratory or purely academic research. Participants typically assumed a straight line between initial questions, a single research project and policy-relevant findings and conclusions, rather than the more fragmented reality in which different projects find out different parts of the answer to question about society, and understanding builds over time. They thought that there was no point in interrogating administrative data unless the researcher has clear hypotheses and a strong expectation that the findings will be used for specific public benefits in the short term.

5) The **use and interpretation of findings**. Few were familiar with the difference between academic researchers and policymakers, or the lack of strong links across these roles. In some locations, participants wanted those who use administrative data in their research to be able to *guarantee* that their findings would be used for the benefit of society, or prove definitively that it can have an impact. However, some did think that impact might only happen over the long-term.

> *"Some things about society we won't be able to see until it goes down the line."*
>
> *Manchester*

Similarly, some participants assumed that data or social research findings cannot lead to solutions for social problems, and thus saw little use for this kind of research.

> *"Data will not cure diseases or come up with solutions."*
>
> *Stirling*

<u>Views of the value of social science research</u>

Those who were the least familiar with social research tended to be the most critical of it, often characterising it as researchers simply *"wasting time".* Sometimes these criticisms were less to do with social research itself and more to do with a lack of trust in government and politicians and how they use research findings (often described simply as *"statistics"*) to justify policies.

> *"You can't take figures at face value."*
>
> *Stirling*

For the most part though, participants attached some value to social research, and generally valued it more as the workshops progressed. They did however tend to criticise it for a number of reasons, outlined below:

- **Profiling or pigeonholing individuals or areas.** This was raised with reference to the case study examples of the National Pupil Database and the Index of Multiple Deprivation. While participants could see the value of both sets of linked data for improving services and allocating resources, they also thought that they could lead to unintended negative outcomes for specific types of people or those living in particular areas.

  "Can be detrimental - you may end up predicting grades for children based on statistical average based on typical, data-driven performance of people from similar backgrounds."

  Cardiff

  "I personally wouldn't move to an area that was deprived, statistics are interesting because they influence people not to go there."

  King's Lynn

- **Low impact.** There was strong suspicion that government does not use research findings properly, and, as such, it often represents wasted money and effort. It was not understood that people other than the government (e.g. academics, charities, businesses) might be interested in the outputs of social science research. Participants reasoned that as they perceived the influence of social research as limited and the pace of social change as slow, there must therefore be little value to greater social understanding.

  "Is research looked at, or acted on? If it is acted on, great, but if people are just looking at it, I don't know."

  London

- **Overlooking context.** Some suggested that research using data can never capture or control for all of the potentially complex causes of social issues or problems, and that relying on data can therefore lead to incorrect conclusions and inappropriate solutions. There was also concern that people's voice is lost when quantitative data in particular becomes too important in developing understanding.

  "I'd be uncomfortable if schools are demonised individually, when there are loads of reasons socially that they're performing badly."

  Cardiff

- **Researching the 'wrong type' of people.** This criticism was raised specifically in relation to the example of the crime and reoffending data linking project. Some resented public spending on social research about people who have done wrong in society.

For some, uncertainty about whether social research is needed at all was a key driver of attitudes towards data linking and the ADRN. These participants tended to be sceptical of the aims and cost of the network, and often proposed stricter rules for the ADRCs than other participants.

Even where understanding of the aims, methods and usefulness of social research was greater, participants tended to argue that it should always be undertaken to deliver an outcome that is of social value, and to criticise what they saw as research with no specific or tangible "point". Blue skies or theory-led academic research was thus usually valued far less than social research that can directly feed into government services or other practical social good.

2.2 The value of data

By contrast with their low awareness of social research, participants were generally familiar with the importance of data to modern society. In their spontaneous discussions of what data is and how it is used, they often mentioned the power of data and its worth to individuals and businesses.

> *"Organisations that are clever [use data]. Ones that want to make services better."*
>
> *Manchester*

> *"Someone sitting in an office is getting your data somewhere. It is like a goldmine."*
>
> *Manchester*

They described how data is collected from them all the time, for example when using companies' services, interacting with the government, and making applications for jobs or courses. In general, they were either uninterested in this, or resigned to it, seeing the modern world as one in which people collecting data from or about you on a regular basis as "just part of life".[10] They thought that having less privacy is an inevitable result of this.

> *"You just get used to knowing you're on databases, on lists, you just roll over and take it."*
>
> *Belfast*

However, some participants thought that too much data is demanded for no good reason, and felt that they no longer had control of their personal data. They spoke about being worried when having to provide personal data (to government or private companies), and concerned about other people who do not understand the implications of doing so, particularly online.

> *"People give too much information these days, all these youngsters with Facebook, everyone knows everything."*
>
> *Belfast*

Perceptions of data and data collection

Across locations, participants recognised the many ways in which data is collected from them on a daily basis (although some younger participants with less experience of filling in forms and applications found this more difficult). Most spontaneously spoke about the provision or collection

---

[10] This reflects EUROBAROMETER findings that 74% of the Europeans see disclosing personal information as an increasing part of modern life. See http://ec.europa.eu/public_opinion/archives/ebs/ebs_359_en.pdf (accessed 070314).

of personal data that happens when a person fills out a form or survey, or gives their data to another person or organisation (whether public or private). They usually had to be prompted to start thinking about more passive forms of data collection, for example cookies[11] or data on travel or purchasing patterns. Younger participants tended to have more knowledge of newer ways that data is provided or collected, giving a wide range of examples from people posting data on Facebook, to schools where children's fingerprints are collected to allow cashless payments for lunch.

Despite seeing data collection as an inevitable part of life, there was concern about the lack of explicit consent for the different uses that organisations might make of the personal data they collect. For example, when discussing the data collected by private companies, younger participants in particular described a quid pro quo whereby they understand that by giving information to a company, you gain something in return. This could be discounts or other loyalty rewards, or a better service.

However, participants were also frustrated at not being able to keep track of where their information 'goes' after they have willingly given it to one particular organisation. For example, many said that they often forget or cannot find the 'tick box' that you have to search for in order to opt out of data sharing.

> *"If you buy a phone from O₂, the next thing you know Vodafone are ringing you if you don't tick a box on a form. It's the small print."*
>
> *Belfast*

In any case, there was a large amount of cynicism about the efficacy of opt-out mechanisms when it comes to data sharing, with many participants citing examples of being contacted when they did not know specifically how organisations had sourced their data.

In general participants had not thought very much about passive data collection, and tended to focus on data that they had consciously 'given' to businesses or government in subsequent discussions. Having said that, there were a few isolated concerns about consent mechanisms for passive data collection, for example the participants who objected to car counting for transport surveys.

> *"I think you should be asked – it should be made obvious if you're part of some type of data collection. If you fill in a form, you know. But people who count cars, no."*
>
> *King's Lynn*

Different issues arose when discussing government data collection. While some saw a similar quid pro quo as in the private sector, whereby your data leads to a direct personal benefit, others pointed out that giving your data to government is usually compulsory. This is not something they usually questioned, and few saw it as a problem when discussing data linking. Indeed for some, giving data

---

[11] A cookie is a small amount of data generated by a website and saved by an individual's web browser. Its purpose is to remember information about the user, to allow web pages to be personalised on repeat visits.

to government was seen as an act of citizenship which allows the government to make better decisions and improve the way things work.

> *"You just have to give your information to use these services, if you want to be treated by a doctor, if you want a job."*
>
> *King's Lynn*

By contrast, a few participants who had multiple interactions with government services repeatedly returned to the issue of lack of consent for further uses. They thought the lack of explicit consent had a bearing on whether data should be shared and linked or not – for example a participant who had been in care strongly disagreed with his data being linked for research purposes without his explicit permission.

> *"If you come out of hospital or prison and you've got nothing, you've got to share your details."*
>
> *London*

Data security

Personal data security was very important to participants, and this framed much of the discussion about linking administrative data. They were particularly concerned about identity theft, and personal data being sold on to other organisations. Often the main objection to the latter was the profit companies make from using their data, rather than the privacy implications.

> *"I don't have any objections to anyone knowing anything. My problem is when they sell the info. Otherwise, I've got nothing to hide."*
>
> *Manchester*

This was often driven by experience (either personal or through friends and family) of their data being shared or sold by private companies without their express permission. In all locations participants described receiving unwanted and annoying insurance and other marketing calls that they were convinced were the result of illegal data sharing or sales. A few participants referred to the Data Protection Act, but there was widespread cynicism about respect for this legislation and how well it is enforced.

Participants were also worried about personal data being leaked, lost, shared or sold by government departments to third parties, particularly commercial companies. Several participants had experiences that made them think that hospitals pass on data about those who have been in accidents to insurance companies. Low trust in government more generally seemed to be driving these views.

> *"The government might sell it to other people for all I know."*
>
> *Wrexham*

General concerns about data security were related to the widespread perception that it is not possible to create completely foolproof data security processes. They pointed to:

- Media stories about government data loss as evidence of how human error can lead to breaches in data security. Participants cited examples of datasets with records for every person in the country being left on trains.

"What about loss of records? I've very concerned about that"

Cardiff

- Potential for data theft by those who have access to individual's records and databases.
- Recent stories about the NSA accessing individual's data, and 'spying' on government officials, and individuals such as Edward Snowden and Julian Assange gaining access to government data. Participants argued that these sort of stories proved that *"anything can be hacked"*, or that the security of any database can be compromised by those who have legitimate access. In their view this means no databases are truly secure.

This meant that they did not think that either the government or private companies have the ability to keep data completely secure, whatever their intentions. Overall, participants trusted government's intentions more than commercial companies on data security and protection from data misuse (such as selling information on to third parties) or fraud. However, a few mentioned the sale of the electoral roll as something that makes them worried about how government treats data. In addition, some government institutions are trusted more than others to deal with data securely. For example, a few participants said they would not trust the police to look after their data properly, because they think that the police are prone to abuse their authority.

Use of administrative data

Throughout the dialogue, participants did not always distinguish between using administrative data for operational purposes (i.e. to improve the way services are delivered), and its use for research purposes. Indeed, many participants were not familiar with any non-operational uses of administrative data before they were shown examples.[12] This was primarily due to low awareness of statistics and social research, and the subsequent low value that they placed on these uses of administrative data. The exception was when discussing examples of specific datasets where participants were more familiar with the research outputs, for example using pupil examination results to feed into school league tables.

However, in all locations, a subset of participants had at least some knowledge of how aggregated data is used in research and planning for government and public services.

"The Government [collect medical data] so they can look at trends of illness in certain areas, diet in certain places, health risks and future burden on the NHS"

King's Lynn

Positive views about the use of administrative data included:

- ✓ Use of administrative data to improve the way services are run or to increase national security was uncontroversial for most.

---

[12] An exception was King's Lynn, where the younger group focussed on research uses of data even in the spontaneous conversations.

- ✓ Those who were more trusting of government generally tended to think that any government use of data would be fairly benign and benefit the general public, or a subsection of the population.
- ✓ There was strong support for the use of administrative data in planning for future service provision.
- ✓ Participants wanted data to be used to reduce fraud in government services.
- ✓ Many were open to the idea of data being used more efficiently to avoid repeat collection of the same information from individuals.

On the other hand, spontaneous concerns around the government's use of administrative data included:

- ❖ That there was too much unnecessary personal data collected and duplicated by departments and services. It is not always clear to them why some information is collected when it does not seem directly relevant to the service received.
- ❖ The potential for profiling certain subsections of the population, or individuals, and potentially treating them differently on that basis.

> *"Sometimes I think [the government] collect all your details because they want to be pigeonholed. Like in the computer world they can put your name in and see what your likes and dislikes are."*
>
> *King's Lynn*

- ❖ Concerns that administrative data could be inaccurate, especially where it is self-reported, and that this could have negative consequences for individuals or groups.
- ❖ Using linked administrative data to justify and implement controversial policies, such as the bedroom tax:

> *"They found out a member of my family had passed away because of the electoral roll, so now I pay bedroom tax."*
>
> *King's Lynn*

- ❖ Government departments holding prejudicial administrative data (e.g. records of arrests) for longer than they should, and sharing that with other people.
- ❖ The ideological and political factors that influence what data is collected by government, which in turn influences the way data is interpreted and the decisions government makes. This point arose in the context of discussions around data about schools and pupils, where the consequences of data-driven targets were seen to have been negative by some participants.

Spontaneous views of administrative data linking

Participants commonly assumed that government administrative data is already linked and shared across government departments. A few participants in each of the workshops thought there was already a central government 'mega-database' where all the information on each individual is stored. Some thought that National Insurance (NI) number is used as a way to link data across departments, and by implication that giving anyone your NI number gives them access to a huge amount of data about you. The more common view was that government departments share data as

and when necessary for operational purposes. However, participants did not spontaneously mention academic or any other non-governmental researchers using linked administrative data – none had heard of this practice until taking part in the dialogue.

This vague expectation about the prevalence of government data linking was in contrast with the equally strong view that the government does not currently use administrative data as well as it could. Participants shared many stories of government departments and public services not acting together in a joined-up manner because they did not have access to the same information. The perception therefore seems to be that linking administrative data is already happening but that it is not being used effectively to improve services.

> *"My son has 13 different disabilities but information is not shared between departments. He gets asked why hasn't he applied to a job, but he couldn't do it…the important things are not shared wisely."*
>
> *Manchester*

In general, participants spontaneously felt that linking administrative data is sensible and can be useful for practical purposes, for example to stop benefit fraud and in general enhance operational efficiency. They wanted government to share data as much as is necessary, as long as they aren't *"doing anything wrong"* with it.

> *"I don't see why that's a problem if I go into a medical branch, and they know all of my data already, that's a good thing. I don't know why you wouldn't want that."*
>
> *Belfast*

A less common view was that data linking can be justified for planning government services, but not for operational purposes, as this would result in government employees being able to access too much information about individuals.

A handful of people were completely opposed to any information sharing without explicit consent, although this was usually based on the expectation that the data to be shared would include personal details.

> *"I don't want my data going anywhere other than the person I give it to."*
>
> *Cardiff*

Data ownership

Views of ownership of data drove many of the more detailed views on the ADRN outlined below. Most importantly, participants were split over what constitutes personal data.

For many participants, any data that does not include name/address/NI number or other identifiers is no longer 'personal'. These participants were more willing to support administrative data linking as they felt data is de-personalised once it is de-identified. Within this overall view though, some particularly sensitive types of information were seen as too personal to be shared outside of the agency that collected it, for example records of domestic violence, or HIV status, because of the potential consequences of the data getting into the wrong hands.

"For domestic violence you don't want anyone to know your info."

Manchester

In addition, 'personal identifiers', the term, meant more than just name/address/NI number to some people. They did not think that these are the only types of data that make someone identifiable i.e. they think it could be possible to trace someone with enough other information. For others, any data relating to 'me', whether identifiable or not, was personal. These participants argued that, even after they have given data to government they do not see themselves as relinquishing ownership. They therefore found it more difficult to accept data linkage, as they felt that any administrative data relating to them that exists (even without name/address/NI number) is still highly 'personal', and, additionally, requires their consent to be used for another purpose other than that for which is was initially collected.

"I'm uneasy about data ownership. It's [tax records] my data, not their [HMRC] data."

Cardiff

Chapter 3: Views of ADRN initiative

This chapter describes participants' views about current plans for the Administrative Data Research Network. It outlines spontaneous views towards the ADRN, before going on to examine the main areas of confusion, issues of continued concern, and points that reassured participants across the two dialogue workshops.[13]

Key points

The table below summarises the main concerns, reassurance points and outstanding questions across the dialogue workshops. It is important to note that some of these issues overlapped, not all of these views were held by everyone. It shows the breadth of views what were encountered across the dialogue workshops.

Concerns	Reassurance points	Remaining questions
❖ De-identification – will it be possible to identify people even without their 'personal details'	✓ ADRCs' independence	? Can this improve operational use of administrative data?
❖ De-identification process – is this secure? Where does the personal data 'go'?	✓ Data could improve public services or allocate funding better	? How much have government departments 'bought in' to this idea?
❖ Security – is it possible to hack? Does that even matter?	✓ Efficiency – Save money, or reduce burden on population	? Who works for the ADS and how are they governed?
❖ Cost – is this £34 million well spent?	✓ Better regulation than current system	? What will be the punishment for any misuse of data by researchers?
	✓ No 'super-database'	
	✓ Limitations on researchers	

3.1 Spontaneous views of the ADRN

Early in the first dialogue workshop, participants were given a short presentation that outlined the purpose of the network and why it was being set up. Throughout the rest of the two days, they were also given some case studies to understand the kind of projects that the ADRN might facilitate, a chance to talk to social researchers who link data and those who will help run the network and associated centres, and a demonstration of the process of linking de-identified data.

---

[13] Please note that some concerns were based on misunderstandings about the process, but often drove views nonetheless. Where misunderstandings were cleared up, concerns tended to decrease or disappear altogether.

As a result of exposure to new information, and the arguments of their fellow participants, views tended to shift throughout the two days, often moving from support for the idea of the ADRN, to concern or opposition to aspects of the plan and back again.

Initial reactions

When the ADRN was first presented in the dialogue workshops, reaction tended to be cautious. A handful of mainly younger participants were quick to support the idea in principle, and questioned why it has taken until now to put this type of system in place. But other views were mixed. Many were initially surprised about the need for the network, based on their assumptions that there are already high levels of data sharing in government. However, they were often reassured by the fact that there are so many barriers to sharing currently in place. Few had realized how difficult it is to gain access to and link administrative data.

> *"I thought that this system was already in place, and it was already happening, I'm quite surprised by this."*
>
> *King's Lynn*

Others were unhappy to find out that this was happening without their notice. In Cardiff in particular, participants were frustrated that there had been no public consultation.

> *"I'm surprised I've never heard of it before…I'm assuming that there were a lot of public opinion surveys beforehand?"*
>
> *Cardiff*

For some, initial rejection of the idea of the network seemed to be based in more general discomfort about change.

The most common reaction from those with initial concerns was a call for more information about the reasons for setting up the ADRN. This tended to be motivated by a general lack of trust in government motives. Participants were often unsure of the value or benefit of creating the network, given their low awareness of the uses of social research, and the low value they place on social research. In Manchester, participants' first reaction was to question the cost, as they were fearful of taxes being increased to pay for it.

Other questions that were raised spontaneously included:

- Is the government doing the linking or the researchers?
- Are the researchers the only people looking at the data?
- Why does it have to be regional? Why can't there just be one centre?
- What is the process to gain access to the data?
- One the linked data has been used, what will happen to it?

Operational vs. research uses

While isolated, there were several points of confusion that emerged at different stages of the dialogue across groups and locations. These point to the potential misunderstandings that could arise when communicating about the ADRN. Most importantly, some participants found it difficult to

understand or bear in mind the difference between operational and research uses of data. This meant that they often held concerns about government employees gaining access to personalized linked datasets.

"Will local government have access to this? Or just central?"

Belfast

Conversely, this confusion made some particularly supportive of the ADRN plans, because they thought that linked data could be used to increase government efficiency and reduce fraud. For example, many of the post-it comments on the "Offending, Employment and Benefits" project case study said that they supported data linking in this case as it might help prevent fraud, although the purpose of the project was not operational.

3.2 Concerns

Many concerns about administrative data linking were raised throughout the two days. These were discussed in more or less detail depending on the nature of the discussion in each group, and the extent to which participants understood and remembered particular information.

One overarching concern that tended to be addressed as the dialogue progressed was the perception that the ADRN is unnecessary. Participants learned, through dialogue with researchers and case study examples, that it is already possible to link administrative data when really necessary. The added efficiency and extra research that may be possible by creating the network were not always seen as particularly valuable.

The strongest concerns, however, related to the *process* of data linking. These are presented below roughly in order of strength of feeling, starting with those that worried participants the most.

De-identification

The de-identification process was described using a visual role-play, with cards representing different pieces of data. The main steps described to participants are outlines in the three slides overleaf, which show the 'data journey' from the individual to the linked, anonymised dataset.

Data linkage journey - Step 1 – Data giving

2011

Adam – data giver

- Name and other personal details
- Work history
- History of claiming JSA

Department for Work and Pensions - data owner

2012

Adam – data giver   Adam's employer

- Name and other personal details
- Earnings
- Place of work, start date of employment

Her Majesty's Revenue and Customs - data owner
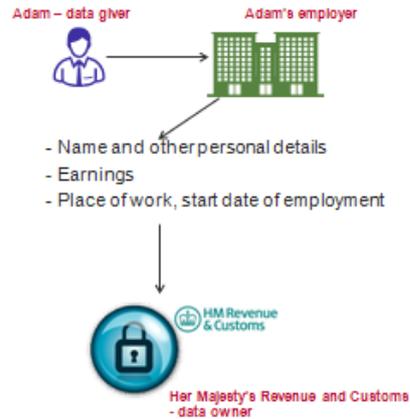
Data linkage journey - Step 2 – Project and researcher approval (ADS)

Ruth - researcher   Ruth's project: long-term cost of unemployment

Application / Approval

Administrative Data Service

Negotiate Permission

DWP Department for Work and Pensions   HM Revenue & Customs
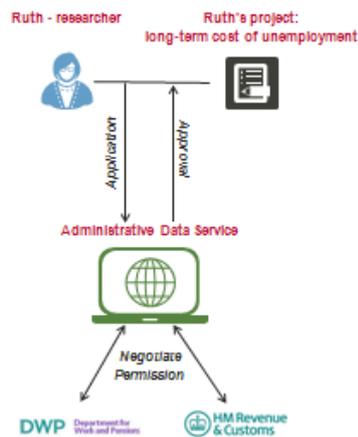
☐ Researcher Approval
- Check skills
- Prescribe training
- Give accreditation

☐ Project Approval:
- Check feasibility, timescales, resource requirements
- Check methodology
- Science panel review
- Ethical Review

☐ Gain permission from Data Owners (the relevant Government Depts)

Data linkage journey - Step 3 – Data Linking (ADRC)

DWP Department for Work and Pensions   HM Revenue & Customs

Transfer Anonymised Datasets

Administrative Data Research Centre

Link Data   Train, Monitor Support

Analyse linked data

Linked Data   Ruth - researcher
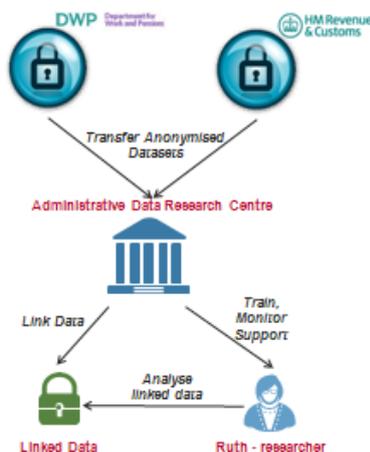
☐ Obtain anonymised data from Data Owners

☐ Link data in a secure setting

☐ Support researcher
- Give training
- Provide secure setting where researcher can access the linked data, and support
- Monitor use of data

☐ Destroy anonymised datasets

Understanding de-identification and trusting that the process as described will adequately protect personal data was the key to support for data linking. Participants wanted to be satisfied that personal data is never put at risk. Working through the de-identification process usually convinced them that this would be the case.

However in some locations, de-identification remained a concern that participants returned to throughout the dialogue. In most cases, this was simply because they forgot that all linked data would be de-identified, and needed to be reminded of this by facilitators and experts. The repeated concerns were driven by the fact that the term 'data' is associated so strongly with personal and thus identifying information.

Concern was sometimes driven by the misunderstanding that the ADRCs would be creating one 'super' database from all linked datasets. Even when the data linking process had been explained in detail, some participants returned to this view, and did not take on board the idea that only the specific data for each research project would be linked.

However, some remained concerned about the de-identification process as described. They thought that it would be possible to identify individuals in some linked datasets, for example if it included information about someone that was unusual and might only apply to a small number of people.

> *"Place of work narrows it down if you have a small number of employees."*
>
> *King's Lynn*

For these participants, it was important that rules be put in place to ensure that not only are people's obvious personal identifiers such as name, address and date of birth removed before linking, but that thought is given to ensuring that nothing else in the remaining data-set allows for re-identification.

Finally, some concerns around de-identification were driven by very low trust in the government and the wider perception that our freedoms and privacy are being reduced. A handful of participants questioned how they could be expected to trust that the process would work as described. There was also concern about how it could change in the future, with these participants making slippery slope arguments about the potential for the scope of data linking to be extended. Some participants returned to the idea of building a 'super database' or were worried about the data being used by for political purposes such as targeting voters.

> *"Yes, where are the guarantees. 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?"*
>
> *Cardiff*

Security

This was a high concern in every location, and participants asked a large number of questions of experts around security policies and processes. Questions raised around security included:

- Who would have access to linked data? Participants wanted to know that the vetting of individuals for access is robust (for example suggesting CRB checks). A handful were still worried that even the strictest checks would not stop "bad apples" who were determined to get access to data for illegitimate reasons.
- How secure is the physical infrastructure? Participants worried that researchers might be able to remove the data once they have access to it. This worry tended to be raised by those who had stopped thinking about the fact that the data would be de-identified.

"Ruth could write [the data] down, or e-mail it to herself."

King's Lynn

- How does the data "get" from one physical location to another, and is it at risk while in transit?

Some participants were simply convinced that "nothing is secure" and were not satisfied by the proposed security measures at ADRCs. However, most of participants' security concerns were addressed by explanations around current security processes for linking data that those of the ADRN may be modelled on. The most reassuring fact for participants was that researchers can currently only access data in a secure physical setting. De-wired computers, bans on phone, pens and paper in these secure rooms were also seen as robust processes. In one area, there were concerns around the university setting, and preference for data not *"leaving"* government departments. In general it was thought that if sufficient infrastructure was in place, universities were secure enough settings for accessing linked data.

Other questions were raised about cyber security, with participants worried about the potential for hacking, and potential identity fraud. On the other hand, some argued that this is possible with any dataset, and the security measures outlined in the data linking process seemed secure enough in their view.

Those who were less worried about security pointed out that there was little danger in hackers gaining access to de-identified linked datasets, as they would not be able to do much with this type of data, which *"can't hurt anyone"*. They pinpointed the initial data before de-identification as much more risky.

"It's only numbers [once it's linked], why would it matter [if anyone gained access]? At the start it's identifiable, that's much worse and that's already there."

Belfast

Indeed a handful thought that proposed ADRC security as outlined in the data linking journey was too high level and potentially expensive, given how low-value the data would be to hackers.

De-identification process

In the latter half of the first dialogue workshop, facilitators and participants discussed the process that data would go through in order to be linked and used within an ADRC. This process covered all of the steps, from an individual handing over the data to a government agency to the researcher

gaining access to it. A key goal of this exercise was to explain and discuss the concept of de-identified data.

The process was slightly simplified in order to ensure that it was accessible to all participants, whatever their level of previous level of knowledge of data and the research process. One aspect of this simplification was the removal of the third party from the de-identification process.

The simplified de-identification process as presented was generally well understood, and for many this was the point at which they understood how ADRCs will work and why they are needed. However, some thought the process wasn't detailed enough. They had questions and concerns about what happens to the data up until the personal identifiers are removed. Participants in several locations wanted to know who actually removed the data, and how they have been vetted. This was seen as a much higher security concern than who has access to the de-identified data once linked.

> *"I'm worried about who handles the data to de-identify it."*
>
> *Manchester*

In these cases, a more detailed explanation of the third party anonymisation process by the experts present satisfied those with questions that the process is robust. For most participants, who had low interest in the process, the introduction on the concept of the 28 digit linking number on Day 2 increased confusion, as they - incorrectly - thought that this meant that the ADRN would use one 28 digit number for each person in the country for all linking of administrative data.[14] Participants thought would increase the risk of their data being identifiable. Some even thought that it might end up being used as an operational identifier, like and NI or NHS number.

> *"I'm not really clear if you're taking my name and giving me a 28 digit number on the basis of my name then you got to jobseekers or pension and I appear again do I get the same 28 digit number, or a different number?"*
>
> *Stirling*

Cost and cost-benefit

Cost was a concern only in some workshops, and one likely to be driven by wider concerns about the need for value for money in public expenditure at a time of austerity. In Manchester, cost was the first question that participants asked after the presentation that described ADRN plans. By contrast in London cost was not discussed at all.

There was some confusion about whether the main purpose of the ADRN is to save money long-term. For a few, this would be the only justification for putting it in place. Those who debated cost and cost benefit in detail tended to be those participants – again, usually older – who were unconvinced of the value of social research, or thought that we already know enough about society.

---

[14] As part of their 'Beyond 2011' research, ONS researchers currently use 28 digit numbers as part of their masked matching technique to match GP and HMRC records. In the second workshop, participants were shown a video of an ONS researcher describing the process.

"What is the benefit? Everything is already logged now anyway? Why are they pouring extra money into it?"

King's Lynn

These participants thought that £34 million was a huge amount of money to be spent on research. They thought that this money could be spent better elsewhere.

"It's the money that bothers me…I think that potholes are a better use of money."

Manchester

For those who were unhappy about the cost of the network, the explanation that this money comes from a ring-fenced research budget was not enough, with participants arguing that all government spending needs to be justified in terms of measurable social good. One suggestion for ensuring that the ADRN and ADRCs lead to socially valuable outcomes was for each linking project to have to prove that government money would be saved before it could go ahead. This was in contrast with those who assumed that there would be an overall social return, even if it is not immediately clear in the short term.

"This will benefit everyone. That's why we pay our taxes."

King's Lynn

Finally, a small number of participants objected to the cost of the ADRN because they thought that this represented funding for research itself. It was not immediately apparent to some that the £34 million is an infrastructure and administration cost, not least because that seemed quite high to participants.

Other concerns

- **Consent and notice:** The principle of de-identification meant that neither consent nor notice were of concern for most participants, because they did not see de-identified information as personal information, and were thus relaxed about how it is used. However, a number of participants – particularly those who saw all data relating to them, de-identified or otherwise, as 'personal' – questioned why people would not be asked to give consent to link their data. The idea that this would be extremely difficult and expensive to do for data that has already been collected was widely understood. A few participants did not see this as a strong enough argument for linking data without consent. Some o argued that consent could be sought or, at the very least, notice should be given from now on for administrative data, especially given that much of it is collected online, where people are used to privacy policies.

  "I'd like to see an online style privacy policy at the point of you giving away the data."

  Cardiff

- **Lack of consultation:** In Cardiff especially, there was some cynicism about the dialogue, and anger that the public had not been consulted about the principle of putting the ADRN in place. Others were more resigned than angry, but shared the view that anything they said would not have any influence on the policies or projects of the ADRN.

"The government are going to do what they want. No one will be 100% happy."

<div align="right">King's Lynn</div>

- **Potential impact on local areas:** This concern was twofold, and tended to be prompted by the case study that explained IMD. The first aspect was that linking data to produce small area statistics would lead to stereotyping and consequent negative effects on those who live in areas, for example not being able to get insurance or a mortgage. The second was that these areas would be too large, and that knowing statistics at one area level (for example a neighbourhood) would miss issues at *a "really local level"* (for example a street). Participants worried that this could lead to government funding not being allocated to very small deprived areas within better off areas.

- **Potential impact on the individual:** This was concern for those who either did not understand or trust the de-identification process. They worried that historic data about an individual could be linked in with their current records and available for inspection by, for example, future employers.

3.3 Reassurance points

There were some key messages that participants thought were important in understanding social research, the work of the ADRN, and why it has been put in place. Those who had become more positive about the plans over the course of the two days identified many of these points, although others were identified through the analysis process. As with concerns, these are presented roughly in order of strength of feeling, starting with those that were most important.

While all of these points came up in at least one workshop, it is important to note that the influence of these reassurance points was different across individuals and workshops.

<u>Research advocates</u>

The biggest reassurance observed across the dialogue was interaction with the social researchers and representatives of the ADS and ADRCs. These interactions significantly affected participants' views, and usually made them more positive about the ADRN plans. Experts were able to provide very detailed answers to participants' questions about the potential work of the ADRCs, and provide reassurance on points of concern.

Often the effect on participants' views was not just down to the information that the experts supplied and the arguments that they made about the network, but the enthusiasm they displayed for social research and its uses.

"Speaking to [ADS representative] has really helped, someone within the service who is passionate about it."

<div align="right">King's Lynn</div>

Advocacy of the ADRN was important here, as was the provision of basic information about their profession and working lives. Speaking to experts and asking basic questions about the nature and uses of social research was much more effective in helping participants to understand than any of the other stimulus material, including case studies.

De-identification

Participants consciously pointed to this as being the key message that reassured them about the ADRN plan.

> *"Now I know more, and it's de-identified, a lot of the bad factors have been cleared up like trust issues."*
>
> *Cardiff*

This completely changed the views of many participants, who saw de-identified data as no longer being personal to them. As discussed in the next chapter on rules, it is important to emphasise that for participants de-identified data meant the removal of name, address (including full post-code), date of birth and NI number, and any other data that would make it possible for a person to be re-identified from the dataset. As pointed about above though, de-identification was not sufficient reassurance for some participants.

Data being published and used

Participants tended to react well to examples of where the findings of research using linked data is being used to allocate funding effectively, and improve public services. Discussion on this issue tended to stray into the areas of decision making, policy and politics. Again, participants thought that there was a very short and straight line between research findings and direct policy implications. They tended to want any findings from linked data acted on, in order to justify the money spent on the network. Indeed some questioned spending money on the ADRN on the ground that it is impossible to force politicians to use results.

> *"Is there government accountability, so if they have data that has told them this, do they have a responsibility to act?"*
>
> *Belfast*

However, that is not to say that participants could not see the benefits of more theory-focussed research as the dialogue progressed. In conversations with working researchers, they were extremely interested by less policy-focussed questions. Examples of theory led research in an area that they found interesting or close to their personal experience usually made the participants much more sympathetic with researchers. Talking to researchers about the difficulty of ensuring that research is listened to also made them more open to the idea that some research has less policy-focussed goals. In Wrexham, having learned more about how researchers work, participants thought that the ADRNs could play a role in helping us to understand society and encouraging students' intellectual development.

In most areas however, participants remained wedded to the idea that administrative data linking projects carried out through the ADRN should lead to an obvious and measurable impact on the general public. After discussion, some came to the compromise conclusion that, in order to access

linked data, researchers should have to have plans for publishing their work for public access, and ensuring that those who work in the relevant policy areas are informed of the results.

Transparency was also thought to be an important part of the quid pro-quo. While participants understood that their consent cannot be asked, they were clear that, in return for data about them being used, they should have the option to access the findings.

Conditions for accessing linked data

As outlined in Chapter 2, lack of familiarity with and therefore lack of intrinsic trust in researchers and their motives meant that some participants had spontaneous questions about who should have access to linked data and the approval process. They strongly believed that the motivations of researchers should be interrogated, and that the process for gaining access should be rigorous. In particular, the need for review by scientific and ethics panels was welcomed.

ADS and ADRC independence

By contrast, some participants had inherent trust in the independence and motivations of academics. Others at least thought that they were more trustworthy than those working directly for the government. Therefore the fact that ADRCs are situated in universities appealed, as it meant that the ARDN was seen as an initiative aimed at gaining societal understanding, rather than the government "prying".

Efficiency and contrast with current arrangements

Arguments that resonated well were those that presented the ADRNs as leading to a more efficient use of data and therefore cost savings for government or reduced burden on the population in terms of data collection. This was thought to be especially important at a time of austerity when the government needs to know more about society in order to better use scarcer resources.

> *"Now that we're in recession, government can only help if they know where they need help. Push money or better teachers into schools…informed decisions"*
>
> *Belfast*

As outlined above, the assumption that most government data is already linked meant that participants initially did not understand why the ADRN was necessary. Hearing from working researchers about the barriers and time-delays that they had faced in previous linking projects meant that they were much more enthusiastic about the idea of streamlining the linking process. Explanations of the high cost and effort involved in data collection through surveys also convinced participants that data linking is an efficient use of resources.

> *"It's a bit messy and complicated at the moment. It should be out there and researchers should have access to the data."*
>
> *Manchester*

In some areas, the most engaged participants had questions around the quality of administrative data over and above survey data; and were interested in the fact that some issues of statistical robustness might be reduced through using administrative data.

Those who were concerned about researcher motivations were attracted by the idea that the ADRN provides a more consistent means of regulating researchers and data linking research than the current ad-hoc arrangements on projects.

Other points of reassurance:

- **No 'super-database'**: Some were reassured by the fact that datasets for each linking project will remain separate, and be destroyed at the end of the project. However it should be noted that even at the end of the second day of workshops, several participants still seemed to assume that there would be such a 'super-database', or that it would be created eventually.
- **Aggregation:** Where participants understood that researchers were not interested in individuals per se but aggregated groups, this tended to make them feel much more supportive of greater linking of administrative data. However, there remained participants who did not understand why individual data needs to be linked if researchers only want to look at aggregated linked data for groups.
- **Limitations on how data can be linked:** In one workshop, explanation of how researchers are monitored while using linked datasets (e.g. not being able to cross-tabulate certain variables that might be sensitive/potentially identifying) significantly reassured participants.
- **Regulation and enforcement:** Some participants were very interested in the potential sanctions for researchers and institutions that misuse data. The most important regulatory lever was thought to be the potential for the institutions – as well as individual researchers - to lose funding and possibly institution-wide permission to access to linked data:

   "Now I know more about the researchers' permissions being removed they've got more motive to keep safe, it's personal to them."

   King's Lynn

   Those with very low trust in researchers asked questions about oversight, and those in one location who were told (by ONS experts) about the existence of the UK Statistics Authority were reassured by the fact that there is already a regulatory body in place.

3.4 Remaining questions

Finally, there were certain questions that came up across workshops that point to the kind of issues that the general public might want to be addressed in communications about ADRN.

- Can this improve operational use of administrative data? Some participants found the distinction between operational and research uses of data difficult to separate conceptually, and thought that any linking through ADRNs should be of more direct benefit to individuals. While this shows a misunderstanding of the de-identification process, it highlights the

potential for confusion, and the desire for greater linking where it will have a positive impact on individuals.

- How much have government departments 'bought in' to this idea? Participants were concerned that without data owners being prepared to hand over the data, the whole initiative will be a failure (and by implication waste money).

- Who works for the ADRN and ADS and how are they governed? Some wanted to know more about how employees are chosen and who is on the governing body. In some workshops there were detailed discussions about the need for strong regulation of and transparency from the ADS. As outlined above, hearing about the Statistics Authority regulatory role in relation to the ONS helped assure participants in one location that the rules and enforcement in this area are sufficiently strong.

- What will the public hear about this and how will they be involved? This is discussed in detail in Chapter 5.

- Will the ARDN create jobs? This was based on the assumption that £34 million of government spending should create additional jobs. In areas where this was discussed, some participants were disappointed to hear that the numbers of jobs created might be low. Some were concerned about the potential for job losses among those working in data collection as a result of increased data linking.

3.5 Considered view of the ADRN

Overall, there seemed to be a correlation between levels of understanding of social research, and levels of positivity about the ADRN plans. Those who had previous knowledge of how research works tended to trust researchers' motivations and therefore had fewer concerns about the ADRN plans. Others who started the day with low trust that was linked to low awareness found the process and journey of learning interesting and reassuring.

> *"The more I talk the more in favour I am. At first I thought this was an invasion of privacy."*
>
> *Manchester*

Others were initially apathetic about researchers, either trusting not untrusting. These participants either remained apathetic or became more trusting of researchers' motives as the dialogue progressed.

By the end of the second day of dialogue, participants tended to have concluded that they would support the increased levels of linking of data via the ADRN, if three main conditions were met:

- The data is linked for socially beneficial purposes

> *"As long as it's used for good, like to develop things, improve services, improve knowledge."*
>
> *Belfast*

- It is fully de-identified

- It is kept secure at all times

The latter two concerns in particular were discussed during the data-linking simulation exercise, while the first came out during broader discussion of the nature and uses of social research.

While the other concerns and reassurance points presented in this chapter were important, these three issues were the key to broad understanding and support of the network. It should be borne in mind however, that participants needed extensive information and discussions with experts and researchers in order to be satisfied that these conditions would be met under the ADRN plans. This suggests that simply publicising these three conditions may not be enough to ensure that the general public are reassured about or support the work of the ADRN.

Chapter 4: Proposed principles for ADRCs

This chapter describes participants' considered views on the principles and rules that should govern access to linked data. These rules represent participants' views after a day and a half of digesting information, dialogue with experts and debates among participants.

Key points

The table below sums up the overall findings about the rules for the ADRCs, based on analysis of participant opinion across the seven dialogue locations. In summary, participants felt that once there are strong researcher approval and security process in place, they were happy for data linking to go ahead using all types of administrative data.

Question	Proposed principles and rules presented as part of the dialogue	Principles and rules suggested by participants
Why	To find out more about society; or To help national or local government to plan or carry out their services	As per proposed rules, provided there is no commercial gain for those carrying out the research, and the research had been through methodological and ethical approval processes.
Who	Researchers (i.e. those affiliated with a university or a government department) who have done the training offered by the ADRCs and gained an accreditation.	Once a researcher had undergone proper academic training and had been accredited, they should have access to linked data, provided they are not accessing the data for solely personal gain, or profit.
What	Any data that is available in datasets collected by the government, where the government department that owns the data gives permission. This could include any information that the government collects on any of the forms they ask the public to fill in.	The overall view was that any *government* data administrative data could be linked provided identifying individuals would not be possible. However, there was also a lack of consensus about particular types of information including religion, full postcode, travel patterns, and medical records.
Where	In safe settings, which in practice would mean in a secure area within the university that hosts the ADRC, on secure servers with access limited to those who have been vetted and checked. In the future the ADRCs may make it possible to access these secure servers remotely.	Few concerns about allowing researchers to have access to data in a controlled and secure setting within the ADRC, once the researcher had been vetted. Significant concerns about allowing remote access to a secure environment because of general worries about internet security.
When	For the duration of their research project	Participants tended towards one of two views: 1) Linked datasets should be stored so that researchers - could have access after the initial research project is complete. 2) Destruction of data after project completion reassuring

4.1 Principles for ADRCs

Throughout the first day, participants spontaneously suggested principles and rules that they thought should be in place in the ADRCs, including:

- Uses of data – There should be rules in place to ensure that the research does no harm, and ideally that it benefits individuals or society.
- Reasons for linking data – Building on the previous point, many participants spontaneously called for specific rules around project impact. They wanted it to be clear to those applying for data to be linked that "something has to come out of it". This related to the underlying principle that research should be socially beneficial and only undertaken if it is going to be used.

> *"They have to substantiate it, give reasons they are doing research, whoever requests the data set, has to explain positives."*
>
> *King's Lynn*

- Strong oversight and approval processes - Participants wanted rules in place to ensure that researchers and their projects are bone fide. Spontaneous suggestions included references from established individuals or institutional affiliation.
- Ban on commercial researchers using the data – Participants did not want anyone to make money using from linked administrative data.

> *"I don't want a commercial gain off my data. I don't mind government making reasonable policy off it."*
>
> *Cardiff*

At the second dialogue workshop, participants were asked to deliberate as a group and come to a consensus on what rules they would put in place if they were running and ADRC. In order to ensure that participants were considering a broad range of possibilities, they were given prompts to consider for each area (who, what, when, where and why). For example, when discussing who should have access to linked research, they were asked to consider private individuals, PhD students, academic researchers, government researchers, researchers working for private companies funded by the government, and researchers working for private companies not funded by the government. This allowed them to talk about where the 'red lines' lay in relation to the ADRCs, and pinpoint the situations in which they were not comfortable with administrative data being linked.

4.2 Why? Appropriate reasons for linking administrative data

After consideration, participants generally agreed that, provided each project obtains methodological and ethical approval, the proposed broad criteria are strong enough to allow administrative data to be linked:

- To find out more about society; or
- To help national or local government to plan or carry out their services.

Participants wanted administrative data (i.e. data about individuals collected by government) to be used for projects that will have positive outcomes for society. They were particularly keen on it being used for projects that help the government better target spending so that it is more effective, as well as projects that improve key public services such as health and education in any way.

At the first dialogue workshop, many participants were quite negative about theory led or academic research, primarily because it seemed too abstract, and it was hard to imagine how such research could lead to socially valuable outcomes. However, after lengthy discussions of the process that researchers have to go through to get access to linked data, most softened their views on this, and were happy to allow for theory led research using linked administrative data once the project fell within the broad set of criteria above.

> *"As long as there's reasonable intent."*
>
> *Wrexham*

While there was some residual concern that "finding out about society" was too vague a criteria, participants generally trusted that the process of gaining approval would ensure that linking projects that do not have any social value would not go ahead. Their main concern was to ensure that projects that are "sheer nosiness" are not approved. This seemed to be driven by the suspicion that allowing projects to go ahead "just because" increases the risk that the data will be misused. As outlined in the previous chapter, some participants wanted more information about the membership of the boards that give approval to projects and assurances that they are impartial.

While agreeing with the fairly broad criteria for allowing administrative data to be linked, some concluded that if there is a limit on the number of projects that can go ahead – for example, because of limited resources within the ADRC, or very high volume of applications for access to linked data - priority should be given to those projects that will lead to the greatest social value.

Within this broad focus on social value, participants argued that research that is only of benefit to the researcher should not be allowed. There was also a strong view that there should not be commercial gain for those carrying out the research, even if it also meets one of the two criteria above. Participants did not like the idea of someone making money from information about them, even if it is de-identified, and advocated banning private companies from using linked administrative data for their own purposes.

> *"It should not be allowed for private companies' private gain."*
>
> *Manchester*

Some, mostly younger participants, were more open to the idea of allowing commercial gain from linked administrative data as long as a social purpose is also served by the research. They thought that the potential commercial benefits could lead to wider societal benefits, such as new jobs. However, even these participants argued strongly that access should not be allowed *solely* to create commercial benefit.

> *"If it is for a positive ending, jobs created for the economy, but if it doesn't help the area then ethically why should you share that information?"*
>
> *King's Lynn*

Finally, a handful of participants suggested that linked administrative data should be used for specific operational purposes e.g. "to weed out false claimants". They saw linked data as a resource that could be used by government departments and local public services to improve the way they serve the public. By contrast, some thought that there should be rules to ensure that linked administrative data *is not* used for operational purposes; these participants characterised this type of use as "spying" on individuals, and worried what the implications for specific groups might be. The on-going lack of clarity about the separation between operational and social research uses of administrative data (after two days of dialogue and discussion) highlights the potential scale of the communications challenge in informing the public about the work of the ADRCs.

4.3 Who? What type of researchers should be allowed to have access to linked data?

Though suggestions here were diverse, many participants came to the conclusion that, once they had undergone proper training and had been accredited, most types of researcher should have access to linked data. There was near consensus across workshops that academic researchers and those working for charities and the government to be allowed to access linked data, provided they met training and accreditation criteria and their project had also been approved.

They wanted researchers who gain approval for access to be "neutral" "unbiased", "qualified". This view was entirely contingent on the approval process being strict enough to ensure that only those who had the right background, sufficient skills and correct motivations would gain access. In Manchester, participants also suggested compulsory confidentiality clause.

> *"Reassure us that it is a difficult and lengthy process."*
>
> *Belfast*

Few thought that researchers working for private companies should be allowed to have access to linked administrative data. Again, the exception here were some younger participants who thought that private companies were working on behalf of the government this should not pose a problem provided they received ethical approval and had been through the accreditation process. A few also thought that private researchers with *"a good enough reason"* should also be allowed to have access, citing supermarkets who might want to know more about the needs in their area. They also thought that the ADRCs could charge private researchers for access, making them even more cost-effective.

A handful of participants were concerned about allowing PhD students to access linked administrative data. This seemed to be driven by the perception that they wouldn't be well-trained enough, or that they would only be doing the project for the benefit of their own personal research (rather than for a short or medium term social good).[15] These participants thought that PhD students should *"make do"* with data that is already available.

---

[15] Most participants were not familiar with the role of PhD students in universities, and how their training differs from undergraduate students.

"We were umm-ing and ah-ing about PhD students. The more doors you open the more dangerous it is. There's no reason for them to do it. They'll only be approved for their own benefit."

Wrexham

Similarly, very few participants thought that private individuals should have access to linked data, even if they had a project that met the criteria for approval. They struggled to think of reasons why individuals would want to do so, making them suspicious of those who did. The backing of an institution that has something to lose if the researcher breaks the rules or misuses the data was thought to add an extra level of security to the process.

"I don't think it should be anyone, should be a structured organization, government body, UN, WHO. Anyone can misbehave but this reduces the security risks."

King's Lynn

Some also argued for allowing researchers to have different levels of access to linked administrative data or types of linked data based on experience, training and expertise. For example, they thought that those who wanted to access medical records should be subject to higher scrutiny and a more rigorous approval process than those who wanted to access less sensitive material. Similarly, if there was more risk of identifying individuals through linking multiple datasets, more engaged participants argued that the rules in place should be stricter.

4.4 What? Which pieces of data are acceptable to link?

In several areas, participants thought that any government administrative data should be linked. In the context of strict project and researcher approval, few wanted to put specific restrictions on specific types of information. They thought that the privacy and ethical implications would be considered by the approval committees in any case. Those who had understood and trusted the efficacy of the de-identification process were more likely to be relaxed about the type of data that can be linked.

"As long as it's anonymized it's ok."

Cardiff

When discussing this question, level of understanding of social research was also important. One specific group argued that most types of data should be excluded from linking, usually on the grounds that "we already know" this information or because they didn't understand how it would be used.

"But why do you need to know about employment details? You can look at how many jobs are being advertised."

London

Even among those who were generally happy for "anything" to be linked, there were concerns around a few specific pieces of information, either because they could not understand how linking

this data could avoid identifying individuals, or because of the very personal nature of the information made them feel uncomfortable.

- **Full postcode:** No one thought that this should be allowed to stay in a dataset after linking, due to the potential for identification on individuals. Many also objected to partial postcode on similar grounds, although others thought that this was an acceptable level of aggregation.

> *"All data should be allowed to be linked but we think that full postcodes are a little too data specific and should be limited to first one or two letters."*
>
> *Manchester*

- **Details of your home:** It was unclear to a many participants what this meant. Concerns raised included privacy (anything about my home seen as "too personal") and potential for identification of individuals.

> *"If it's just a general, 'how many people live in 2 bed flats?', that's okay but if it's really specific like how many TVs have you got, that's too much."*
>
> *London*

- **Religion:** Participants who did not want this included said that they could not understand how it could be relevant to any research project.
- **Travel patterns:** In some locations, participants had detailed discussion of the use geo-location data. They were doubtful of the possibility of ever properly de-identifying such information, and as a result thought it should not be linked to other administrative data.
- **Date of birth:** Again, this was seen as identifying information particularly when combined with other data. Year or age range was thought to be an appropriate replacement, if age is needed for analysis.
- **Medical records:** Here it depended on the amount of data to be shared as a full medical record was seen as potentially identifying.

Those who discussed the potential for linking data from private companies (for example, details of what you buy) to administrative data instinctively disliked the idea. They could not see the potential benefit of this, and were concerned about the potential privacy implications.

> *"Never personal financial details like bank accounts."*
>
> *Stirling*

4.5 Where? In what location should access to data linking happen?

Overall, participants were happy for researchers to have access to data in a controlled and secure setting within the ADRC, once the researcher had been vetted and accredited according to a strict process.

As described in the previous chapter, security was one of the biggest concerns that participants had about the ADRN plans. However, most of the concerns were with the way data was handled during

the linking process. The post-linking, de-identified data was seen as much lower risk. A handful of participants were completely unconcerned about the security of de-identified linked data and argued that spending money on high-security settings for it is wasteful.

On the other hand, most thought that security remained important even after de-identification, and the fewer points of access the better in order to minimize any risk that the data could be misused or hacked, and also allow for greater accountability, as all times the data has been accessed can be logged.

Participants wanted various guarantees with regard to the security of the physical setting, some of which they picked up from questioning researchers about current arrangements for linking.

- In high security buildings
- No way to extract data from the system (e.g. no USB ports on computers)
- No digital equipment allowed
- Limited numbers of people in the secure area at once

A few questioned whether this level of security is possible within a university setting, suggesting that the data shouldn't leave government facilities, even after linking and de-identification. However, having had the opportunity to speak and question the ADRN representatives, most were happy for the access to take place within one of the four current ADRCs, provided that the guarantees above were in place.

In Cardiff, participants argued that researchers should only have access to linked data within the ONS i.e. they were against the ADRCs operating as a physical environment in which researchers access linked data. This was because they had spoken to an ONS researcher in detail about the security provision that is in place for linking data within the ONS, and they thought that it would be costly and inefficient to replicate this across the country.

> *"Why do we need new places? We have the structure already there within the ONS."*
>
> *Cardiff*

Participants were generally against allowing virtual access to a secure environment containing linked administrative data. The concept of secure remote access is difficult to explain and perhaps was not explained consistently across the workshops. Even in areas where participants asked the experts numerous questions about the concept, they said they still didn't fully understand how it would work.

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, who were convinced that this mode of access is less secure than access within an un-networked physical location. 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 on this topic.

Opposition to remote access was in part driven by a slippery slope concern. Some participants questioned how, once the data is being transferred from one source to another online, can we be sure that it won't be transferred to those who would misuse it?

Those participants who thought that de-identified data is very low risk would be happy for remote access to the linked data, with certain caveats (limited time passwords, logging of all actions). Indeed they thought it would be necessary to avoid capacity issues at the centres and to allow open access to centres and thus more efficient use of the money that had been spent creating them.

4.6 When? How long should the linked data be available to researchers?

Views were divided on this question. Those who were completely happy with security at the ADRCs and the de-identification process tended to argue that linked datasets should be stored. This was so that researchers - both those who carried out the original project and others who wanted to replicate or build on their work - could have access to the linked data after the initial research project for which it was linked is complete.

They pointed out that it would be inefficient and costly to destroy all linked datasets after they have been used once.

> *"That's a waste of money to start all over again for future use. There's no personal information there."*
>
> *London*

Some who advocated this view argued that the linked data should be thought of as a government resource to be continuously exploited by researchers as and when needed. Those who had the best understanding of the social research process argued this particularly strongly, as they thought that the data might be needed for peer review of replication of a study.

Others found the destruction of linked data after project completion reassuring from a security perspective. Among those who had this view, some thought that it should be very strict, with the data destroyed immediately after the researcher has had access to it while some said maybe it could be kept for a few months afterwards to allow the researcher to come back to do further analyses or check their work. The debate on this issue was made difficult by lack of awareness of the timescales of social research projects.

> *"For as long as they are in the room. If you forget something, that's your fault. If you want more information you have to link it again."*
>
> *Stirling*

One concern that was driving this view was relevance. As outlined in Chapter 2, it was hard for participants to imagine how historical or 'out-of-date' data could be of relevance to researchers, and therefore presumed that it is always 'real-time' data that is linked. This meant that they didn't understand why it would be useful to store linked data for more than a few months.

However, after debate, some participants settled for a rule that allow researchers to access the linked data for *"as long as they need it",* perhaps subject to reapplication to the ADS or ADRC if a lot of time had lapsed. This would necessitate keeping the linked datasets indefinitely.

> *"They [databases] should be archived but you still have to go through the same permissions - same checks and application."*
>
> *Stirling*

Indeed, some participants reflected back to the reasons for data linking when discussing this question, suggesting that there is no point in having strict rules. Instead they argued that the length of time that the data is held should be project-dependent. They thought that only those with a good understanding of how long the data would be necessary for could make that decision.

In some areas, participants favoured the compromise idea of retaining the code or syntax that allowed the linking of two datasets so that they could be easily and cheaply relinked, without having to do the same work a second time.

4.7 Considered views of principles and rules within the ADRCs

Participants were generally satisfied with the suggested rules that were presented to them. They agreed a broad 'why' criteria from which all the other rules follow i.e. they agreed that once that project has social value, then it should be allowed to go ahead, with most of the other rules being project dependent.

Participants often suggested deferring to the authority of experts to set detailed rules for the data linking process within ADRCs. This was because the specifics of the detailed data linking process were of low salience, and because the participants thought that they didn't have enough information about all of the types of projects that could happen to decide on strict criteria. Once participants were convinced that the overall process is rigorous, and approval committees are run by unbiased qualified people, then they didn't mind if those approval committees made project by project decisions on who exactly has access to linked data, what particular data they can link and the length of time for which the data can be held.

The main exception to this general preference for expert decisions was in relation to the involvement of commercial companies, where participants argued for strict and specific rules. More research would need to be done to fully understand attitudes towards allowing business data to be linked or allowing businesses to link administrative data – while there is instinctive distrust, attitudes tended to change once potential benefits were discussed.[16]

Further public acceptability research would also need to be undertaken if researchers want to be allowed to link passively collected administrative data. Participants in this dialogue questioned how such data can be aggregated and properly de-identified. Concern about whether this is possible meant that those who discussed this issue tended to want passively collected administrative data to be excluded from the ADRC initiative.

---

[16] While the potential for commercial companies to access linked data was discussed across workshop locations, this issue was not deliberated in depth, and therefore the finding should be seen as spontaneous/topline only. Views may have changed if participants had heard from researchers working for commercial companies, or had had access to case studies that outlined the benefits and potential drawbacks of commercial access to linked administrative data.

The most important conclusion participants came to was that it is vital for the ADS and ADRCs to create rules to ensure that all data held and linked within the ADRN is safe and secure at all times. Even where they were convinced that the linked administrative data would be fully de-identified – and not all were – this usually meant that they were strongly in favour of secure physical settings and against remote access.

Chapter 5: Information and Engagement

All ADRCs have been tasked with developing a public engagement strategy, and must employ a public engagement and communications officer. Currently these strategies are at an early stage of development. Participants in the dialogue were invited to share their overall views on how to inform and engage the public about the ADRN and its work, and to comment on some initial ideas that have been suggested by the ADRCs.

This chapter describes the information that participants thought would be most useful to the wider public in explaining ADRN plans. It also examines participant views on most appropriate level of engagement and information for the wider public.

Key findings

- Participants were split between arguing that the ADRN should actively communicate as much as possible about the ADRN initiative, and suggesting that only minimal information should be disseminated.
- De-identification, the benefits of linking, security and the similarities with what was possible under the previous system were the main ideas that participants felt were important to communicate.
- In line with their view that public interest in this topic is low, most participants suggested that ADRC information plans should be broad rather than deep.
- Participants were similarly split on the level of public engagement that is required for ADRCs. Some feel that only those who have a strong interest should be involved, whilst others favoured wider public participation.
- There was consensus that whatever the level of public engagement, the work on the ADRN should be as transparent as possible so that anyone who seeks out information should be able to find it easily.

5.1 Informing the public?

The relationship between the level of information participants had about data linking and their views of the proposed ADRC network was more complicated than might have been expected.

Over the first day of the workshops, some participants became increasingly concerned as they found out more information about the data linking process, but this often this subsided once they heard about the principle of de-identification, and had the opportunity to discuss the process in full. Beyond this point few participants opposed the ADRN initiative in principle, or had significant concerns about the data linking process.

Based on their experience of the dialogue process, opinions varied as to the amount of information that about data linking and ADRCs that should be disseminated to the public. Two main views emerged; that people should either be fully informed, or not given much information at all.

The main argument in favour of limited information was that the subject matter is so complicated that it would be difficult to achieve a level of understanding amongst the general public that would prevent people from worrying unduly about their privacy and the security of their personal information. It was noted that some information about the data linking process, for example security, had made them feel more distrustful initially, until they had had a chance to discuss fully and ask clarification questions.

> *"I wouldn't [tell people]. I think you're going to have so many obstacles to get around, hurdles to get over, if you just tell someone about this."*
>
> *Belfast*

Younger participants were particularly likely to make this claim about older people, who they feared were less likely to understand the process and more likely to worry unduly.

It was argued that most people would find this topic uninteresting and too complicated to engage with. Participants thought that the topic has little relevance to everyday life, and that people are likely to ignore any information about it. The complexity of the area, combined with a lack of interest, would make it difficult and counterproductive to try to inform people about the proposed changes in detail.

> *"It's taken the best part of eight hours to get [data linking] across to us, who are actively interested. It'll be more difficult for a member of the public"*
>
> *Cardiff*

However, participants also thought that communicating nothing about the ADRCs would ultimately be counter-productive. They expected that people would hear about it eventually, and may come to suspect the government of trying to hide something. Some argued that people are increasingly educated about and interested in this topic, and as such would want to know more.

Therefore, participants were clear that there should be the facility for people to know what is going on in ADRCs, but were less sure on how much of this information needed to be broadcast widely. Transparency was seen as vital. Participants were unanimous in their opinion that detailed information about projects, processes security and impact should be available to those who are interested, for example via ADRC websites.

5.2 Key messages to communicate

Participants in a number of locations (where there was time) were invited to create newspaper headlines that they felt would succinctly explain the key points about data linking and ADRCs to the general public. A number of key themes emerged across the headlines they created, including the facts that:

- The data is de-identified
- There are security safeguards to prevent information being leaked or stolen
- Research conducted from data linking is socially beneficial

- The new system does not represent a major change from what is currently possible, but will make things more efficient.

De-identification

Across all groups, participants felt that the most important piece of information to communicate about the data linking process was that all personal data is de-identified or, more understandably, "anonymised". Reflecting on their own experience during the dialogue workshops, they felt that it was only once the concept of de-identification had been fully demonstrated that they felt comfortable with the ADRN plans.

> *"Basically they need to know that they are not an individual and they are part and parcel of some correlation… they're just figures."*
>
> *Wrexham*

Security and location

Another important message was the security of the systems the ADRCs would operate. Participants said that it was important to reassure the public by showing that all aspects of the data linking and access process are secure, including the stringent application process and the physical security of the facility.

> *"[Tell the public] it can only be accessed by certain people – and that there will be checks in them."*
>
> *Belfast*

Reflecting their general cynicism about the way that private companies use data, they suggested reassuring the general public by emphasising the fact that private companies are not allowed to access the data.

> *"And let them know it's not third party services that are getting access to the data."*
>
> *Belfast*

Benefits

Reflecting on their own experience throughout the workshops, participants suggested that the most interesting information to provide would be how the data is used. *"Good news"* stories about research could be published to provide positive examples of how data linking could be used. Many of the good news stories suggested by participants related to the application of data linking to public services, particularly health and schooling.

> *"I'd like to see some positive outcomes from the results of the result. Something good coming of it. Not just for someone's personal research."*
>
> *Stirling*

Participants thought that the public would, like they were, be open to finding out more about research, but that this would more interesting and relevant when it is tied to a real life outcome.

Some also suggested that this would in some way atone for the fact that consent was not sought for the use of the data for this purpose.

<u>Other messages</u>

There were a variety of other factors that some participants felt were important to mention, although these were discussed less often:

- o Highlight the continuity between the new and old systems. Some participants were concerned that members of the public would have a hostile reaction to a new system, so thought that it would be important to emphasise the fact that the system was streamlining a process that already exists.

> *"Tell them that all this is [the old system] coming together…"*
>
> *Wrexham*

- o Explain that this system would not be particularly costly, and that it may actually save the government money through more efficient use of data. Others wanted ADRCs to be transparent about cost and economic impact, because they were sceptical about the suggested savings.

> *"The data is already being used but it's costing so much…need an explanation of how it works and expenses compared with doing [the research] without data linking."*
>
> *King's Lynn*

- o Publish details of what studies are underway, and who is conducting the research, and for what purpose.

In Belfast participants were of the opinion that other countries hon the UK, and suggested that the creation of the ADRN could be advertised as a move to keep up with other countries.

> *"[Tell them] that it's moving forward, keeping up with other countries… in some countries they have so much data already."*
>
> *Belfast*

5.3 Methods of informing the public

When asked about the best methods of informing the general public about the ADRN plan, television and radio advertising were popular suggestions. However, when this idea was explored further, participants were unsure what such television advertising should say or aim to achieve. Other suggestions included national pamphlet campaigns, roadside billboards, and banner advertising on the internet. Any mass campaign would need to be explained in *"clear, simple English"*.

The main suggestions put forward by participants are outlines below:

- A national newspaper campaign that would let people know about the role of ADRCs and the sort of information they would be handling, alongside reassurances about security and scope. There should be links to a website with more information for the small proportion who might want to find out more.

> *"You should put it in a national newspaper – the Times, not the Lynn news!"*
>
> *King's Lynn*

- A more local approach, with an information sheet or pamphlet handed out with annual council tax leaflets or local papers. Those who suggested such an approach worried that a national newspaper campaign might backfire, causing concerns about data linking amongst the public, who might find out enough to be aware and worried, but not enough to be reassured.
- Providing information through local libraries and schools to reach as wide audience who might not be engaged by less personal methods.

> *"Someone from the library could be trained on this and if someone has any questions to ask them."*
>
> *Manchester*

- Engagement through the publication of success stories, linking the work of ADRCs to progress in feel-good areas such as education and health.

> *"People want to know what the benefits are, e.g. improving education, health services. Share the results and outcomes with the public."*
>
> *Wrexham*

Reflecting on their own experiences during the dialogue, participants suggested that researchers themselves should be involved in any information campaign, as hearing about real researchers' motivations, and how seriously they take ethics and security had really helped them understand this issue.

Some participants completely opposed all of these suggestions on the grounds of costs, and thought that a website for each ADRC would suffice. Many agreed with keeping costs low in principle, but also argued strongly for a Freephone helpline/FAQ number for those who did not have access to the internet.

5.4 Engaging the public

Participants were split as to whether it is necessary for the public to be directly engaged in the work of the ADRCs, and the extent to which they should be engaged. Overall, while demand for personal engagement with the Administrative Date Research Centres was low, there was some interest in participation in decision-making from a few individuals, including many older participants in Cardiff. By contrast, in the younger King's Lynn group especially, a number of participants felt that there was no need for the public take part in decision making within ADRCs.

> *"I don't think people should be involved in ADRCs. It would be beneficial to have a spokesperson telling them what's happening, but I don't think the general public should be involved."*
>
> *King's Lynn*

However, participants for the most part - whatever their level of interest in being involved personally - thought that there should be some place for the public in the running of ADRCs, particularly for those members of the public who feel strongly about it. They were reassured by the idea that the public/layman would be represented in some way, making sure that their interests were being served.

> *"If you know there are a couple of members of Joe public in that committee…then you'll know your voice is going to be spoken for."*
>
> *Belfast*

In several areas it was suggested that if ADRCs want public representatives involved in decision-making, they should be pay members of the public to do so.

Participants also worried that the level of complexity of the topic was such that members of the public taking part in decision-making would need to be carefully picked, or perhaps have a professional interest in the area in some way.

> *"They would have to have some clue what is going on…ask them some questions to work out are you lucid enough, are you discriminating?"*
>
> *Wrexham*

<u>Engagement methods</u>

Amongst the few who supported a comprehensive engagement approach, popular ideas included general public participation on the board of the ADRC, or of a *"citizens' panel"*. There was also some support for a discussion forum on ADRC websites as a low-cost way of engaging people across the country who might be interested in having input into the work of the ADRCs.

Participants in Wrexham and Belfast were particularly keen to get people who *"live in the real world"* onto a panel that could oversee research projects, suggesting mass participation and the establishment of a jury service-style system.

Others thought that a more detailed advertising campaign would be necessary to generate interest. In several groups an analogy was drawn with tax self-assessment; self-assessment is an important and complicated, but not particularly exciting process which many people have to complete:

> *"If you look at how the tax man campaigns for self-assessment… why not just do the same, and say all this will be used for the betterment of society and it is to use the money in a better way."*
>
> *Stirling*

Another suggestion along this line was the use of interested national bodies and impartial observers who are trusted by the public to advertise the ADRN initiative and encourage a wide range of the public to engage.

Other more active methods of engagement were less appealing to participants, who thought that they would not be interesting to many members of the public. There was some support for ADRC open days, although the usual participant reaction to this suggestion would be to say that there wouldn't be much to see, owing to the strict security requirements of ADRCs.

> *"I don't know whether ADRC open days are a good idea. People who are strongly opposed might turn up but people in favour wouldn't. Put it on the weekend and people would rather watch football!"*
>
> *King's Lynn*

Chapter 6: Conclusions and recommendations

This chapter outlines the most important findings from the dialogue process, and the lessons that the ADRCs, ADS, ESRC and ONS can draw from them. It should be noted that there were few major differences across the locations where the dialogue workshops were held, and all of the conclusions here are pertinent to each of the four countries where the ADRCs are based.

<u>The context for the ADRN</u>

Participants had **little prior familiarity with or understanding of the purpose and processes of social research**. Even by the end of the dialogue workshops, social research was perceived has having limited relevance for most people's everyday lives. In addition, participants found the operational and research uses of data difficult to separate conceptually.

This implies that the general public would need to be informed about how research works at the broadest level before they could begin to understand how linked administrative data might be useful for research purposes. This demonstrates the **scale of the challenge of fully informing the general public** about the work of the ADRN.

As dialogue participants were shown information about research projects that have used linked administrative data, asked questions of experts, and heard the opinions of other participants, their opinions of the ADRN plans changed, sometimes moving between negative and positive perceptions several times. The broad overall pattern was that **as participants gained *understanding* of social research and the ADRN plans, they tended to become less concerned and more supportive**.[17] However, becoming informed enough to gain this level of understanding took at least a day of intensive discussion, which would be unrealistic to replicate on a broader scale.

Participants did take a **strong interest in the uses and outcomes of social research**. Any general public communications around the ADRN should therefore focus on the societal benefits of social research, which could be very broadly defined. This would also help the public to understand the need for and the value of the network.

The centres should ensure that they develop strong **case studies** that show how the projects that they have enabled have led to **policy change, or service improvement**. Projects that have most relevance and interest to the general public - participants mentioned the subject areas of health and education - should also be highlighted. Similarly the value of historical and longitudinal research is not obvious and needs to be demonstrated in a simple, clear way.

<u>Informing and reassuring the public about administrative data linking</u>

The findings suggest that the public would be **broadly happy with administrative data linking for research projects provided (i) those projects have social value, broadly defined (ii) data is de-identified, (iii) data is kept secure, and (iv) businesses are not able to access the data for profit**. Messaging around the ADRN should focus on addressing the first three issues, and further work should be done to fully understand the public's view on business access (as this was not explored in any depth in this dialogue).

---

[17] This observed correlation would be useful to investigate using quantitative research.

Underpinning acceptance of the ADRN plans was a desire to see the government make better, more efficient and cost-effective use of data, avoiding burden on the population resulting from repeatedly collecting the same information. Support for the ADRN was also driven by how bounded its remit is. By the end of the dialogue, participants could see little danger in their de-identified administrative data being used for individual projects by academics or government researchers. It is therefore important to explain clearly how the ADRN is designed to streamline and make an existing process more efficient, cost-effective and even more secure.

Participants wanted the **research using linked data to have impact as its goal**. They suggested that researchers who want to use linked data should submit concrete plans for publishing their work, as well as ensuring that relevant policymakers are aware of the results.

The key area where there could be public concern about the ADRN plans is around **de-identification**. Confidence in the process by which this happens is crucial to creating support for linking administrative data. But the dialogue shows that this process is very difficult and time-consuming to explain.

The simplified version discussed in the dialogue was not detailed enough for some participants. The network should undertake more work to develop the several ways to explain de-identification in a way that different audiences readily understand. Language will be important here; for example, participants in the dialogue used the term "anonymous" interchangeably with the technically correct term "de-identified", as the former is more familiar to them. Method of explanation will also be key. The visual technique used in the dialogue workshops to explain the data linking process was effective in helping participants conceptualise the process. This suggests that explaining data linking visually (e.g. through an animated video or interactive diagram) may be a useful way to help people understand this issue.

The public also need to be **convinced that the security of their data (including their de-identified data) is of the utmost concern** to the ADS and the ADRN. They were clear that rigorous policies and processes should be put in place to ensure this. Participants were generally happy that the potential processes as described were sufficiently rigorous, particularly in ensuring that only those with the right motivations and qualifications would gain access to linked data.

Security was especially important for those who were unclear about the concept of de-identification, and those who were not convinced that full de-identification is really possible for some types of administrative data. This high level of concern around security should be reflected in ADRCs' and ADRN operational plans. It was the one area where participants were unwilling to defer detailed policies to experts or authorities.[18] For example, any move towards virtual access to this data would need to be clearly demonstrated to be *as secure* as access through a secure physical setting.

Views of the ADRN plans

Overall, participants were reassured by the idea that the ADS would have strong governance and be regulated. However, there were some clear 'red lines' in participants' views about the limits they

---

[18] On other issues, for example the length of time that linked data can be held, participants were generally happy for that to be decided on a project by project basis by *"the experts"*.

would like to see imposed on the ADRN. **Further public dialogue would be needed for any expansion of the ADRN's remit**, in particular with relation to:

- Creating large databases containing many variables/data from a large number of public sector sources
- Allowing administrative data to be linked with business data
- Linking of passively collected administrative data, in particular geo-location data

All of the above were seen as having potential privacy implications or allowing the possibility of re-identification of individuals within datasets. The other 'red-line' for some participants was allowing researchers for private companies to access data, either to deliver a public service or in order to make profit. Trust in private companies' motivations were low. More research would be required to understand detailed views on this issue.

Some participants in the dialogue were frustrated that the public had not been involved earlier in the process of planning and setting up the ADRN. This suggests that, to ensure credibility, any further public acceptability research or dialogue should happen as early as possible in the planning process, while the public view can still influence policy and procedures.

Engaging the public with the ADRN

The views of dialogue participants suggest that there may be low interest from members of the general public in becoming actively involved with the ADRCs or the ADS. Even so, ensuring **public representation in the decision-making process** could be important in proving the transparency of these bodies. Recruiting those with pre-existing interest or experience would be more efficient, but might make them less representative of the public at large. On the other hand, members of the public would need to be brought up to speed about research methods and processes before being able to engage meaningfully with proposals for projects. Dialogue participants were split on this issue, with some favouring previous knowledge and experience and others representativeness.

The overall findings of this dialogue, particularly with relation to lack of familiarity and public confusion about social research, suggest that the communications strategy for the ADRN should focus on transparency. Many participants saw **transparency as both important and sufficient**, and were concerned about too much effort and money being used to inform the general public about such a complex topic. A comprehensive website for each ADRC with as much information as possible - in clear, plain English - about the linking projects and their impact would most likely be enough to satisfy the information needs of the general public, allowing those who are most interested to find out as much as they want. The website would also need to have sufficient detail on linking processes and security.

Appendix: Contents of Methodological Annexe

Please see the methodological annexe, published in conjunction with this document, for further technical details. The contents are as follows:

- Project governance

- Project design

- Sample design and recruitment

    o Locations

    o Participants

- Stimulus and workshop design

- Fieldwork

- Experts

- Analysis and reporting

    o Reconvened reporting workshop

- Workshop materials

Sarah Pope
Research Manager
Ipsos MORI Social Research Institute
sarah.pope@ipsos.com

Daniel Cameron
Associate Director
Ipsos MORI Social Research Institute
daniel.cameron@ipsos.com

For more information

About Ipsos MORI's Social Research Institute

The Social Research Institute works closely with national governments, local public services and the not-for-profit sector. Its c.200 research staff focus on public service and policy issues. Each has expertise in a particular part of the public sector, ensuring we have a detailed understanding of specific sectors and policy challenges. This, combined with our methodological and communications expertise, helps ensure that our research makes a difference for decision makers and communities.