



# 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

## Table of contents

1. Project governance .....	2
2. Project Design .....	3
3. Sample design and recruitment.....	3
4. Workshop and stimulus design .....	5
5. Fieldwork .....	5
6. Experts .....	6
7. Analysis and reporting .....	7
Appendices: Workshop materials .....	9

This annexe outlines all of the key methodological details for the Dialogue on Data carried out by Ipsos MORI on behalf of the Economic and Social Research Council and the Office for National Statistics. Any queries should be directed to the project team at Ipsos MORI: [Sarah Pope](#), [Daniel Cameron](#) and [Michael Clemence](#).

## Project governance

This dialogue formed part of the wider Public Attitudes to Science (PAS) 2014 project commissioned by the Department for Business, Innovation and Skills (BIS).<sup>1</sup> That project included a survey module which provided nationally representative statistics on the general public's views on a range of potential uses of big data. This dialogue was commissioned to gain a more detailed understanding of the *range* of public views around this issue.

The Dialogue on Data was conducted by Ipsos MORI on behalf of the Economic and Social Research Council (ESRC) and the Office for National Statistics (ONS). Its two key purposes were to examine public attitudes towards the use of government administrative data linking for research purposes, as well as public opinion to potential revisions to the methodology employed in the UK Census.<sup>2</sup>

The direction of the project was overseen by a Steering Group incorporating stakeholders from ESRC, ONS, the Administrative Data Service, BIS, academia and other interested organisations. The steering group met three times over the course of the project:

Meeting 1: Discussion of dialogue design including objectives, sampling and materials design process

Meeting 2: Comment on dialogue materials

Meeting 3: Comment on dialogue findings and first draft of the report

The members of the steering group are as follows:

- **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

<sup>1</sup> The full PAS report can be found at [www.ipsos-mori.com/pas2014](http://www.ipsos-mori.com/pas2014)

<sup>2</sup> The findings for the portion of the dialogue examining the Census will be published separately later in 2014.

- **Maria Sigala**, Senior Policy Manager, ESRC

Members of the core team at Ipsos MORI responsible for conducting the research attended the steering group meetings and incorporated comments and suggestions from the group into project design, materials and reporting.

## **Project design**

Public dialogues bring together members of the public, policy makers, scientists and other expert stakeholders to deliberate and come to conclusions on complex policy issues. This approach was chosen as the best way to engage the public on this area, given the complexity and low everyday salience of social science research.

The dialogue required members of the public to suggest guidelines for acceptable behaviour in complicated policy areas. As participants typically came to the events with low existing knowledge of the relevant issues, it was important to ensure that participants achieved a level of fluency in these areas before they could make suggestions. Holding dialogue events over two full days meant that in the first event the emphasis could be on educating participants about data, administrative data, and data linking, whilst the second event could focus on investigating what participants thought was acceptable and unacceptable about administrative data linking. This approach also allowed observation of how participants' views on the topic changed as new information was provided, and allowed the use of a variety of qualitative approaches to gather additional data, including plenary groups, small group discussions and activities.

The involvement of experts in the discussion was an important feature of the dialogue. Experts provided presentations and background information to help educate participants in the topic area, and were able to field questions that the facilitators could not answer.

The two dialogue workshops in each location occurred on Saturdays spaced a fortnight apart. In addition to providing more time for discussion, this structure allowed greater consideration of the topic by participants, who could think about the topics at home in the intervening fortnight and talk about it with friends and family.

## **Sample design and recruitment**

This project employed a purposive sampling approach to ensure that the location of the workshops and the demographic make-up of the participants provide an overview of the range of people living in the United Kingdom.

### Locations

It was important that the locations selected reflect the diversity of opinion across the UK. As a result, the location sampling distribution needed to encompass all four nations. The seven workshops were spread across the UK in a 3-2-1-1 pattern, with three in England (north, south and rural), two in Wales (north and south Wales), and one each in Scotland and Northern Ireland.

Workshops were concentrated in England and Wales because the census-related portion of the dialogue was only relevant in these countries. Three sets of workshops were held in England and two in Wales due to the need to reflect geographical diversity in both countries.

Country	Location
England	London
	Manchester
	King's Lynn
Wales	Cardiff
	Wrexham
Scotland	Stirling
Northern Ireland	Belfast

### Participants

Within each location 18-20 participants were recruited face-to-face by specialised and accredited qualitative recruiters. Participants were given a cash incentive to take part. Quotas were set on various demographic attributes (age, gender, social grade and ethnicity) to ensure that the participants were broadly reflective of the population in each of the locations. The exceptions were the workshops in Cardiff where all participants were recruited to be aged over 55, and King's Lynn, where all participants were recruited were aged under 35. This approach was taken to allow researchers to analyse differences in attitudes by age.

Quotas were also set key attitudinal questions taken from the Public Attitudes to Science 2014 survey, to ensure that a range of attitudes towards 'big data' were reflected in each workshop. The selected questions were:

1. **"How much, if at all, do you trust researchers to follow any rules and regulations which apply to their profession?"** At least 3, up to a maximum of 6, of those who said either "not very much" or "not much at all" were recruited to each group.
2. **People's data can be used in many ways to change public services. In each of these instances, the data is anonymised, so it can't be linked back to individuals. To what extent do you support or oppose combining the data held by multiple government departments and using them to better tailor public services?** At least 3, up to a maximum of 5, of those who said either "tend to oppose" or "strongly oppose" were recruited to each group. This quota was not applied in King's Lynn or in Cardiff, to avoid adding additional complications to the focus on recruiting to age quotas.
3. **Here are some services through which organisations can collect data about people. Which, if any, of these services have you decided not to take up because of concerns about how your data would be used?** At least 7, up to a maximum of 9, of those who said they had decided not to take up at least one service were recruited to each group. This quota was not

applied in King's Lynn or in Cardiff, to avoid adding additional complications to the focus on recruiting to age quotas.

The target quotas and demographic profile of the participants who took part are detailed in the table overleaf:

	Total number	Age	Social grade	Gender	Number of participant from BME backgrounds	Q1	Q2	Q3
<b>Quotas</b>	>126	18-24: >23 25-34: >23 35-44: >15 45-54: >15 55-64: >23 65+: >23	ABC1: >42 C2DE: >42	Male: >56 Female: >56	>32	15–30 as above	15–30 as above	35-45 as above
<b>Attendees</b>	136	18-24: 24 25-34: 27 35-44: 16 45-54: 17 55-64: 23 65+: 26	ABC1: 70 C2DE: 54	Female: 68 Male: 68	22	38	28	54

Most quotas were met, although recruiting participants from black and minority ethnic backgrounds proved difficult in some areas; particularly in Wrexham and Belfast.

### Stimulus and workshop design

The dialogue workshops and accompanying stimulus were both designed by the core Ipsos MORI team, in close communication with the clients and the steering group, and after telephone consultations with several interested stakeholders who were not part of the steering group. Prior to the initial design of the materials, extensive background research into the area was carried out to see where the project could contribute to knowledge most effectively.

After conducting background research the next step was to design the discussion guide and stimulus for the project; both of which required assistance from the clients and steering group. The stimulus for the first dialogue event required examples of administrative data research, and both ESRC and ONS provided examples that Ipsos MORI turned into stimulus. After the pilot workshop some materials were reviewed in collaboration with the clients and steering group.

All stimulus and workshop materials were reviewed and approved by the steering group prior to their use.

## Fieldwork

Fieldwork took place on Saturdays between the 5<sup>th</sup> October and 9<sup>th</sup> November 2013. The London group was conducted first, acting as a running pilot group for later groups. The order and dates of the groups were as follows.

Each workshop was facilitated by four Ipsos MORI researchers; two lead facilitators and two junior facilitators and note takers. In addition a number of observers and experts were also present at each group. These visitors were briefed prior to their participation in the group and all dialogue participants were made aware of their identities and roles.

Location	Dialogue group 1 date	Dialogue group 2 date
London	5 <sup>th</sup> October 2013	19 <sup>th</sup> October 2013
King's Lynn	12 <sup>th</sup> October 2013	26 <sup>th</sup> October 2013
Stirling	12 <sup>th</sup> October 2013	26 <sup>th</sup> October 2013
Manchester	19 <sup>th</sup> October 2013	2 <sup>nd</sup> November 2013
Cardiff	19 <sup>th</sup> October 2013	2 <sup>nd</sup> November 2013
Belfast	26 <sup>th</sup> October 2013	9 <sup>th</sup> November 2013
Wrexham	26 <sup>th</sup> October 2013	9 <sup>th</sup> November 2013

## Experts

The presence of expert witness and observers is one of the key strengths of a dialogue approach. Every dialogue event was attended by a number of experts, clients, and steering group members; each provided valuable insight and an alternative perspective on the issues under discussion. The full list is provided below:

Date	Location	ADRC representative	Other expert	Steering Group and observers
5 <sup>th</sup> October	London (1)	Tanvi Desai	Tak Wing Chan, Adil Deedat	Melanie Knetsch; Genevieve Groom
12 <sup>th</sup> October	Stirling (1)	Chris Dibbon	John Frank,	Sarah Cunningham-Burley; Paul Grice
12 <sup>th</sup> October	King's Lynn (1)	Melanie Wright	Peter Jones	
19 <sup>th</sup> October	London (2)	Peter Smith	Elpida Prasapoulou	Maria Sigala; Genevieve Groom
19 <sup>th</sup> October	Manchester (1)	Elaine Mackey	Dr Andy Fohrmann, Jennifer Wall	Lynn McKeague; Neil Serougi
19 <sup>th</sup> October	Cardiff (1)	David Ford	Adil Deedat	Vanessa Cuthill
26 <sup>th</sup> October	Belfast (1)	Dermot O'Reilly, Helen Dolk	Dr Mark McCann	Christina Rowley

<b>26<sup>th</sup> October</b>	<b>Wrexham (1)</b>	David Ford	Pete Jones	Greg Meredith
<b>26<sup>th</sup> October</b>	<b>Stirling (2)</b>	Chris Dibbon	NA	Sarah Cunningham-Burley
<b>26<sup>th</sup> October</b>	<b>King's Lynn (2)</b>	Melanie Wright	Joanna Lake	David Walker; Genevieve Groom
<b>2<sup>nd</sup> November</b>	<b>Manchester (2)</b>	Mark Elliot	Minda Phillips	David Carr; Naomi Beaumont, Rebecca Fairburn
<b>2<sup>nd</sup> November</b>	<b>Cardiff (2)</b>	Kerina Jones		Genevieve Groom; Jacky Clake
<b>9<sup>th</sup> November</b>	<b>Belfast (2)</b>	Dermot O'Reilly; Helen Dolk	NA	Melanie Knetsch
<b>9<sup>th</sup> November</b>	<b>Wrexham (2)</b>	David Ford	NA	Vanessa Cuthill, Genevieve Groom

### Analysis and reporting

For most of the day the dialogues were conducted in two groups of 8-10 participants, with one facilitator and one note taker in attendance on each table. The note takers took verbatim notes of discussions in this setting, and these notes served as the raw data used in analysis. Additionally, all focus groups were audio recorded with digital audio recorders – this allowed researchers to refer back when notes were unclear or were interrupted for any reason. Where possible, small group activities and standing exercises were recorded by note-takers or on the digital recorders. Participant data generated on post it notes were transcribed for analysis.

Analysis was carried out on an iterative basis throughout the project. The first analysis session occurred immediately after the pilot London workshops for both days. There the team discussed topline themes with the client and attendant experts in order to judge how well the session went and to agree changes to the order of the workshops and the stimulus materials.

Within Ipsos MORI, weekly analysis sessions were conducted throughout the fieldwork period with the facilitators who had recently attended or were about to attend workshops present. Emerging findings from the different focus groups were discussed to help establish the central project narrative, and to provide those about to facilitate workshops with additional prompts or new areas to explore based on the findings to date. At the end of the fieldwork period a larger analysis session was held with all facilitators and the core team, where the findings from all the groups were analysed and compared.

Based on the analysis sessions a topline findings document was drawn up and circulated for comment. The structure of the report, based on these findings, was then discussed and agreed between Ipsos MORI, ESRC and ONS. At this point full data from the workshops were analysed in detail in order to draw out the range of experiences and views, identifying similarities and differences and interrogating the data to seek to explain emergent patterns and findings. During this time a reconvened participant reporting workshop was also carried out (see below).



The event evaluations were also digitised, and this was analysed in a short, separate, report looking at how participants felt that the event went.

The first completed draft of the report was sent to the steering group for their thoughts, and a full steering group meeting was held to discuss the content of the report and its findings.

#### Reconvened reporting workshop

During the reporting process ten participants from the dialogues were invited to Ipsos MORI's London office to discuss the dialogue findings with the Ipsos MORI team and representatives from ONS and the Administrative Data Service (ADS). The intention was to evaluate the research and reporting process, to see if the findings tallied with their impressions of the event and their personal opinions, and also to provide transparency by ensuring that participants that their opinions were being represented faithfully. Participants attended from King's Lynn, Cardiff and London, and their comments were used to refine the analysis.

## Appendices: Workshop Materials

The dialogue events were structured to take advantage of the reconvened element. Participants' were informed about the subject areas under discussion during the first event, prior to discussing their views of administrative data linking and the proposed ADS plans, as well as changing the Census methodology.

Full discussion guides and materials are included below. The London workshops functioned as pilots, with final materials agreed after these had taken place.

## Appendix 1: Workshop 1 Discussion Guide

### Discussion Guide for Workshop 1 of Public Dialogue on Data

#### V2 151013

This discussion guide outlines the discussion that will take place between members of the public, facilitators and experts in the field at the first of a pair of workshops. There are 7 workshops taking place around the country over the course of October and November 2013. The dialogue has been commissioned by the Economic and Social Research Council (ESRC) and the Office for National Statistics (ONS), with the following objectives:

#### Project objectives

1. **To better understand people's views on the linking of administrative data**, and work out how different messages about the process by which the data is linked affects attitudes.
  - Specifically, to explore how the following aspects of the linkage process affect views:
    - Why - The end use of data (administrative/statistical/operational) and people's ability to understand the difference between different potential uses; how do people understand 'benefit' and what kind of trade-offs that may exist between data protection and privacy and the value of data analysis.
    - 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 (matching records from public and private sector providers).
    - Who – The identity/qualifications/sector/motivations of the people carrying out the linking
    - Where – The impact of the environment in which the linkage is to take place
    - 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
    - How – is data-linking acceptable in the case of a potential 2021 administrative data Census?
2. **To begin the process of creating a language around the re-use of administrative data and data linking that is understandable to the general public**
  - What sorts of assurances are required, and what sort of language is necessary, to help the public to understand data linking. Are there any 'red flags', or key words we need to think about when talking about this area?
  - What concepts need to be conveyed to get across the key principles, dilemmas and arguments in this area?
  - What do those who work in this area need to do to earn the public's trust?

3. To help inform the development of the governance and operational procedures that ADRCs will adopt, and provide data on public attitudes for ADRCs to help inform their future strategies and priorities for public engagement
- How much do people want/need to know in order to be properly engaged in the process? How should any arrangements be future-proofed?
  - Are there regional or demographic differences in the level of interest in, opposition to, and engagement with data linking?
  - Who needs to be prioritised for engagement, why and how?

### Using this guide

We use several conventions to explain to you how this guide will be used. These are described below:

Timings	Questions and Prompts	Objectives
5 mins	<p><b><u>Underlined = Title:</u></b> This provides a heading for a sub-section</p> <p>Plain text = Information This is information for the facilitator to give to the participants (in facilitators on words)</p> <p><i>Italics = Instructions.</i> This tells the facilitators how to organise activities or record participant responses</p> <p><b>Bold = Question or read out statement:</b> Questions that will be asked to the group if relevant. Not all questions are asked during the workshop based on the facilitators view of progress in answering the research questions</p> <p>- Bullet = prompt: Prompts are not questions – they are there to provide guidance to the facilitator if required.</p>	<p>This area is used to summarise the objectives or each section - for information only.</p> <p><i>It also shows what material (e.g. slides, handouts) is needed for each section</i></p>
How long it takes	Typically, the researcher will ask <b>questions</b> and use the prompts to guide where necessary. NB: Not all questions or prompts will necessarily be used in an interview	

Time	Process	Objectives and materials needed
0930-0940	Facilitators to welcome experts and observers, run through their roles and rules of engagement	Guidelines for observers/experts
0940-1000	Meet and greet participants	
1000-1020	<p><b><u>Welcome and Introductions</u></b></p> <p>IN PLENARY:</p> <p><i>Use Slides 1-2. N.B. this is not a script but points to be talked through</i></p> <ul style="list-style-type: none"> <li>• Welcome from Ipsos MORI – independent research company. Introduce team.</li> <li>• Today we are talking about how people’s information (DATA) that is collected by the government (ADMINISTRATIVE DATA) is linked together to find out more about society.</li> <li>• Explain clients and aims – ESRC (understand public’s views on administrative data linkage, proposed rules for new ADRCs and how the public think they should be engaged on an on-going basis) and ONS (explore public views on potential changes to the census in 2021 – talk about this next week)</li> <li>• Challenge for the day is to think about how people’s information should be used by researchers to help understand society better</li> <li>• <b>Reassure participants.</b> Participation does not need to be based on your knowledge of research or how data is used; the aim of this discussion is to understand how you think that researchers should use data to understand society and plan/run public services, and the principles that should be followed when linking up data.</li> <li>• How the session will run e.g. plenary, group discussion and exercises - run through agenda (slide 3)</li> <li>• Introduce observers /experts.</li> <li>• Hand over to tables</li> </ul>	<p>Set scene, housekeeping, allow participants to introduce themselves.</p> <p><i>Plenary Slides 1-3</i></p> <p><i>Participant Books pg 2</i></p>

	<p><i>IN TABLES (9 participants per table)</i></p> <ul style="list-style-type: none"> <li>• Introduce self and note-taker and any experts/observers at the table (participants free to ask questions at any time)</li> <li>• All views valid; please speak up and respond/agree/disagree to other points of view; try not to talk over one another; may need to interrupt to move discussion on.</li> <li>• Confidential with no direct attribution.</li> <li>• MRS Code of Conduct Permission to record</li> <li>• Housekeeping (phones on silent, location of toilets, any scheduled fire alarms, fire exits)</li> <li>• Use page 2 of participant booklet to explain the concept of data</li> <li>• Introductions (split into pairs for 5 mins, then introduce partner to rest of <b>table - First name, where live, who with, what do with your days, last time you gave someone some data about yourself</b>)</li> </ul>	
1020-1040	<p><b><u>Data: Brainstorming and Grouping Exercise</u></b></p> <p>AT TABLES:</p> <p><i>Ask participants to brainstorm as many different types of data that they can think of.</i></p> <p><i>FLIPCHART ANSWERS for this and the questions below</i></p> <p><b>What kind of people or organisations collect data?</b></p> <p>PROBE</p> <ul style="list-style-type: none"> <li>– Private companies (some of whom work for the government, some of whom don't)</li> <li>– The government/government bodies like the DVLA/companies delivering government services</li> <li>– Charities/social enterprises/third sector organisations</li> </ul> <p><b>How do companies, charities or the government collect data:</b></p> <ul style="list-style-type: none"> <li>- Surveys/polls</li> <li>- Focus groups/interview</li> <li>- Application forms</li> </ul>	<p>Warm-up exercise – help participants to better understand what data is, and to start thinking about it in terms of:</p> <ul style="list-style-type: none"> <li>-How it is collected</li> <li>-By whom</li> <li>-For what reasons</li> <li>-How is it used</li> </ul> <p>Allow participants to talk broadly about data and how it is used.</p>

	<ul style="list-style-type: none"> <li>- Through monitoring the use of services/products</li> </ul> <p><b>How do you give information to companies, charities or the government?</b></p> <p>PROBE FOR PASSIVE VS ACTIVE GIVING:</p> <ul style="list-style-type: none"> <li>– Direct to them</li> <li>– By giving it to someone else</li> <li>– By buying or consuming products</li> <li>– Through online/mobile activity</li> </ul> <p><b>How is this data used by companies? Charities? The government?</b></p> <p>PROBE</p> <ul style="list-style-type: none"> <li>– In business decisions</li> <li>– In government planning</li> <li>– To make sense of the world/understand it better</li> <li>– To advertise products or services</li> </ul>	
1040-1105	<p><b>Administrative Data</b></p> <p>AT TABLES:</p> <p><i>Use information on pg 3 of the participant workbook to explain administrative data. Check comprehension and ask participants to shout out examples of administrative data. Work through examples on page 5 of the participant workbook if this proves difficult.</i></p> <p>Work through the list of different types of administrative data below (lead facilitator work from the top of the list, other facilitator to work from the bottom of the list), and ask the following questions. Try to cover at least two types of data:</p>	<p>Understand levels of knowledge and top of mind views about what administrative data is, how it is collected, how it is used and stored, in particular how and whether people see government and private use of data differently</p> <p><i>Participant Books pg 3-5</i></p>

	<p><b>How is this data collected from people?</b></p> <p><b>Why is it collected?</b></p> <p><b>How personal is this data?</b></p> <p><b>How is this data currently used?</b></p> <p><b>How could this data be better used to plan services or understand society?</b></p> <ul style="list-style-type: none"> <li>• Tax records – the data that the government collects - through your employer or the tax return you fill out - about how much you work and earn, to work out how much tax you need to pay.</li> <li>• School progress records – the data the government collects about how well pupils are doing in schools, for example their results in their Key Stage tests or GCSEs.</li> <li>• Social security records - the data the government collects about those who claim government benefits, such as Job Seekers Allowance, child benefit, or the state pension.</li> <li>• Electoral registration records – the information the government holds about who lives where, in order to allow people to vote in the right constituency.</li> <li>• Vital events records (births, marriages, deaths) – the information the government holds about births, deaths and marriages that have been registered.</li> </ul> <p><i>Overall questions to ask once you have worked through a few examples:</i></p> <p><b>How ‘personal’ is administrative data?</b> What does that mean for how it should be used to plan services and understand society?</p> <p><b>How well or badly do you think administrative data is currently used?</b> What makes you say that? Can you think of any examples? How do you think this differs from how information about you is used by private companies?</p>	
--	--	--



	<p>IF APPROPRIATE:</p> <p><b>Do you think the government collects the right amount of administrative data?</b> What type of data or information should it collect more or less of?</p> <p><b>Do you think that the government and researchers should be using administrative data differently?</b> Why? Why not?</p> <p><b>How would you like to see it better used in the future?</b> How would that affect society/how public services are run?</p>	
1105-1125	<p><b><u>Presentation on Data Linkage and ADRCs and Q&amp;A</u></b></p> <p>IN PLENARY:</p> <p><i>Lead facilitator to outline ADRC plans and allow for question and answers to the facilitator and experts present.</i></p> <p><i>Key point that participants need to understand at this stage is that, for the rest of the day, we will be talking about using data for STATISTICAL and RESEARCH purposes i.e. how we understand society and how government services are planned at an overall level. We are not talking about linking up data for OPERATIONAL purposes related to the individual and their interactions with government services.</i></p>	<p>Introduce participants to the main concepts they need over the two days of workshops, and to the outline of the plans for ADRCs</p> <p>Q&amp;A allows for any necessary myth-busting</p> <p><i>Plenary Slide 5-10</i></p>
1125 – 1140	<b>MORNING COFFEE BREAK</b>	
1140 – 1315	<p><b><u>Discussion on Data Linkage Using Case Studies</u></b></p> <p><i>Check participants' comprehension of the pre-break presentation, in particular the difference between operational and statistical/research use of data – we are only talking about the latter for the rest of the day.</i></p> <p>IN TABLES (15 mins)</p>	<p>Help participants understand why researchers might want to link up data and the uses that the linked data might be put to. This will allow a more informed discussion of what is acceptable, which will be continued on Day</p>

1155	<p><b>Before the break, we explained the concept of linking administrative data. Overall, do you think that this type of data linking should be done? Why? Why not?</b></p> <p><b>Having heard about the plans for the Admin Data Research Network and Centres, what is your first reaction? Good idea? Bad idea? Why?</b></p> <p><b>What do you think will be the impact of having these centres? Can you think of any benefits? What about drawbacks?</b></p> <p><b>At this stage, what else do you think you need to know about this topic to fully understand it?</b></p> <p><i>IN PLENARY (30 MINS)</i></p> <p><i>Split participants in half into groups of 4/5. Tell them that we are now going to consider some case studies about data linking, to understand a bit more about why researchers want to link up administrative data.</i></p> <p><i>4 'stations' centred around the room. Give each group one to start with, then they will work around clockwise, spending 5 minutes at each. The case studies will be printed out on A2 laminated card. One card describe the case study, while the 'aims' and 'uses' cards have more detail on the aims and outcomes of the example.</i></p> <p><i>First participants read the description of the case study and add post its with any spontaneous thoughts on what strikes them about the case study, and how they feel about it. Once they have done that, they move on to read more about the aims and uses and add post its with any questions, comments or concerns.</i></p> <p><i>One facilitator or notetaker stands at each station and can read out the information to facilitate engagement, encourage participants to add post-its and take notes of conversations. DO NOT ASK DIRECT QUESTIONS AT THIS STAGE.</i></p>	<p>2.</p> <p><i>Case Studies A3 Slides – each set of three slides set up in a separate area around the room</i></p> <p><i>4 printouts of the case study slides for reference</i></p> <p><i>Post its and pens for each participant</i></p>
------	--	---

1225	<p>IN TABLES</p> <p><i>Distribute a set of case study printouts per 2/3 participants, so that they can refer back to them if needed</i></p> <p><b>Having read through all the case studies, what is your overall reaction?</b> Why? Were you surprised by the case studies in any way? Had you heard about this type of research before?</p> <p><b>Was there anything confusing in any of the case studies?</b> How could they have been better explained?</p> <p><i>Explore in detail which words/phrases/concepts that participants found difficult to understand and why.</i></p> <p><b>What did you think about the reasons for the data linkage across the case studies?</b> Were all the reasons for linking the data together acceptable? Why? Why not?</p> <p>PROBES:</p> <ul style="list-style-type: none"> <li>– Clarity of the end use of the linked data</li> <li>– Purpose of the end use of the linked data</li> <li>– Effect on the individual</li> <li>– Effect on society/how it is governed</li> </ul> <p><i>Facilitator reminder of the reasons in each case:</i></p> <ul style="list-style-type: none"> <li>• <i>Crime + Benefits – Understand the long term costs of reoffending using existing datasets rather than surveys</i></li> <li>• <i>IMD – Create an easily updated measure of multi-dimensional disadvantage at a small area level</i></li> <li>• <i>NPD – Understand how different issues effect educational achievement, set targets for schools</i></li> <li>• <i>Scottish life events – Academic interest</i></li> </ul> <p><b>Did you have any concerns about data linking as used in the case studies?</b> What were they? Why does</p>	
------	---	--

	<p>this concern you? What would have to change for you to stop being concerned about that issue?</p> <p>PROBES:</p> <ul style="list-style-type: none"> <li>– Personal privacy</li> <li>– Data security</li> <li>– Unauthorised access</li> <li>– Implications of the uses of the data</li> </ul> <p><i>Facilitator reminder of the uses in each case:</i></p> <ul style="list-style-type: none"> <li>• <i>Crime + Benefits – Policy implementation planning, policy evaluation, targeting spending</i></li> <li>• <i>IMD – Resource allocation for many policies (e.g. Regeneration, Sure Start), allocate charitable funds</i></li> <li>• <i>NPD – School self-evaluation, contextual measurement of school progress, school funding, other research on health and employment</i></li> <li>• <i>Scottish life events – Understanding social mobility, understanding risks associated with certain jobs and how these have changed over time, calculate genetic bases of some illnesses</i></li> </ul> <p><i>Look out for following reasons and PROBE fully if they come up, but not otherwise:</i></p> <ul style="list-style-type: none"> <li>- <i>Objections to targetted spending (on equality grounds)</i></li> <li>- <i>Objections to population 'profiling' on grounds it might lead to stigma etc.</i></li> <li>- <i>Objections to genetic profiling</i></li> <li>- <i>Worry that researchers might draw incorrect conclusions from the data</i></li> </ul> <p><b>Should government/academic researchers be able to link administrative data for any reason they want? Why? Why not?</b></p>	
--	---	--

	<p><b>Are there areas of government where this kind of data linking is not appropriate? Why do you say that?</b></p> <p><b>Do you think that the public needs to know about the government or researchers linking up data in these ways? Why/why not? If so, what do they need to know and how should they be told?</b></p> <p><b>What rules do you think should be in place in order for researchers to carry out this kind of data linking? FLIPCHART ANSWERS</b> <i>Spontaneous at this stage – explored in a lot more depth in the second session</i></p> <p>PROBES:</p> <ul style="list-style-type: none"> <li>– Who</li> <li>– What</li> <li>– When</li> <li>– Where</li> <li>– Why</li> </ul> <p>IN PLENARY</p> <p><i>Facilitators (or participant, if comfortable) to feed back the main points coming out of the discussion on their table to the rest of the room</i></p>	
<b>1315 – 1345</b>	<b>LUNCH</b>	
<b>1345 – 1430</b>	<p><b><u>Ask a researcher/ADRC rep</u></b></p> <p>AT TABLES</p> <p><u>Table 1</u></p> <p>A social science researcher will, for 5/10 minutes, present a piece of work where data linking has been or would be useful. Participants can then ask questions for 10 minutes – either about that researcher’s work specifically, or about the general implications of data linking.</p>	<p>Further information for participants</p> <p>Understanding of what participants are most interested in questioning spontaneously when given the chance to ask the experts.</p>

	<p><i>Try to keep the participants focussed on the data aspect of the researchers' work – why do they need administrative data for their work? Could they use something else instead? How did they gain access to the data they use? How did that process affect their work? What would happen if they didn't have access to it? How will the ADRCs help? What would happen to research in their field more widely? What areas of research might easier admin data linkage open up?</i></p> <p><u>Table 2</u></p> <p>A rep from one of the ADRCs will for 5/10 minutes, outline why their organisation bid to become and ADRC and how they plan to put safe places and training in place. Participants can then ask questions for 10 minutes – either about what the rep has said specifically or about data linking and the ADRC plans more generally.</p> <p>After 20 minutes the researcher and ADRC rep to swap tables</p> <p><i>Facilitators to ask probing questions where useful.</i></p> <p>IN PLENARY (if time)</p> <p><i>Facilitators (or participants, if comfortable) to feed back the main points coming out of the discussion on their table to the rest of the room</i></p>	<p>Allow researchers who may be using the ADRCs, and ADRCs reps to understand public views on their practice and aims.</p>
1430 – 1445	<b>AFTERNOON COFFEE BREAK</b>	
1445 – 1545	<p><b><u>The process behind creating and using administrative data – Data journey exercise</u></b></p> <p>AT TABLES</p> <p><i>Let participants know that we are going to spend an hour talking through and physically demonstrating the data journey process. Encourage participants to comment and ask questions at any step or sub-step throughout journey. The experts present may be able to help with technical questions.</i></p>	<p>This session is primarily educative, and aims at giving the participants a better understanding of how the data will be linked in practice and data anonymity (a poorly understood area).</p>

	<p><i>Explain to participants the purpose of the exercise, and stress that this is just an example – this is the same process that could happen with any type of administrative data, for example school records, tax credit records, data about peoples travel habits etc.</i></p> <p><i>Talk through Slide 6, which outlines each of the people involved in the journey. Assign the roles but it is not necessary to make each person read their role if you think the participants would struggle:</i></p> <ul style="list-style-type: none"> <li>- Adam</li> <li>- DWP</li> <li>- HMRC</li> <li>- Ruth</li> <li>- ADS</li> <li>- ADRC</li> </ul> <p><u>After Step 1 (Slide 7)</u></p> <p><b>Can you think of other situations in which you give information about yourself or household to the government?</b></p> <p><i>Get participants to call out as many as possible and FLIPCHART</i></p> <p><b>Who do you think it belongs to?</b> What should the ‘data owners’ be allowed to do with it? Should they ever have to ask permission of the people who gave them the data in order to use it again?</p> <p><b>Do you ever think about what’s going to happen to that information?</b> Does it matter to you? Where do you think it is stored? Do you think it is secure?</p> <p><b>Would you be happy for all of these different types of information you give to the government to be</b></p>	<p>Help start to isolate which parts of this process are essential for the public to</p> <ul style="list-style-type: none"> <li>a) Know about</li> <li>b) Be involved in helping to set the rules for</li> <li>c) Be involved in regulating</li> </ul> <p><i>Participant handout slide 6-9</i></p> <p><i>JSA forms (4 per table)</i></p> <p><i>Secure ‘boxes’/envelopes</i></p> <p><i>Cards representing data of Adam and a few other people</i></p>
--	---	--

	<p><b>linked? Why? Why not? In what circumstance?</b></p> <p><u>After Step 2 – Researcher and project approval (Slide 8)</u></p> <p><b>What researchers do you think should be allowed to ask the Administrative Data Service to allow them to link this kind of data?</b> Are there any sorts you think should <u>not</u> be allowed?</p> <p>PROBE</p> <ul style="list-style-type: none"> <li>- Researchers working for government departments</li> <li>- Researchers working for academic institutions</li> <li>- Researchers working for commercial research organisations (working for government/ commercial ends)</li> <li>- Researchers working for private businesses</li> <li>- Private individuals who want to carry out research</li> </ul> <p><b>What do you think about Ruth’s researcher’s reason for wanting to link the data?</b> Why do you say that?</p> <p>What kind of reasons or aims should researchers need to have in order to get permission to link administrative data? Is it okay to do academic research with linked data or does it have to be used to improve services?</p> <p>PROBE</p> <ul style="list-style-type: none"> <li>- Curiosity</li> <li>- To test a theory</li> <li>- To plan a government policy</li> <li>- To work out how to allocate government spending</li> <li>- To work out how well a government programme is working</li> </ul>	
--	---	--



1525	<p><b>Are there particular areas of government policy/types of government services where you think this type of data linkage should always be allowed?</b> Any areas where it should never be?</p> <p><b>What do you think of the steps that must be gone through before the researcher gains permission for their data linking project?</b> Do they seem like the right/wrong ones? Is anything missing?</p> <p><u>After Step 3 – Data linking (Slide 10)</u></p> <p><b>What do you think about the process for actually linking the data together?</b> Does it seem secure or not secure? Should anything be added to the process?</p> <p><b>Would your views change if the first three letters of your postcode remained in the dataset?</b> What about your date of birth? What about both?</p> <p><u>After all the steps have been talked through</u></p> <p><b>Having looked at the whole process, have your views on administrative data linkage changed at all?</b> Why? Why not?</p> <p><b>Did anything in particular concern or reassure you?</b> What was that?</p> <p><b>Do you think the process is the right one?</b> Should anything be changed? What?</p> <p><b>Do you trust this process?</b> If not, what would anything have to change to make you trust it?</p> <p><b>How much, if anything, do you think the general public need to know about this process?</b> How far, if at all, should they be involved in ensuring that it all happens as described here?</p> <p>IN PLENARY (if time)</p> <p><i>Facilitators (or participants, if comfortable) to feed back the main points coming out of the discussion on their table to the rest of the room</i></p>	
------	--	--

1545 – 1600	<p><b><u>Wind Down and Evaluation</u></b></p> <p>IN PLENARY (Slide 11)</p> <p><i>Thank respondents for taking part today, and remind them how their contributions will be used (to help ESRC understand public views, and the new ADRCs to plan how best to involve the public in their work from now on) – allow questions.</i></p> <p><i>Explain that next session we will concentrate on Census in the morning and come back to this topic in the afternoon – remind them of the date and that recruiters will call to confirm attendance.</i></p> <p><i>Explain homework task</i></p> <p><i>Give out end of event questionnaire and allow participants 10 minutes to fill it in <u>before incentives</u>.</i></p> <p><i>Give out incentives and sign form</i></p> <p><i>THANK AND CLOSE</i></p>	<p>Allow time for participants to evaluate the day.</p> <p>Homework task will prepare participants for discussing the census in greater depth in Day 2, and also help us to understand the types of arguments around the changes that resonate.</p> <p><i>Plenary Slide 11</i></p> <p><i>Homework task</i></p> <p><i>Evaluation forms</i></p>
-------------	---	---

## Appendix 2: Workshop 1 Plenary Presentation


# WELCOME!

ABZIFESAVABGIGATFVA  
AIINIDQPCAIAINFQPCA  
QNJOKQEAQNJOKQXARQ  
DATA LUGIAKEYPLBGNAI  
PPBLMHOVDPPBLMSOVUP  
NLIDE-IDENTIFIEDDHN  
EANYTOBAJUNNYTCBSJE  
WAOGSEEBOHATGSUEZOW  
HQPIALYEEJQPIARYIEH  
CONFIDENTIALKFEEUEQ  
ZHWLXYAEREPWLXGAPRI

Version 1 | Internal Use Only

## Public Dialogue on Data – Workshop 1

Ipsos MORI  
Social Research Institute



### Why are we here today?

- Today we are talking about how information or data about the public that is collected by the government ('administrative data') is linked together to find out more about society, and plan government services.
- We're here today on behalf of two organisations who are really interested in what you have to say:
  - **The Economic and Social Research Council (ESRC)**
    - Funds research on economic and social issues, and training for researchers
    - Funded by but independent of government
  - **The Office for National Statistics (ONS)**
    - Produce official government statistics
    - Carry out the census



## What will happen today?

- We want your views on two big questions:
  - How should social researchers link up Government data and use it to understand society and possibly inform public services?
  - What rules or principles should be followed when linking up people's data?
- Don't worry! You don't need to know anything about research or how it's used!
- We will be doing lots of different things to help you along.



## Agenda for the rest of the day

**Now:** What is data, administrative data, and data linking?

**BREAK (11.25-11.40)**

**Before Lunch:** What does this mean for my data?  
Looking at real examples of where data linking is used and discussing how you feel about them

**LUNCH (13.15-13.45)**

**After Lunch:** Ask a researcher  
Hearing from some researchers who want to link data and asking them questions

**BREAK (14.30-14.45)**

**Last Session:** Data Linkage Journey  
Learning about the journey that data goes on from members of the public to a linked dataset

**Final 15 minutes:** Feedback and take-home exercise for next week

## What is social research and how is it done?

Social research involves finding things out about society, understanding how it works and explaining why.

- It can help us understand topics like:
  - causes of unemployment/poor educational attainment
  - what helps economic growth
  - how and why people vote
  - what makes people healthy and happy
- Social research uses data/information about large numbers of people to find out the answer to these types of questions at a societal level.
- Researchers use lots of different tools to gather data, including surveys, experiments, personal interviews and group discussions

## Administrative Data: A new way forward for social research

**Administrative Data** is data collected by government departments and other government organisations to help them deliver a service to you – for example the number of children you have to work out what child benefit you are entitled to.



- Examples include data from **social security, tax and educational records**
- The government now collects more data than ever. How?
  - When people fill in forms to access services
  - When businesses give information to government, for example about their workforce or pay
  - Through measuring use of services e.g. number of people using roads

## What is 'data linkage' and how is it done?

- Administrative data can be particularly useful to researchers when **linked**
- **Data linkage is when we connect data from two or more different sources or datasets.**
- This is done using **de-identified administrative data**: data from which personal identifying information (name, address, NI number) has been removed.
- **A new dataset** is then created that **contains more information** about the same individual, family, event or place.
- This linkage is done on very big datasets – researchers are not interested in what the linked data tells them about individuals, but in what the whole dataset tells them about society, or specific groups



## Why and how is administrative data used for research?

Administrative data is useful for research because:

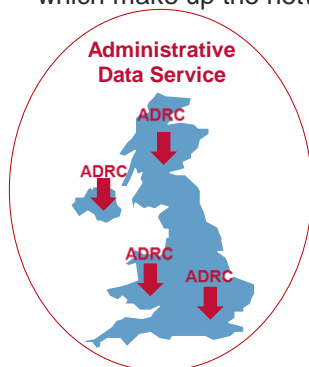
- ✓ It is often good quality and consistent
- ✓ It covers everyone who uses a service
- ✓ It is cost-efficient as it is collected as part of a service

However, it is not currently used or linked as much as it could be because

- ✗ Administrative data is not specifically for the purpose of research, so it is not always collected in useful formats
- ✗ It can be difficult, time consuming and sometimes impossible for researchers to access the datasets
- ✗ Researchers don't always have the skills or resources to link the data properly
- ✗ It is not clear what the public think about linking this type of data

## The future of linking administrative data - the Administrative Data Research Network

- A new **Administrative Data Research Network** has just been set up to help social researchers to get access to and work with linked administrative data.
- It consists of an overarching **Administrative Data Service**, and then 4 **Administrative Data Research Centres** (one in each nation of the UK), all of which make up the network.



The network will:

- Set up safe places to access the data and procedures for researchers to work with de-identified administrative data
- Train scientists to work correctly, securely, ethically with this data
- Work with government to access data and influence policy
- Work with the public to inform and involve them
- ADRC staff may also undertake research on the linked data

## When and how all this will happen?

**Now**



- **October 2013** – announce who will run each ADRC and the ADS
- **January 2014** – set up the procedures for researchers to work with and access the data; and agree how the public will be part of this
- **April 2014** – launch the new network and start the first research projects, start the programme of training
- **October 2014** – expand to create more places where researchers can access the data (researchers don't then need to travel to the centres)
- **2014-2018** – lots of data linking projects!
- **October 2015** – Review the centres and Data Service

**2018**

- **September 2018** – end of the first five year grants to ADRCs and ADS

**Thanks and see you soon!**

- **Thank you so much** for all your contributions today!
- **We look forward to seeing you in two weeks** – same time, same place
- **Please fill in your evaluation forms.** Be honest! All feedback – good and bad – is useful
- **Remember to take your homework task with you** – it shouldn't take long at all, and will help you get ready for our discussion of the census next week

**ENJOY THE REST OF THE WEEKEND!**



# Participant Information and Workbook

## Public Dialogue on Data – Workshop 1

Ipsos MORI  
Social Research Institute

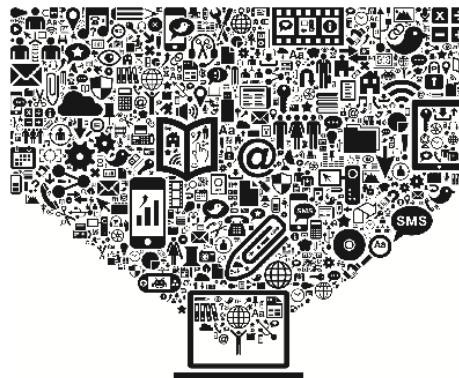


### What is Data?

**Data** is a collection of facts. It can be numbers, words, observations or even just descriptions of things. Nowadays a great deal of data is stored on a computer or electronically.

Examples of data might include:

- Energy used by households
- Records of items bought in shops
- The number of people at an exhibition
- Notes from an interview
- The results of a survey
- Electoral registration records
- A tweet



## What is Administrative Data?

**Administrative Data** is data collected by government departments and other government organisations to help them deliver a service to you – for example the number of children you have to work out how much child benefit you are entitled to.

It is therefore data that has been collected for administrative purposes, and not necessarily for research purposes.

**SHOUT OUT AS MANY EXAMPLES OF ADMINISTRATIVE DATA AS YOU CAN THINK OF**

## Examples of administrative data

- Examples of administrative data include:
  - Social security records
  - Tax records
  - Educational records
  - Medical records
  - Vital events records (births, deaths, marriages)
  - Electoral registration
  - Crime records
  - Local authority records (use of services e.g. parks, social care)



# Data Linkage Journey

## Public Dialogue on Data – Workshop 1

Ipsos MORI  
Social Research Institute



### Data linkage journey - who is involved?



- **Members of the public**– people like you, who give data to government departments when using government services  
*Adam – Claimed JSA in 2011, got a job and paid tax in 2012*



- **Data owners**– the government departments or institutions that then 'own' and store the data  
*DWP – owns benefits data      HMRC - owns tax data*



- **Researchers** –people that analyse the data from linked administrative datasets e.g. academics or people who work for the government  
*Ruth – a PhD student who wants to understand the link between long-term unemployment and the amount people subsequently earn*



- **Administrative Data Service** – The overarching service that controls the data linkage process in all four countries



- **Administrative Data Research Centres** – One in each four countries that enables secure data linkage and supports researchers



## Data linkage journey - Step 1 – Data giving

**2011**

Adam – data giver



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



DWP Department for Work and Pensions

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



HM Revenue & Customs

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



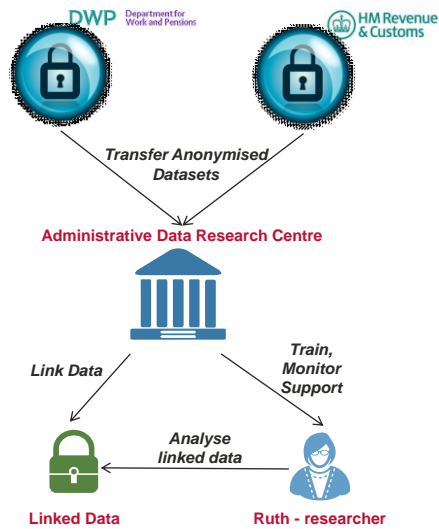
Project Approval:

- 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)



- ☐ 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

## Appendix 4 – Workshop 1 Case Studies

### Offending, Employment and Benefits project

Case study  
1



- Links administrative data between three government departments: Ministry of Justice (MoJ), the Department for Work and Pensions (DWP) and Her Majesty's Revenue and Customs (HMRC)
- Linking criminal records to employment and benefit information allows the government to see what happens to offenders once they leave prison
- The matched dataset holds 40m pieces of data on 3.6m offenders from 2000 to 2010

Ipsos MORI  
Social Research Institute  
© Ipsos MORI Version 1 | Internal Use Only



### Aims of Offending, Employment and Benefits project

Case study  
1

#### Why was this data linked?

- **Reoffending is expensive for society and the Government** – this research is meant to help understand the long term job prospects for offenders and re-offenders.
- Before this project, the only information on the links between employment, benefits and offending was from surveys, which are expensive and do not provide accurate information on all offenders.



Ipsos MORI  
Social Research Institute  
© Ipsos MORI Version 1 | Internal Use Only

Source: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/217428/offending-employment-benefits-emerging-findings-1111.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/217428/offending-employment-benefits-emerging-findings-1111.pdf)



## Uses of Offending, Employment and Benefits project

Case study  
1

### How has the linked data been used?

- The data has been used to create previously unknown statistics, such as the ones below:



- Analysis of this data is helping government departments to:
  - understand** how successful Government policies are
  - target spending** to where there is most need or where it will be most useful
  - develop** policy to help offenders and former prisoners into work (for example, a policy to get ex-prisoners claiming Job Seekers Allowance onto the Work Programme)

## Understanding of disadvantage at a small area level (Index of Multiple Deprivation)

Case study  
2

- The University of Oxford has produced and updated the “**Index of Multiple Deprivation**” for the last ten years.

- The Index links data from most government departments, including:

- Health records from the NHS
- Unemployment and tax data from HMRC
- Benefits information from DWP



*In 2010 the most deprived area in England was in Tendring District, Essex*

- By linking together a lot of government-held population data, the index provides a measure of “deprivation” for very small areas which includes unemployment, life expectancy, crime, housing, income, and the quality of education, as contributing factors

## Aims of the Index of Multiple Deprivation

Case study  
2

### Why was this data linked?

- It is very difficult to measure how disadvantaged an area is because:



- “Deprivation” means a lot of things (what people earn, their health, the type of housing they live in, the quality of education available).
- The level of deprivation in an area can change from year to year as people move in and out of an area.
- Before this index, local and national government used Census data to measure disadvantage. There are problems with this because Census data is not very detailed, is only collected every ten years, and isn’t available at a very small area level. By contrast, the index provides a very sophisticated way of measuring deprivation

## Uses of the Index of Multiple Deprivation

Case study  
2

### How has the linked data been used?

- The Index **helps Local Council and national Government to allocate resources and funds** to the most disadvantaged areas e.g.



- neighbourhood regeneration programmes
- health services
- the Sure Start educational programme
- transport planning
- emergency services
- sports and arts facilities for children



- It is also used to **allocate charitable funds** such as National Lottery funding.



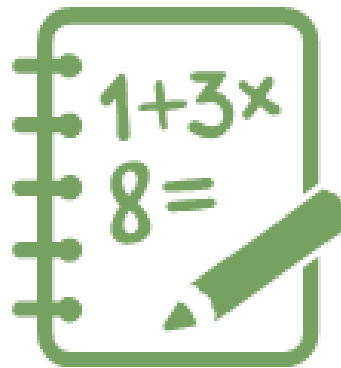
- The **National Pupil Database (NPD)** is a dataset owned by the Department of Education, which contains detailed information on every child in school in England



- It combines all kinds of data on **pupils from different sources** – for instance examination results, attendance records, the name of their school – **into a single database**
- These datasets are linked using a unique pupil identifier (a type of reference number), rather than the pupils' name or personal details

### Why was this data linked?

- The National Pupil Database was created to help researchers and others who work in the education sector to better understand **how different issues affect pupils' educational achievement**.
- It was also developed to help the Government to **set targets for schools** and measure how well they meet them.



## Uses of the National Pupil Database

Case study  
3

### How has the linked data been used?

- To help schools **identify particular groups of pupils** who are underperforming.
- To produce public statistics - for example **how well a school is doing in relation to the number of its children who are eligible for Free School Meals**
- To decide **how much money from the national education budget is to be given to particular local authorities and schools**
- **To link it with external survey datasets**, which can help with further research on other issues, for example health or employment.

## Linking life events records in Scotland

Case study  
4

- The 'Digitising Scotland' project aims to **computerise all birth, marriage and death certificates in Scotland from 1850 to 1974.**
- Birth, death and marriage documents hold important information on the people involved
  - Birth and marriage certificates have the occupation and age of parents/newlyweds
  - Death certificates detail a person's age at death and the cause of death
- Digitising and linking these datasets will enable researchers to **produce anonymised family trees for everyone in Scotland between 1850 and 1974**

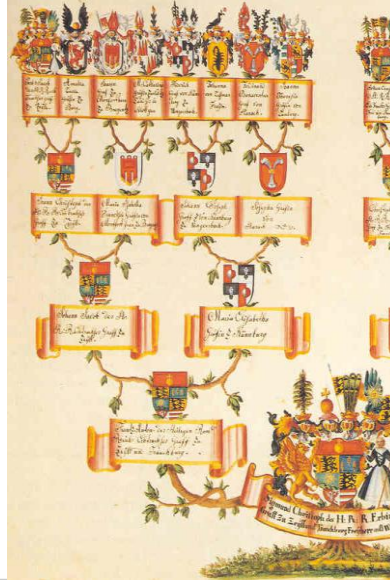


## Aims of linking vital events records in Scotland

Case study  
4

### Why is this data being linked?

- Currently this data has been linked without a definite aim of improving a specific public service.
- However, the linked data may be useful for:
  - Genetic and public health research
  - Research into patterns of diseases and causes of death.
- It could also be useful for studies of the family and social mobility, and environmental and historical research.



## Uses of linking vital events records in Scotland

Case study  
4

### Potential uses of this linked dataset, yet to be confirmed:



- It could first be used by researchers to answer important medical and historical questions, for example:
- Comparing the similarity of causes of death in related individuals will allow researchers to calculate the genetic basis of illnesses
- Looking at age and cause of death by occupation helps to understand risks associated with certain types of jobs and how these changed over 150 years
- By looking at change in occupational social class – it will be possible to better understand changes in social mobility over time and in different parts of Scotland.

## Appendix 5 – Workshop 2 Discussion Guide

### Discussion Guide for Workshop 2 of Public Dialogue on Data

#### FINAL Post Pilot 231013

This guide outlines the discussion that will take place between members of the public, facilitators and experts in the field at the second of a pair of workshops. There are 7 workshops taking place around the country over the course of October and November 2012. The dialogue has been commissioned by the Economic and Social Research Council (ESRC) and the Office for National Statistics (ONS), with the following objectives:

#### Project objectives

1. **To better understand people's views on the linking of administrative data**, and work out how different messages about the process by which the data is linked affects attitudes.
  - Specifically, to explore how the following aspects of the linkage process affect views:
    - Why - The end use of data (administrative/statistical/operational) and people's ability to understand the difference between different potential uses; how do people understand 'benefit' and what kind of trade-offs that may exist between data protection and privacy and the value of data analysis.
    - 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 (matching records from public and private sector providers).
    - Who – The identity/qualifications/sector/motivations of the people carrying out the linking
    - Where – The impact of the environment in which the linkage is to take place
    - 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
    - How – is data-linking acceptable in the case of a potential 2021 administrative data Census?
2. **To begin the process of creating a language around the re-use of administrative data and data linking that is understandable to the general public**
  - What sort of assurances are required, and what sort of language is necessary, to help the public to understand data linking. Are there any 'red flags', or key words we need to think about when talking about this area?
  - What concepts need to be conveyed to get across the key principles, dilemmas and arguments in this area?
  - What do those who work in this area need to do to earn the public's trust?

3. To help inform the development of the governance and operational procedures that ADRCs will adopt, and provide data on public attitudes for ADRCs to help inform their future strategies and priorities for public engagement
  - How much do people want/need to know in order to be properly engaged in the process? How should any arrangements be future-proofed?
  - Are there regional or demographic differences in the level of interest in, opposition to, and engagement with data linking?
  - Who needs to be prioritised for engagement, why and how?

### Using this guide

We use several conventions to explain to you how this guide will be used. These are described below:

Timings	Questions and Prompts	Objectives
5 mins	<p><b><u>Underlined = Title:</u></b> This provides a heading for a sub-section</p> <p>Plain text = Information This is information for the facilitator to give to the participants (in facilitators on words)</p> <p><i>Italic or CAPS = Instructions.</i> This tells the facilitators how to organise activities or record participant responses</p> <p><b>Bold = Question or read out statement:</b> Questions that will be asked to the group if relevant. Not all questions are asked during the workshop based on the facilitators view of progress in answering the research questions</p> <p>- Bullet = prompt: Prompts are not questions – they are there to provide guidance to the facilitator if required.</p>	<p>This area is used to summarise the objectives or each section - for information only.</p> <p><i>It also shows what material (e.g. slides, handouts) is needed for each section</i></p>
How long it takes	Typically, the researcher will ask <b>questions</b> and use the prompts to guide where necessary. NB: Not all questions or prompts will necessarily be used in an interview	

Time	Process	Objectives and materials needed
1000-1005	<p><b>Welcome and Introduction</b></p> <p>IN PLENARY:</p> <p><i>NB this is not a script but points to be talked through</i></p> <ul style="list-style-type: none"> <li>• Welcome from Ipsos MORI – (re)introduce team</li> <li>• Objectives for the day and how findings will be used.</li> <li>• Two weeks ago we presented you with a lot of information about data and data linkage, and we had a broad discussion of your views on this topic</li> <li>• We also asked you to do a short homework task about the Census and bring it with you</li> <li>• Today we'll discuss your homework task, and build on what we presented last time to talk about the future of data collection and linkage, and what principles you think are important</li> <li>• As with the first workshop there will be a mix of plenary, group discussion and exercises; Introduce observers /experts.</li> <li>• All views valid; please speak up and respond/agree/disagree to other points of view; try not to talk over one another; may need to interrupt to move discussion on.</li> <li>• Confidential with no direct attribution.</li> <li>• MRS Code of Conduct.</li> <li>• Housekeeping (turn phones off, any scheduled fire alarms, fire exits).</li> <li>• Ask everyone to reintroduce themselves and say what they found most interesting from the last session</li> <li>• Work through agenda for the day</li> </ul>	<p>Set scene, housekeeping, allow participants to re-introduce themselves.</p> <p><i>Participant Book Slide 2</i></p>

<p><b>10-05-1020</b></p>	<p><b><u>Recap of Day 1 and Admin Data Linkage Process</u></b></p> <p>IN PLENARY</p> <p><i>Work through Plenary Slides 3-4 that recap Day 1, to make sure all participants are clear on the</i></p> <ul style="list-style-type: none"> <li>• <i>What Data linking is (remind them of de-identification)</i></li> <li>• <i>Why it is carried out – the point of social research (remind them of the case studies and the experts they heard from)</i></li> <li>• <i>The ADRN plans</i></li> </ul> <p><i>Remind participants briefly about the data linking journey that we went through, and then let them know that we're going to watch a short video to recap the process.</i></p> <p><i>Show video of Adil from ONS explaining the data linking process and ask for participant reactions.</i></p> <p><i>Check for understanding and allow any time for questions. Remind participants that researchers who want linked data for their projects <u>will only have access to the data after it has been linked.</u></i></p> <p><b>Is there anything else you might need to know about how researchers might go about linking administrative data?</b></p> <p>SHOUT OUT</p> <p><b>How do you feel about this process?</b> Any questions? Concerns? Suggestions for changes?</p> <p><b>How much do you think the general public need to know about this process?</b> Why?</p>	<p>Recap from the previous week.</p> <p><i>Participant Book Slide 3-4</i></p>
<p><b>1020-1120</b></p>	<p><b><u>Data linkage Issues</u></b></p> <p>AT TABLES</p> <p><b>Taking everything you have heard last week and the video this morning into account, please discuss in</b></p>	<p>Allow the range of views around data linking to form part of the dialogue, to ensure that participants are taking all</p>

	<p><b>pairs any issues or questions you still have about data linkage.</b> <i>Ask participants to write on post its</i></p> <p><b>What issues or questions did you come up with?</b></p> <p><i>Stick post-its on flipchart, grouping similar ones – then ask them to rank the areas in order of level of concern (high, low or medium). If they struggle to come up with anything, run through prompts below, and get them to rank the areas in order of concern.</i></p> <ul style="list-style-type: none"> <li>• <i>Administrative data will be linked without consent, and usually without notice</i></li> <li>• <i>Privacy – are there certain types of information they think should be excluded e.g. sexual orientation</i></li> <li>• <i>Potential for huge databases (in terms of number of records)</i></li> <li>• <i>Security of databases</i></li> <li>• <i>Records may be linked across long time periods giving service user histories</i></li> <li>• <i>Research on specific groups that could be targeted differently (for better or worse)</i></li> <li>• <i>Results of the research findings could be something they don't agree with</i></li> <li>• <i>Data could be collected differently in the future, particularly passively through mobile devices – should it still be linked?</i></li> </ul> <p><i>Then work from the area of most concern downwards, trying to work out what is driving the concerns, and what it would take to allay these concerns.</i></p> <p><i>Cover as many as you can in the hour, but do not worry if that means you only cover a few – spend as long on each issue as the participants want. Further information on each of the areas above to allow for prompting and probing. Prioritise the areas your group are most concerned about. No need to cover and prompt on all.</i></p> <p><b>Lack of consent and notice – does it matter?</b></p> <p><i>It is completely impossible to ask for consent for the linking of administrative data, for financial and practical reasons, given the size of the datasets. It is also not possible for Govt Depts to know at the point</i></p>	<p>potential implications into account in the design activities after lunch.</p> <p><i>Post Its, Pens</i></p>
--	---	---



*of data collection, how the data will be used – think about when you fill in a government form – it would be impossible for them to run through all of the thousands of potential research uses of each of the pieces of data. In addition, some of the data that researchers want to link could have been collected years ago, and no one ever thought that it would be one day used for research (remember the case study about births, deaths and marriages last time). This data will be linked without consent or notice.*

Do you think that the public need to know about this? For what reasons? What do they need to know?  
How should the ADRCs let people know about this?

**Privacy**

*Even though all data will be linked, it will be possible to put together a whole range of data about service users that they never thought would be looked at when collected together. So, for example, while people would never be asked for their income while at their doctor's office or the hospital, it may be that researchers will want to link medical data with income data to understand the links between earnings and health outcomes. So while those pieces of information would never be collected together, they could be linked and analysed together.*

How would you feel about your health records being linked with details of your income, for example?

Does this matter/make a difference to your views?

**Type of databases:**

*There is potential for a lot of different types of databases, for example:*

- *Ones in which cover huge numbers of people/most of the population (e.g. the ones ONS create to estimate population statistics) – link back to Adil's video*
- *Ones which cover a huge number of variables about a each individuals characteristics/service use/outcomes (e.g. the National Pupil Database, which is very rich, containing exam and attendance records across each child's entire school career) – ALL DE-IDENTIFIED*

	<p>Does the size of a linked database matter?</p> <p><b>Security</b></p> <p><i>We talked through the security of these databases on Day 1 – essentially only accredited researchers will have access to pre-linked de-identified data for the time-span of their projects. ADRC staff may also have access in order to help the accredited researchers. Access will be possible either in a physical secure environment, or a virtual, password-protected one (i.e. via VPN, which means that the data itself remains in the secure physical environment, and is never sent on the internet).</i></p> <p>Are there still concerns about the level of security of these databases?</p> <p><b>Linking records across years</b></p> <p><i>Besides linking together datasets from different departments, researchers may link together datasets from different times, which would allow them to build up a picture of service usage and life outcomes for groups over a long period of time (it could be over the entire lifetime of a person). An example might be linking school attendance records with employment records from 30 years later, to see what effect poor attendance has on employment and income in later life. NB this kind of linking happens already e.g. the Longitudinal study links Census data from 1971 to various admin records for health research – ADRCs will just make it easier for researchers to do these kind of projects.</i></p> <p>Does this make a difference to views on linking?</p> <p><b>Potential uses that may harm specific groups/lead to cuts</b></p> <p><i>Research findings from using linked data may lead to decisions about public services that you may not agree with, or changes that are seen as controversial e.g. the bedroom tax.</i></p> <p><i>DO NOT READ OUT: This is of course true of all research, however, linking administrative data may lead to much greater understanding of how services work and how different factors affect them. For example, it could lead to much greater personalisation of services, or targeting at those most in need, which could</i></p>	
--	--	--

	<p><i>conflict with some people's concepts of fairness and equality in access to public service.</i></p> <p>If the research results are counterintuitive or against your personal values do you still support the data being linked?</p> <p><b><i>Research on specific groups</i></b></p> <p><i>Greater linking of admin data may allow for extremely detailed research into service use or outcomes for specific groups e.g. ethnic minorities, those with disabilities, offenders, those with no qualifications etc. For example, there could be a research project into whether a lack of qualifications is correlated with long-term benefit take-up.</i></p> <p>How do you feel about the possibility of knowing a lot more about very specific small groups?</p> <p><b><i>The future</i></b></p> <p><i>We've mostly been talking about data that is collected from people via forms or interviews. In the future we may be able to collect far more data, and in many different ways. One example might be collecting data from smartphones or when they log into an online service account. Or there could be greater collection of geographic location data using smartphones – for example how people move around a specific area, which could make it easier to plan and locate services.</i></p> <p>Would your view change if data were to be collected differently?</p> <p>IF TIME</p> <p><i>Facilitator (or participant, if comfortable) to feedback their tables concerns to the rest of the room.</i></p>	
1120 – 1135	<b>MORNING COFFEE BREAK</b>	
1135-1150	<b><u>ADRC plans- how have views evolved</u></b>	Understand whether and how opinions have changed, based on

	<p>IN PLENARY</p> <p><b>We’ve just talked about some potential issues people might have with data linking, while last week we spoke about the potential uses and benefits. As a reminder, the Admin Data Research Network has been set up to ensure this is all done efficiently, securely and ethically. (Refer to Slide 4 in pack)</b></p> <p><b>Based on everything we’ve discussed today and last week, please can you stand at this end of the room if you are concerned about the plans for the Admin Data Research Network, and at this end of the room if you are not concerned at all.</b></p> <p><i>WHILE STILL LINED UP (10 MINS)</i></p> <p><b>Have your views about this changed based on what we just discussed?</b> Have your views on the ADRC plans changed at all since the last workshop? Why?</p> <p><b>What specific concerns do you have about the plans for the network?</b> Why?</p> <p><i>Ask those who have concerns?</i></p> <p><b>What would have to change for you to have no concerns?</b></p> <p><i>Ask those who do not</i></p> <p><b>How would you explain these plans for the Network to other people? What would you say are the benefits?</b></p>	<p>the information presented before the break.</p> <p><i>Participant Book Slide 4</i></p>
1150 – 1250	<p><b><u>Principles for data linkage</u></b></p> <p>AT TABLES (10 MINS)</p> <p><i>Divide participants into two groups and give each an A3 pro-forma to fill in. Give each group a set of “who, what, when, where why” cards with ideas/prompts on them. Remind participants that the ADRCs have just come into being, and they have yet to work out final rules around access to linked data. The</i></p>	<p>Allow participants to reflect on everything they have learned across the two days, and use that to design the system that they would like to see</p>

	<p><i>point of this session is to let them know what you think should be the rules, so that they can consider the public's views when they are making their final plans. They will need to consider:</i></p> <ul style="list-style-type: none"> <li><i>• The kind of people who should be allowed to access the linked data (WHO)</i></li> <li><i>• The different reasons why they should be allowed to do this (WHY)</i></li> <li><i>• What kind of data can and cannot be linked? (WHAT)</i></li> <li><i>• Where access to linked data could happen i.e. the physical location (WHERE)</i></li> <li><i>• Options for how long this data might be held (WHEN)</i></li> </ul> <p><i>Talk participants through the ADRCs provisional plans for WHO WHAT WHEN WHERE WHY (on the back of the blank pro-forma) and allow them to give any spontaneous reactions/ask any questions before they start working on their own.</i></p> <p><b>GIVE 20 MINS TO COMPLETE THE TASK</b></p> <p><i>Facilitator to sit with one group, and note-taker to sit with another to answer questions, probe and take (handwritten) notes in order to capture data. Please do make sure that this exercise is done in small groups as it gives the less vocal participants a chance to contribute more.</i></p> <p><b>BRING BACK TO TABLES OF 9, PUT A FRESH A3 PROFORMA ON THE FLIPCHART FOR THE FACILITATOR TO FILL IN (20 MINS)</b></p> <p><i>NB this part may not be possible if participants take a long time to do the activity. Use your judgement as to whether it would be useful for them to debate across the whole table.</i></p> <p><b>Now I'd like you to work as a whole table, and discuss the principles you came up with in your smaller groups, and why. Try to work out a combined set of principles for your table</b></p> <p><b>How do these differ from the ADRC principles on the back of your pro-forma?</b> Does it matter? Do you think the ADRCs should follow different principles, in light of the discussions we've had throughout the day?</p>	<p>The debates within the small, bigger and whole workshop groups about what rules should be in place will give us a sense of where the 'red lines' are – which points participants are not willing to compromise on</p> <p>We will also have an output from each area representing the consensus/compromise that was reached around principles for data linkage, which can be compared across areas</p> <p>Gets fully informed feedback on the proposed rules for ADRCs</p> <p><i>Double sided pro-forma, one with ADRC principles, one without</i></p>
--	--	--

	<p>IN PLENARY (10 mins)</p> <p>ASK EACH TABLE TO PRESENT THEIR IDEAS TO THE OTHER TABLE</p> <p><b>Which principles are the most important? Why?</b></p> <p><b>What elements did you have in common? Why?</b></p> <p><b>Is there anything you disagreed on and why?</b></p>	<p><i>“Who, what, when, where why” cards with ideas/prompts on them</i></p>
1250-1325	LUNCH	
1325-1405	<p><b><u>ADRC engagement</u></b></p> <p>AT TABLES</p> <p><b>Now you know a lot more about how the ADRCs are going to work, what will people in general want to know?</b></p> <p><i>Explain to participants that the overarching administrative data service (ADS) is based at the University of Essex, and which institutions will be running the ADRC in their country:</i></p> <ul style="list-style-type: none"> <li>• <i>England: University of Southampton</i></li> <li>• <i>Scotland: University of Edinburgh</i></li> <li>• <i>Wales: Swansea University</i></li> <li>• <i>Northern Ireland: Queens University Belfast</i></li> </ul> <p><b>Do you trust the centres more now they are associated with a known university or does it not make any difference?</b></p> <p><b>If you were in charge of the ADRC in your country, how would you keep the public informed and interested in your work? What would it look like? <i>Flipchart responses</i></b></p>	<p><i>Allow participants to give their ideas for ADRC public engagement, and discuss the early ideas that the centres have put forward, with representatives from those centres.</i></p> <p><i>Slide 5 + 6</i></p>

	<ul style="list-style-type: none"> <li>• What would you tell the general public about your work? How would you make them interested? Imagine the ADRCs were going to put an ad in the paper telling people the most important things they needed to know. What would be the headline? <i>Allow participants to work in pairs to come up with some, then the table can pick the best one. COLLECT IDEAS.</i></li> <li>• How would you actively communicate what you do and how you plan to involve the public? <i>PROBE</i> <i>Websites</i> <i>Social media</i> <i>Videos</i></li> <li>• How would you demonstrate you are listening to people's views? <i>PROBE</i> <i>Meetings</i> <i>Open days</i></li> <li>• How would you like the ADRCs deal with concerns raised by any particular data linking research project?</li> </ul> <p><b>Do you think that the ADRCs should get the public more actively involved by having them be part of the process rather than just informing them about their work (such as on boards)? Why/why not? In what ways?</b> <i>Explain that getting the public involved is a key part of the ADRCs work, and that they will each employ someone full time to do this.</i></p> <ul style="list-style-type: none"> <li>• Is that something you would be interested in? What about your friends/family/peers? Why/why not?</li> <li>• Would you get involved and would it be in an active way or more passive (largely comms and website)?</li> <li>• If not interested in being involved, it is still important for ADRCs to involve people at a collective level. What ways do you think will appeal to most people you know?</li> </ul> <p><i>Depending on availability, a rep from each ADRC can present their plans for engagement, and be allowed</i></p>	
--	---	--

	<p><i>to ask questions of participants. Participant Slide 6 can also be used instead to give participants an idea of the kind of things that are being proposed – give participants 5 minutes to read through and ask them to individually write down on a post-it, in order, what they think are the three best ideas on the slide.</i></p> <p><i>COLLECT POST ITS.</i></p> <p><b>Do these seem like good or bad ways involve the public?</b> Which strike you as the best? Which would you be most likely to take part in?</p> <p>IN PLENARY (if time)</p> <p><i>Facilitators (or participants, if comfortable) to feed back the main ideas for engagement that came out of their table</i></p>	
--	---	--



# Participant Information and Workbook

Version 1 | Internal Use Only

## Public Dialogue on Data – Workshop 2

Ipsos MORI  
Social Research Institute



### Agenda for the day

**Now:** Recap of Day 1 and discussion of Data Linkage Issue

#### **BREAK (11.20-11.35)**

**Before Lunch:** ADRC Plans and Principles for linking data  
Coming up with principles for the who, what, when, where and why of linking data

#### **LUNCH (12.50-13.25)**

**After Lunch:** A changing Census  
Discussing what the census means to you and what you think of proposed changes to it

#### **BREAK (14.45-15.00)**

**Last Session:** What effect might the changes have?  
Learning about potential implications of changes to the census to help decide if they are a good or bad idea

**Final 15 minutes:** Feedback and thanks



## What are we talking about this morning?

We're here on behalf of our clients- The **Economic and Social Research Council**. Our main question this morning is:



**What rules or principles should be followed when gaining access to linked administrative data?**

**Social research** involves finding things out about society, understanding how it works and explaining why.

**Administrative Data** is data collected by government departments and other government organisations to help them deliver a service to you.

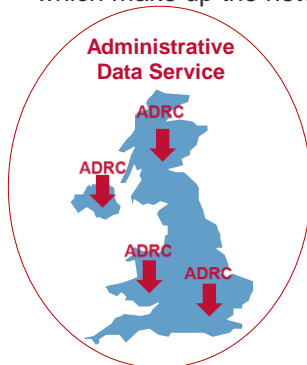
**Data linkage** is when we connect data from two or more different sources or datasets. This is done using **de-identified administrative data**: data from which personal identifying information (name, address, NI number) has been removed.

Ipsos MORI  
Social Research Institute  
© Ipsos MORI Version 1 | Internal Use Only



## The future of linking administrative data - the Administrative Data Research Network

- A new **Administrative Data Research Network** has just been set up to help social researchers to get access to and work with linked administrative data.
- It consists of an overarching **Administrative Data Service**, and then 4 **Administrative Data Research Centres** (one in each nation of the UK), all of which make up the network.



The network will:

- Set up safe places to access the data and procedures for researchers to work with de-identified administrative data
- Train scientists to work correctly, securely, ethically with this data
- Work with government to access data and influence policy
- Work with the public to inform and involve them
- ADRC staff may also undertake research on the linked data

Ipsos MORI  
Social Research Institute  
© Ipsos MORI Version 1 | Internal Use Only



## Public Engagement Plans – Example Activities

### UK WIDE ACTIVITIES

**Media stories** about the benefits of administrative data research

**Public event** debating issues of data confidentiality

**Public debating space** on ADS website

**Campaign** to promote understanding of the privacy safeguards

Work with the organisations concerned about **data linking privacy**

Report to work out the **economic benefit** to society

### ADRC ACTIVITIES IN THE NATIONS

**Learning:** interviews with experts, research on data linking examples worldwide

**Public opinion research:** Surveys, focus groups, public panel

**Awareness raising and events:** Research findings publicised and explained in laymans terms e.g. leaflets, presentations to public, government, academics;

ADRC open days and events to discuss and suggest data linking projects;

**On-going public representation:** General public reps on scientific boards

## An example of keeping research participants informed

The screenshot shows the 'Participant area' of the Understanding Society website. At the top, there is a navigation bar with links: 'About the study', 'News', 'Participants' (highlighted), 'Research, findings and impact', 'Data and documentation', and 'Contact'. Below the navigation bar is a search bar. The main heading is 'Participant area' with the subtext 'Your data – real stories'. A large orange box titled 'Participants features' contains the text: 'This area of the website is dedicated to you, the members of the Understanding Society survey. Here, you'll find the latest research to come out of the study. By now, you should have received a postcard explaining just one of the ways the data has been used. We hope you liked it. To read more, just click on the links below. We'll be sending you these kind of updates several times each year and making them available online – just in case it's not reached you for any reason. If you have any feedback, we'd love to hear from you. Please get in touch. We're excited by these latest findings and hope you and your family are too.' Below this, there are three featured articles, each with a date of '09 Sep 2013': 1. 'Poverty: time to take a fresh look' with the subtext 'Your input helped leading think tank Demos rethink traditional measures of poverty'. 2. 'GDP and beyond: developing a wider picture of well-being' with the subtext 'Information you provided about your income, health, spare time and family helps ONS produce a richer picture on "how the UK is doing"'. 3. 'Do you rely on grandparents for childcare?' with the subtext 'One in four of you do – and that's of big interest to policy makers, says recent report from leading charities.'

## Appendix 7 – Workshop 2 Participant Handout

### Design your own rules for data linkage (pro-forma)

Work in groups and use the cards for ideas (but you can use your own too!)

When should researchers be allowed to access and use administrative data that has been linked?

Who should be allowed to access linked administrative data?

What admin data would you allow to be linked for approved research?

Where should access to de-identified linked data be allowed to happen?

For how long should researchers be allowed to have access to linked data?

### Proposed ADRC principles for data linkage

When should researchers be allowed to access and use administrative data that has been linked?

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

Who should be allowed to access linked administrative data?

Researchers (i.e. those affiliated with a university or a government department – this could be a PhD student, an academic, or a government researcher) who have done the training offered by the ADRCs and gained an accreditation.

What admin data would you allow to be linked for approved research?

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.

Where should access to de-identified linked data be allowed to happen?

In safe settings, which in practice would mean in a secure area within the University that hosts the ADRC, on secure servers with limited access for those who have been vetted and checked. In the future the ADRCs may make it possible to access these secure servers remotely. However, the linked data would remain in the secure server – it would never be transferred via the internet

For how long should researchers be allowed to have access to linked data?

Just for the duration of their research project

## Appendix 8 – Workshop Evaluation Questionnaires

Thank you for taking part in today's workshop. We really appreciate your input. Please take five minutes to complete this questionnaire.

Your **answers will be completely confidential**. Your **views and feedback are really important to us**, and will help us make future events better.

Please answer the questions by ticking or putting an x in the box that best describes your view.  
**Many thanks!**

Q1. What was your main impression of the day?

--

Q2. How far do you agree or disagree with the following statements about the workshop?

PLEASE TICK ONE BOX ON EACH LINE	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree
a. There was enough time to fully discuss the issues	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. The information provided was fair and balanced	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. I understood the information provided	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. I understood the purpose of the workshop	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. I understood how the results of the workshop will be used	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. Attending this meeting has changed my views	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g. I learnt something I did not know before	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
h. I enjoyed taking part	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

PLEASE TICK ONE BOX ON EACH LINE	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree
i. I was able to discuss the issues that concern me	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
j. All participants were treated equally and respectfully	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
k. My views have been listened to	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
l. My views have been treated with respect by the experts	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Q3. Is there anything we could have done to improve the workshop?

Q4. Is there anything else you would like to add about today's event?

**Thank you very much for your help**

## Appendix 9: Public Attitudes to Science 2014 Data Module Topline Results

This topline shows results from the Public Attitudes to Science 2014 survey, which consisted of a UK-wide survey of adults aged 16+ and a booster survey of 16-24 year-olds. Interviews were conducted face-to-face, in-home, from 15 July to 18 November 2013. Data are weighted to reflect the UK population profile.

<b>BIG DATA MODULE</b> <i>Asked of a quarter of the sample in 2014 (446 adults; 117 16-24 year-olds)</i>			2014 all adults	2014 16-24 YOs
<b>Q36</b>	<b>Which, if any, of these services have you decided not to take up because of concerns about how your data would be used?</b>	An account with a social networking site (such as Facebook, Twitter or Instagram)	28	7
		An electronic travelcard that allows you to "touch in" on buses or at train stations (such as an Oyster card)	13	11
		A loyalty card with a supermarket or shop (including online stores)	12	14
		A free email account (such as Yahoo Mail or Gmail)	10	7
		An account with a mobile phone network	9	8
		An account with an internet service provider to access the internet at home	8	7
		A bank account	6	7
		Other	*	0
		None of these/no answer	52	59
		Don't know	2	6
<b>Q37</b>	<b>And which, if any, of these services have you previously stopped using or changed to be with a different provider because of concerns about how your data was used?</b>	An account with a social networking site (such as Facebook, Twitter or Instagram)	7	9
		A loyalty card with a supermarket or shop (including online stores)	5	2
		An account with a mobile phone network	4	8
		A free email account (such as Yahoo Mail or Gmail)	3	5
		An account with an internet service provider to access the internet at home	2	1
		A bank account	2	1
		An electronic travelcard that allows you to "touch in" on buses or at train stations (such as an Oyster card)	1	1
		Other	0	0
		None of these/no answer	83	73
		Don't know	2	6

Q38	To what extent do you agree or disagree with the following 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.	Strongly agree	21	30
		Tend to agree	40	38
		Neither agree nor disagree	10	7
		Tend to disagree	15	13
		Strongly disagree	11	5
		Don't know	2	7
		Agree	61	68
		Disagree	27	19
Q39	Here are some specific ways in which people's data can be used. In each of these instances, the data is anonymised, so it can't be linked back to individuals. To what extent do you support or oppose each of these uses of people's data			
Q39A	Using data from shop loyalty cards to target products at people who are more likely to want them	Strongly support	7	8
		Tend to support	29	31
		Neither support nor oppose	19	20
		Tend to oppose	23	18
		Strongly oppose	21	19
		Don't know	2	4
		Support	36	39
		Oppose	43	37
Q39B	Using data from electronic travelcards (such as Oyster cards) to help improve the scheduling of buses or trains for passengers	Strongly support	24	30
		Tend to support	49	39
		Neither support nor oppose	12	13
		Tend to oppose	7	9
		Strongly oppose	4	5
		Don't know	5	5
		Support	73	68
		Oppose	11	14
Q39C	Websites using people's online browsing histories to create personalised adverts for products that people are more likely to be interested in	Strongly support	2	5
		Tend to support	16	19
		Neither support nor oppose	17	15
		Tend to oppose	28	28
		Strongly oppose	33	29
		Don't know	4	4
		Support	18	24
		Oppose	62	57
Q39D	Combining the data held by multiple government departments and using them to better tailor public services to individuals	Strongly support	13	13
		Tend to support	43	45
		Neither support nor oppose	19	23
		Tend to oppose	14	10
		Strongly oppose	6	3
		Don't know	4	7
		Support	56	58
		Oppose	20	12



Q39E	Using police and crime data to predict and plan for crimes that might take place in the future	Strongly support	24	32
		Tend to support	46	42
		Neither support nor oppose	14	14
		Tend to oppose	6	4
		Strongly oppose	6	5
		Don't know	4	3
		Support	70	74
		Oppose	13	9
Q39F	Offering discounted mobile phone calls and texts, funded by personalised adverts based on the content of people's text messages	Strongly support	2	5
		Tend to support	7	20
		Neither support nor oppose	16	18
		Tend to oppose	29	28
		Strongly oppose	41	25
		Don't know	5	5
		Support	9	25
		Oppose	70	53
Q39G	Creating a DNA database of cancer patients, in order to help develop more effective treatments for cancer	Strongly support	51	52
		Tend to support	37	34
		Neither support nor oppose	5	8
		Tend to oppose	3	2
		Strongly oppose	2	2
		Don't know	2	3
		Support	88	86
		Oppose	6	3
Q40	You said you oppose people's data being used in some of these ways. What makes you oppose these uses? <i>Base: all who oppose at least one of the uses of people's data mentioned (382 adults; 88 16-24 year-olds)</i>	Abuse of personal information (such as bank details)/ identity theft	40	34
		People have a right to privacy	32	35
		Being sent spam/ junk mail	18	27
		I don't trust private companies/ don't want them to profit	18	14
		Haven't got people's consent	17	15
		I don't know what the information is used for	15	8
		Hackers/ other people getting hold of data	14	6
		I don't want people/ organisations to know that much about me	13	10
		I can't see what information is held on me	10	10
		Depends on what information they want	8	2
		I don't trust the Government	6	5
		I don't trust the police	2	3
		Nothing in particular/ that's my view	1	1
		Don't know	3	2

<b>Q41</b>	<b>The analysis of large datasets often requires the use of supercomputers that use electrical power. How much of an impact, if any, do you think these supercomputers will have on the UK's energy consumption in the future.</b>	A very big impact	19	18
		A fairly big impact	27	33
		Not a very big impact	34	29
		No impact at all	8	8
		Don't know	12	13
		Very/ fairly big impact	46	51
		Not very big/ no impact	42	36

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

Ipsos MORI  
79-81 Borough Road  
London SE1 1FY

t: +44 (0)20 7347 3000  
f: +44 (0)20 7347 3800

[www.ipsos-mori.com](http://www.ipsos-mori.com)  
[www.twitter.com/IpsosMORI](https://www.twitter.com/IpsosMORI)

### ***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.