Public views on using government administrative data for research: public dialogue method and findings

Sarah Pope, Ipsos MORI

16.05.14
Survey research (see www.ipsos.com/pas2014)

- Longitudinal survey of public attitudes towards science – 5th in a series
- 1,749 UK adults aged 16+ and a booster sample of 315 16-24 year-olds interviewed in-home from July to November 2013

Changes for the 2014 study:
- Move to probability sampling, funded by ESRC (quota sampling in previous years)
- Move to a modular approach, with questions asked on:
  - Big data
  - Robotics
  - Agri-science
  - Energy technologies
Why Dialogue?

Not very well known or understood topics! Participants would need:
- Lots of time to learn, discuss and assimilate (Event 1)
- Lots of time to deliberate, debate and prioritise (Event 2)
- Access to experts who could answer their questions and help them make sense of it all

Key part of the remit of the ADRN and ADRCs is public engagement. Dialogue allowed representatives to:
- Meet the public and debate the issues with them
- Have their own views challenged
- Allow the public’s knowledge & views to shape PE strategy

“Public Dialogue brings together members of the public, policy makers, scientists and other expert stakeholders to deliberate and come to conclusions on national public policy issues involving science and technology”

Having funded the ADRN, the ESRC wanted to understand public views on the work that they would be undertaking.
As part of the wider consultation on the future of the Census, the ONS wanted to add to their evidence base about attitudes towards the front running options for Beyond 2011.
Methodology
So what did we do, and where?

7 sets of day-long, reconvened deliberative dialogue workshops in October & November 2013

- Wide geographic spread, covering all of the UK
  - Wales (Cardiff and Wrexham)
  - Northern Ireland (Belfast)
  - Scotland (Stirling)
  - England (Manchester, London, King’s Lynn)
- 18-20 participants per workshop, reflecting a range of criteria
  - Age
  - Gender
  - Ethnicity
  - Social grade
  - Attitudes (trust in researchers, views on uses of data)
- Included age specific workshops in:
  - Cardiff (all over 55)
  - King’s Lynn (all under 35)
Who helped us?

OUTSIDE OF THE WORKSHOPS: STEERING GROUP
Stakeholders from ESRC, ONS, the ADS, BIS, academia and other interested organisations.

Discussion of objectives, sampling and materials design process.

MATERIALS DESIGN: INFORMATION FROM ESRC & ONS

Comment on dialogue materials

Comment on dialogue findings and draft report

DURING THE WORKSHOPS: ROLE OF EXPERTS

Explain their work and how they use data linking

Answer participants’ questions about social science research, the census and data linking

Engage in dialogue and debate with participants

Workshop with participants, ADS and ONS to comment on process and key findings
How did we explain it all to the participants?

Previous research indicates that the general public have very low awareness of the scientific method, the production of statistics and data protection. So we had to take it slow and build knowledge across the two days!

- Spontaneous group discussion & brainstorming
- Presentations (Event 1 and 2)
- Case studies of data linking projects
- ‘Ask a researcher’ session
- Data linking ‘journey’ visualisation
- Video about changes to the Census
Example of materials presented – very simple, lots of examples

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

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)
Example – case studies allowed participants to understand the real world uses of linked data

Aims of Offending, Employment and Benefits project

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.

Uses of Offending, Employment and Benefits project

How has the linked data been used?

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

  26% of out-of-work benefits are claimed by offenders

  5% were claimed by ex-prisoners

• 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)
Example - the ‘data journey’ – took over an hour to explain in full!

Data linkage journey - who is involved?

- **Members of the public** – people like you, who give data to government departments when using government services. Example: 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. Example: 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 country that enables secure data linkage and supports researchers.

Data linkage journey - Step 1 – Data giving

- **2011**
  - Adam – data giver
  - Department for Work and Pensions – data owner
  - Jobseeker’s Allowance
  - Name and other personal details
  - Work history
  - History of claiming JSA

- **2012**
  - Adam – data giver
  - HMRC – data owner
  - Name and other personal details
  - Earnings
  - Place of work, start date of employment

Data linkage journey - Step 2 – Project and researcher approval

- **Researcher Approval**
  - Check skills
  - Prescribe training
  - Give accreditation

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

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

Data linkage journey - Step 4 – ADS

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

Department for Work and Pensions - data owner

2012

Adam – data giver
- Name and other personal details
- Earnings
- Place of work, start date of employment

Adam’s employer

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

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
Findings
### Understanding of social research

Across all workshops participants’ lack of familiarity with the aims and methods of social research meant that they made many incorrect assumptions about:

- The extent of what is not known or properly evidenced about society
- The process by which research happens.
- The aggregation of data.
- How understanding is generated from data
- The use and interpretation of findings

### Value of social research

For the most part though, participants attached some value to social research. However they tended to criticise it for a number of perceived failings, outlined below:

- Profiling or pigeonholing individuals or areas
- Low impact
- Overlooking context
- Researching the ‘wrong type’ of people

Early on, there were many questions about the ‘point’ of research and concerns about wasted effort, and scepticism about whether social research always leads to social value.
Value of data

**Views**

- Resignation to idea that their data is used in myriad ways
- Concerned about identity theft, and personal data being sold on
- Overall trusted government’s intentions (if not ability) on data security and protection from data misuse or fraud
- Assumed that governmental administrative data is already linked and shared across departments

**Implications**

- Private sector use of data for marketing conflated with use of data for research
- Views on personal data security framed much of the discussion
- Concern about what happens when data ‘leaves’ government - need reassurance ADRCs can keep secure
- Difficult to understand why ADRN is needed, and subsequent frustration that data will be linked without operational benefits

---

*I don’t have any objections to anyone knowing anything. I’ve got nothing to hide. My problem is when they sell the info.*

*What about loss of records? I’m very concerned about that*

*I don’t want my data going anywhere other than the person I give it to*
Q. 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
- % tend to agree: 40
- % neither agree nor disagree: 10
- % tend to disagree: 11
- % strongly disagree: 2
- % don’t know: 15

Base: 446 UK adults aged 16+

General consensus in workshops that data linking for social research is permissible once fully de-identified – mirrors PAS survey findings
Q. To what extent do you support or oppose each of the following specific uses of people’s data?

- Using police and crime data to predict and plan for crimes that might take place in the future: 70% support, 30% oppose
- Offering discounted mobile phone calls and texts, funded by personalised adverts based on the content of people’s text messages: 62% support, 38% oppose
- Creating a DNA database of cancer patients, in order to help develop more effective treatments for cancer: 88% support, 12% oppose
- Using data from electronic travel cards (such as Oyster cards) to improve the scheduling of buses or trains for passengers: 11% support, 89% oppose
- Combining the data held in multiple government departments and using them to better tailor public services to individuals: 20% support, 80% oppose
- Websites using people’s online browsing histories to create personalised adverts for products that people are more likely to be interested in: 43% support, 57% oppose
- Using data from shop loyalty cards to target products at people who are more likely to want them: 56% support, 44% oppose
- PAS 2014 found that purpose is hugely important, and this was a key point of discussion during the dialogue who wanted to know about impact.
## Views of the specifics of ADRN plans

<table>
<thead>
<tr>
<th>Concerns</th>
<th>Reassurance points</th>
<th>Remaining questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>- De-identification – will it be possible to identify people even without their ‘personal details’</td>
<td>✓ ADRCs’ independence</td>
<td>? Can this improve operational use of administrative data?</td>
</tr>
<tr>
<td>- De-identification process – is this secure? Where does the personal data ‘go’?</td>
<td>✓ Data could improve public services or allocate funding better</td>
<td>? How much have government departments ‘bought in’ to this idea?</td>
</tr>
<tr>
<td>- Security – is it possible to hack? Does that even matter?</td>
<td>✓ Efficiency – Save money, or reduce data collection burden</td>
<td>? Who works for the ADS and how are they governed?</td>
</tr>
<tr>
<td>- Cost – is this £34 million well spent?</td>
<td>✓ Better regulation than current system</td>
<td>? What will be the punishment for any misuse of data by researchers?</td>
</tr>
<tr>
<td></td>
<td>✓ No ‘super-database’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>✓ Limitations on researchers</td>
<td></td>
</tr>
</tbody>
</table>
Q. What makes you oppose people’s data being used in some of these ways?

Abuse of personal information/identity theft (40%)

Don’t trust private companies/don’t want them to profit (18%)

Don’t know what the information is used for (15%)

Don’t want people/organisations to know that much about me (13%)

Can’t see what information is held on me (10%)

Haven’t got people’s consent (17%)

People have a right to privacy (32%)

Being sent junk mail/spam (18%)

Don’t trust the police (2%)

Depends what information they want (8%)

Hackers/others getting hold of data (14%)

Base: 382 adults who oppose one of the potential uses of people’s data asked about

N.B. word clouds are illustrative of data, not statistically representative; only codes registering 2% or more are shown
Suggested ‘rules’ for ADRCs
When should researchers be allowed to access and use administrative data that has been linked?

PARTICIPANT VIEWS

- As per proposed rules in blue box above, provided and the research had been through methodological and ethical approval processes.
- No commercial gain for those who do the research
- If there’s a limit on the number of projects then prioritise those with the greatest social value

PROPOSED RULES PRESENTED TO PARTICIPANTS

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

As long as there’s reasonable intent

It should not be allowed for private companies’ private gain

Fine if it is for a positive ending, [but otherwise], why should you share that information?
Who should be allowed to access linked administrative data?

**PROPOSED RULES**

Researchers (i.e. those affiliated with a university or a government department) who have done the training offered by the ADRCs and gained an accreditation

**PARTICIPANT VIEWS**

- Once a researcher had undergone proper academic training and has been accredited, they should be allowed to access linked data for specific research projects, provided they are not accessing the data for solely personal gain, or profit.
- Even if they meet these criteria, researchers working for private companies should not be allowed to have access.
- Are PhD students, well-trained enough? Are they only doing the project for the benefit of their own personal research?
- Should researchers have different levels of access to linked data based on experience, training and expertise?
What admin data would you allow to be linked for approved research?

PROPOSED RULES
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.

PARTICIPANT VIEWS
• The overall view was that any government administrative data could be linked provided it would be impossible to identify individuals.
• However, there was also a lack of consensus about particular types of information including:
  • Religion
  • Full postcode
  • Travel patterns
  • Medical records
  • Financial details

As long as it’s anonymized it’s ok
Never personal financial details like bank accounts
All data should be allowed to be linked but we think that full postcodes are a little too data specific and should be limited to first one or two letters.
Where should access to de-identified linked data be allowed to happen?

PROPOSED RULES

In safe settings, which would mean in a secure area within the university that hosts the ADRC, on secure servers with access limited to those who have been vetted and checked. In the future the ADRCs may make it possible to access these secure servers remotely.

PARTICIPANT VIEWS

- Few concerns about allowing researchers to have access to data in a controlled and secure setting within the ADRC
- Significant concerns about allowing remote access to a secure environment because of general worries about internet security, and difficulty understanding how remote access works
**For how long should researchers be allowed to have access to linked data?**

**PROPOSED RULES**
For the duration of their research project

**PARTICIPANT VIEWS**
Participants tended towards one of two views:

- Linked datasets should be stored so that researchers could have access after the initial research project is complete

  - That’s a waste of money to start all over again for future use. There’s no personal information there

- Destruction of data after project completion reassuring and therefore desirable

  - For as long as they are in the room. If you forget something, that’s your fault. If you want more information you have to link it again

- A compromise suggestion was archiving linked data but researchers reapplying for permission to work with it again

  - They [databases] should be archived but you still have to go through the same checks
Participants recognised a pattern in their own reactions to ADRN plans...

But split on comms implications...

**Tell people nothing**
- It’s just too complicated
- Might panic people unnecessarily
- Telling everyone would be costly

**BUT**
- This could lead to suspicion
- Some will be interested, and understand

**Therefore**
- Be as transparent as possible
- Keep active comms minimal, but stress the messages of de-identification, security, benefits
Beyond 2011 – Views on changes to the census

Numerous questions around proposed administrative data route
Why the 4% survey? What can that tell you about the whole population? Why not more or less? Why face to face? Who are the 4%? How are they selected?

COST
✓ Sounds great - Let’s save money where we can!
❖ Suspicion re: projected saving
  ? Are the figures right?
  ? Do they take everything into account?
? Where will the saving be spent

FREQUENCY
✓ Statistics should be based on up to date data – the current census dates too quickly
❖ Why not just do the census more frequently?
❖ Society doesn’t change that often – do we need this?

LOSS OF DETAIL
? What will be the impact
? Why can’t administrative data be used to replace this detail?

FAMILY HISTORY
✓ We will have other ways of researching this in future
❖ But this is part of the value of the census to society
Conclusions
Conclusions and implications 1 – ADRN context

Little prior familiarity with or understanding of the purpose and processes of social research, which is crucial to understanding the reason for linked data.

- Demonstrates the scale of the challenge of fully informing the general public about ADRN

As participants gained understanding of social research and the ADRN plans, they tended to become less concerned and more supportive.

- For many, getting to a place of full understanding takes a day of intensive discussion, unrealistic to replicate on a broader scale

Participants did take a strong interest in the uses and outcomes of social research, particularly in areas they see as most interesting and relevant (e.g. health and education)

- General public communications around the ADRN should focus on the societal benefits of social research. The centres should ensure that they develop strong case studies that show how the projects that they have enabled have led to policy change, or service improvement
Conclusions and implications 2 – Informing and reassuring the public

• Public likely to be broadly happy with administrative data linking for research provided:
  • those projects have social value, broadly defined
  • data is de-identified
  • data is kept secure
  • businesses are not able to access the data for profit

• Messaging around the ADRN should focus on addressing the first three issues, and further work should be done to fully understand the public’s view on business access:
  • Researchers who want to use linked data should submit concrete plans for publishing their work, as well as ensuring that relevant policymakers are aware of the results
  • The network should develop the several ways to explain de-identification for different audiences. Language will be important (e.g. anonymised instead of de-identified). In workshops, visual techniques worked well
  • This high level of concern around data security should be reflected in ADRCs’ and ADRN operational plans. Virtual access need to be clearly demonstrated to be as secure and as easy to monitor as access through a secure physical setting

• Support for the ADRN was also driven by the view that its remit is sufficiently bounded, and by a desire to see better, more efficient and cost-effective use of administrative data.
  • Messaging should explain clearly how the ADRN is designed to streamline and make an existing process more efficient, cost-effective and even more secure.
Conclusions and implications 3 – ADRN remit and public engagement

• Overall, participants were reassured by strong governance and be regulation. However, there were some clear ‘red lines’ in terms of the ADRN remit
  
  • Further public dialogue would be needed for any expansion of the ADRN’s remit, in particular with relation to:
    • Creating large databases containing many variables/data from a large number of public sector sources
    • Allowing administrative data to be linked with business data
    • Linking of passively collected administrative data, in particular geo-location data
    • Allowing researchers for private companies to access data, either to deliver a public service or in order to make profit

    The first three above were seen as having potential privacy implications or allowing the possibility of re-identification of individuals within datasets

• There may be low interest from general public in becoming actively involved with the ADRCs or the ADS. Transparency was seen as sufficient
  
  • Communications strategy for the ADRN should focus on transparency
  • However, ensuring some public representation in the decision-making process for ADS and ADRCs could be important in proving the transparency of these bodies
Conclusions and implications 4 – Participants’ views

• Feedback was very positive - participants found the dialogue enjoyable, interesting, informative/educational, well structured, thought-provoking and interactive.

• Participants’ enjoyed:
  • Learning something new
  • Hearing from experts
  • Listening to views of peers and other participants
  • Voicing their views, asking questions and being listened to
  • Feeling involved and valued
  • Feeling like their opinion had made a difference.

• A handful thought the day was too long, a bit repetitive, or the topic too constrained.

Source: Word cloud based on c.100 answers to ‘What was the best thing about being a participant in the evaluation forms’