Smart Data Research UK

APPENDIX A&B

Smart Data Research UK Public Dialogue 2025

sciencewise

Thinks Insight & Strategy

Contents

Appendix A: Methodology

1. Detailed methodology

1.1 Why public dialogue

Public dialogue is a process that provides members of the public, who would otherwise not engage with a topic, with information and the space to consider an issue. It usually takes place over a series of workshops over a period of weeks, to allow participants to consider information before deliberating on the issue with fellow participants. Public dialogue works best with topics that have an impact on society, are not well understood, are controversial or difficult to navigate.

Exploring public views on smart data research through public dialogue is effective as it is a complex topic, and the use of data can be contentious. This form of deliberative research gives time for participants to delve into their values and beliefs, hear information from those working in the field and views of other participants, to go beyond their front of mind thoughts and develop informed recommendations.

Through public dialogue we can explore areas of agreement and disagreement and the wider values and beliefs that underpin participant expectations. It isn't necessarily about reaching full consensus, in fact, participants are encouraged to disagree and debate with one another. This method of engagement is also not representative of the views of the population.

Role of the oversight group

The project was steered by an oversight group, that brought together specialists to help shape the design of the dialogue, provide specialist input to workshops and ensure findings are relevant and can feed into other strands of work within this space. Further details on membership and meetings of the oversight group are in section 1.2.3 below.

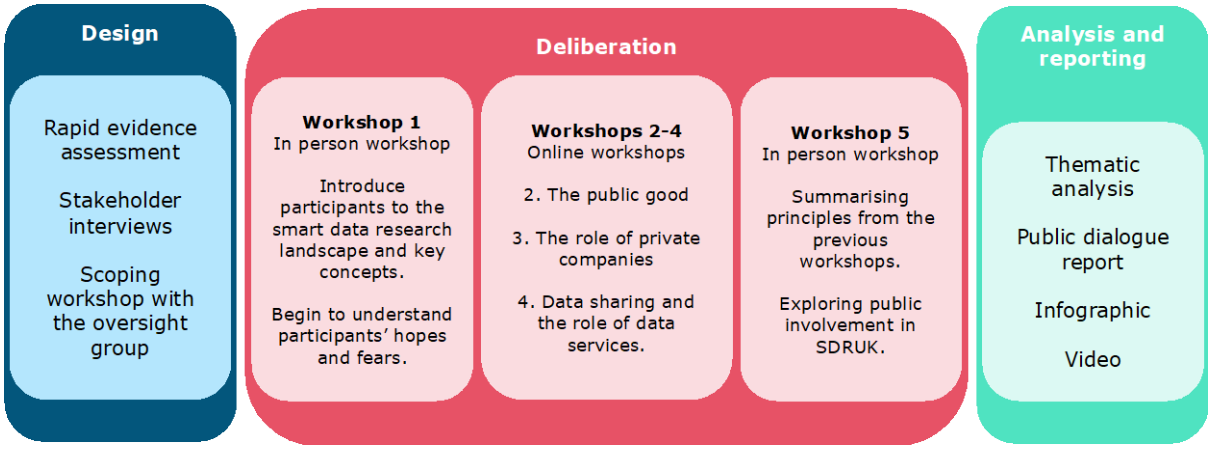
Role of the evaluator

An independent evaluator was commissioned to help ensure the dialogue follows Sciencewise principles. Navigator Consulting attended all project team meetings, oversight group meetings and a significant number of the dialogue events. Insights from the evaluator informed the design and analysis of the dialogue. Surveys and interviews with dialogue participants, oversight group members and the project team were also conducted as part of the evaluation. The full evaluation method and findings will be available in a separate evaluation report.

1.1 Dialogue design

An overview of the dialogue process is provided below, in Figure 1.

Figure 1. Overview of the dialogue process



Scoping phase

Thinks Insight & Strategy conducted a rapid evidence assessment in the scoping phase of the dialogue. 21 documents were reviewed with the aim to:

- better understand the potential risks and benefits of smart data research
- gather case studies
- understand what is already known about public attitudes to data sharing in comparable contexts
- explore what is known from previous engagement on data access programmes.

The findings of the rapid evidence assessment were used to inform the design of the dialogue. The full rapid evidence assessment can be found in appendix D.

Scoping interviews were also conducted with 6 specialists (in the table below) with expertise and perspectives including digital rights, governance and innovation, data science, consumer protection and equality. The interviews were used to help shape the dialogue, with the aim of the scoping phase to:

- Define the scope of smart data research.
- Understand similarities and differences with other data research units and the public engagement conducted on their behalf.
- Understand the ethical, legal, and social issues related to smart data research.
- Hear from experts to understand what they think the public needs to know to have an informed discussion about SDR.

Table 1. Specialist scoping interviews

Name	Role / Organisation
Javier Ruiz Diaz	Digital Policy Consultant, Centre for Inclusive Trade Policy
Giles Pavey	Global Director – Data Science, Unilever

Tilly Cook	Senior Policy Researcher, Citizens Advice
Shayda Kashef	Senior Communications Manager, Public Engagement: ADR UK
Dr Ros Williams	Senior Lecturer in Digital Media and Society, University of Sheffield
Oliver O'Brien	Researcher and Software Developer, University College London

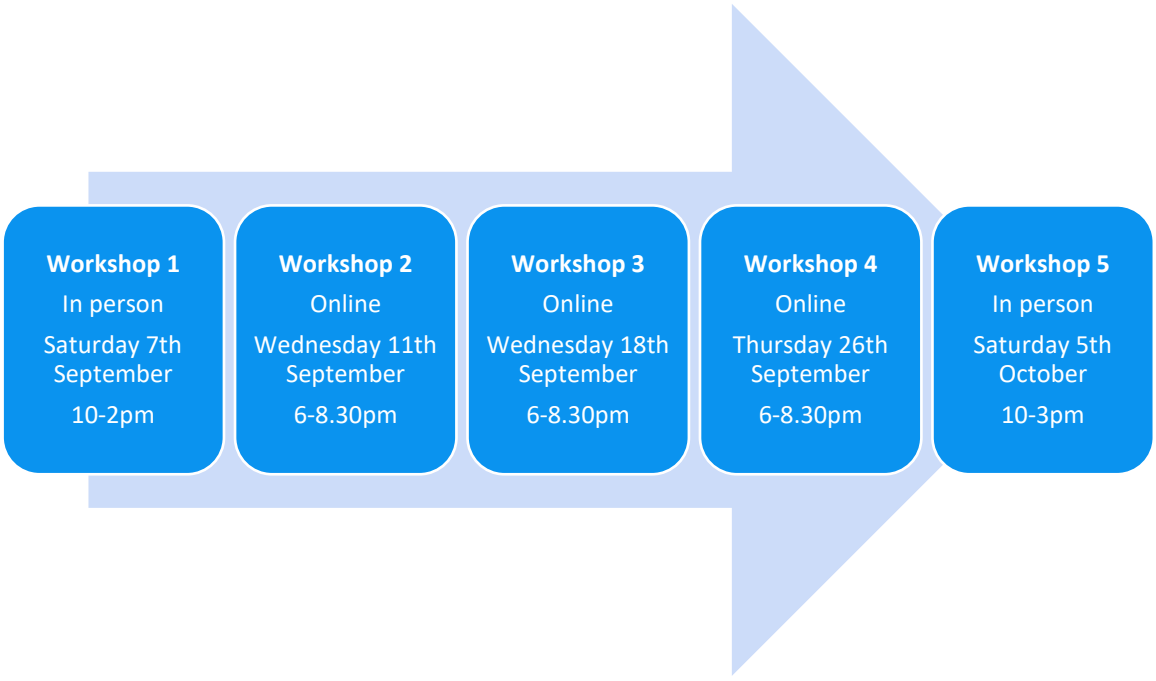
Approach to dialogue design

The dialogue design was led by Thinks Insight & Strategy. The SDR UK, Thinks and Sciencewise project team met weekly throughout the course of the project, feeding into design of the workshops. The findings from the scoping interviews and rapid evidence assessment were also used to inform the design. We also held a design workshop with the oversight group, to get their views and steer on design across the dialogue sessions and to inform the requirements for specialist involvement in the workshops (alongside stakeholder scoping interviews).

Design of the public dialogue

The workshops took place over the course of a month, from September 7th to October 5th 2024. It consisted of five workshops, split between full day face-to-face and evening online workshops (see Figure 2). Participants deliberated for a total of 16.5 hours.

Figure 2. Overview of the dialogue workshops



The in-person workshops took place in five cities across the UK (more detail in section 1.2.1 below) and the online workshops brought together participants from across the regions in one online workshop. Longer in person sessions at start and end were used to allow for group formation and confidence building within the participant group. At the end of the dialogue, the longer session was used to effectively use a wider range of engagement techniques than are ordinarily implemented online to build the recommendations. The three shorter online sessions gave participants access to information in a more accessible format with shorter discussions to build knowledge and understanding.

Each workshop was led by a senior member of the Thinks Insight & Strategy project team. During the workshops, participants were split into small breakout groups of up to 8 people to discuss and deliberate on the information they had received. Each breakout group was led by an experienced Thinks facilitator.

Table 2 outlines the structure of each workshop. Full discussion guides and the stimulus used are available in appendix B.

Table 2. Workshop structure

Workshop	Topics covered	Stimulus
1	**Purpose:** Introduce participants and build understanding of smart data and smart data research, publicly funded and social research and the key concepts of data access and regulatory baseline conditions. Begin to understand participants' spontaneous hopes and fears in relation to smart data research.	
	Topics: • What is smart data and SDRUK? • GDPR • Publicly funded research • Understanding the SDR eco system	Stimulus: • Presentations from SDRUK • Video testimony from specialists in the sector • Ranking exercises • Explanation videos
2	**Purpose:** Understand spontaneous views of what defines "research for the public good" and priorities for future research. Explore what rules participants would put in place to ensure smart data research works for the public good.	
	Topics: • Research for public good • Inequalities	Stimulus: • Presentation from SDRUK • Specialist presentation and Q&A

<table>
<tr><td></td><td>Potential harms of smart data research
Rules and priorities for smart data research</td><td>Case studies</td></tr>
<tr><td rowspan="3">3</td><td colspan="2">Purpose:
Explore participant views on drivers and barriers for private companies to become involved in smart data research, explore potential harms related to private company sharing of smart data. Discuss rules participants would put in place to ensure commercial relationships are fair.</td></tr>
<tr><td>Topics:
Private benefits
Motivations of private companies to share data
Key risks associated with using data from private companies</td><td>Stimulus:
Case studies
Specialist presentation and Q&A</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td rowspan="3">4</td><td colspan="2">Purpose:
Explore principles around data sharing and governance from other data services and trusted research environments and how they would apply to smart data research. Discuss the principles participants would put in place to ensure data is handled safely and securely.</td></tr>
<tr><td>Topics:
Reminder of smart data research ecosystem and key terms
Privacy and data security
Principles for data sharing</td><td>Stimulus:
SDRUK Presentation
Specialist presentations and Q&A</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td rowspan="3">5</td><td colspan="2">Purpose:
Take stock on hopes and priorities for smart data research and SDR UK. Formulate recommendations for SDR UK across the themes of public good, public-private collaboration, data handling and security, and public engagement. Understand people’s expectations for public engagement in smart data research going forward.</td></tr>
<tr><td>Topics:</td><td>Stimulus:
Personas</td></tr>
<tr><td colspan="2"></td></tr>
</table>

	• Benefits and potential harms of smart data research • Research for the public good • The role of private companies in smart data research • Data sharing and the role of data services • Public involvement • Hopes and expectations for SDR UK	• Specialist presentation and Q&A

1.2 Who was involved in the dialogue?

1.2.1 Members of the public

72 members of the public were recruited from five geographical locations across the UK, with representation from each of the four nations (Belfast, Newport, Inverness, Gateshead and London).

Participants were recruited using professional recruiters (accredited by the Market Research Society). Participants were recruited to be broadly reflective of the general population in terms of age, gender and ethnicity.

Quotas were used to boost for people with a long-term health condition or disability and to ensure the sample covered a range of experiences and attitudes including digital footprint and connectedness (use of social media and smart devices) and attitudes to data sharing (a mix of those who feel more or less comfortable). Participants who had participated in market research within the last 6 months or research on this topic before were excluded from the research.

A total of 72 participants took part in the workshops, with 52 participants participating in every workshop (absences from individual workshops occurred due to e.g. participant illness). A detailed breakdown of the demographics of participants can be found in table 3 and attrition in table 4.

Participants received an incentive payment of £450 for participation in all workshops as a thank you for sharing their time and views. Given the length of

engagement, participants were paid a portion of their incentive after attending each workshop.

Table 3: Public sample

Demographic	Criteria	Intended	Achieved
Location	London	16	14
	Gateshead	16	16
	Newport	16	16
	Inverness	16	15
	Belfast	8	8
Gender	Male	32	29
	Female	32	40
Age	18-24	15	16
	25-44	15	22
	45-64	15	17
	65+	15	14
Socio-economic group (SEG)	ABC1	20	32
	C2DE	30	37
Ethnicity	White	20	41
	White non-British / Irish / Northern Irish	5	4
	Mixed or multiple ethnic groups	5	3
	Asian or Asian British	10	12
	Black, Black British, Caribbean or African	9	9
Health	Long term health condition / disability with a significant impact on daily life.	15	20

Extent of their digital footprint	High	18	26
	Medium	18	25
	Low	18	18
Family	With school age children	25	32
	Children in primary school (aged 5-11)	7	13
	Children in secondary school (aged 11-16)	7	11
	Children in sixth form or equivalent (aged 16-17)	7	8
	With no children	25	28

Table 4: Participant attendance numbers across the dialogue

Workshop 1	69
Workshop 2	66
Workshop 3	60
Workshop 4	59
Workshop 5	58

1.2.2 Specialists

Each specialist was briefed by a senior member of the Thinks Insight & Strategy's team on the content of their presentation and role at the workshops.

Specialists were identified in collaboration with the SDR UK project team, drawing on their own contacts and the rapid evidence review. Specialists were recruited from a wide range of professional backgrounds, covering the range of issues to be explored in the dialogue. This included professionals from commercial organisations, academia, data centres and public engagement. The specialists' role involved presenting information during the dialogue and engaging in Q&A's with participants (facilitated by Thinks' facilitation team). Each specialist was briefed by a senior member of the Thinks Insight & Strategy's team on the content of their presentation and role at the workshops.

Table 5. Specialist involvement across the dialogue workshops

Workshop	Topic	Speaker
1	What is smart data and SDR?	Joe Cuddeford, Director: SDR UK
1	Impact of public and social research on policy/government expert	Paul Monks, Chief Scientific Advisor: Department for Energy Security and Net Zero Dr Rachel Oldroyd, Lecturer in Geographic Data Science: Consumer Data Research Centre
1	Understanding SDR eco-system	SDR UK team member onsite (per workshop location)
2	Introducing 4 thematic pillars	Mark Gardner, Head of Communications and Public Engagement: SDR UK

2	Inequality and exclusion	Dr Maxine Mackintosh, Programme Lead – Diverse Data, Genomics England: The Alan Turing Institute
3	Private business representatives	Andy Morris, VP Chief Data Officer: Boots
4	Consumer protection	Cassie Smith, Head of Legal, Trust and Ethics: Health Data Research UK
4	Researcher on TRE's	Pete Stokes, Director of Platform Development: Bennet Institute for Applied Data Science
5	Public involvement expert	Fionnuala Ratcliffe, Dialogue and Engagement Specialist: Sciencewise Lucy Farrow, Partner: Thinks Insight & Strategy Catherine Joynson, Head of Participant Engagement: UK Biobank

1.2.3 Oversight group

Table 6. List of oversight group members

Name	Role / Organisation
Colin Griffiths	Policy Manager, Citizens Advice
Andy Morris	Chief Data Officer, Boots
Sofi Nickson	Head of Research, OSR
Alison Park (Chair)	Deputy Executive Chair, Economic and Social Research Council
Nick Bailey	Director, Urban Big Data Centre University of Glasgow
Peter Stokes	Director of Platform Development, Bennett Institute for Applied Data Science
Roger Halliday	Chief Executive Officer, Research Data Scotland

Dr Ros Williams	Senior Lecturer in Digital Media and Society, University of Sheffield
Stephanie Borthwick	Senior Policy Adviser, Which
Jared Keller	Independent Consultant
Javier Ruiz Diaz	Digital Policy Consultant, Centre for Inclusive Trade Policy
Octavia Field Reid	Associate Director, Ada Lovelace Institute
Tim Davies	Research and Practice Director, Connected By Data
Fay Skevington	Digital Strategy, AI in Education, Department for Education
Teodora Kaneva	Head of Smart Infrastructure and Systems, techUK
Alison Preston	Co-Director and Head of Research, Ofcom
Richard Syers	Principal Policy Advisor, Information Commissioner's Office
Kathy Peach	Director of the Centre for Collective Intelligence Design, Nesta
Yves-Alexandre de MontJoye	Associate Professor, Imperial College London

Table 7. Oversight Group meetings

Date	Activity	Expectation / Purpose
09/04/2024	Oversight Group meeting - 1	Introduction and project scoping • Key documents for the REA • Stakeholder interviews • Experts to provide testimony during fieldwork
05/06/2024	Oversight Group meeting - 2	Workshop design • Gathering feedback on workshop outlines and suggestions on stimulus

07/09/2024	Fieldwork begins - 3	Opportunity to attend/observe workshops
22/10/2024	Interim Debrief - 4	Emerging findings • Q&A with OG members who attended the workshops • Feedback on the findings with invitation to feed into the final report
14/05/2025	Publication of report	Oversight Group Members will take part in panel session discussing the themes and implications of the research.

Appendix B: Workshop Materials

Workshop 1: Discussion Guide and Materials

Workshop objectives:

The specific objectives for this workshop are to:

- Encourage participants to engage with their own digital footprint and introduce the idea of smart data as "footprint data".
- Introduce participants to key concepts of data access and regulatory/legal baseline conditions.
- Familiarise participants with publicly funded and social research.
- Understand participants' spontaneous hopes and fears in relation to smart data research.

Workshop guide (in-person) | September 2024		
Section and aim	**Key questions and probes**	**Time**
Arrival	*Participants arriving* *Participants will be asked to arrive from 9:45am* *Sign people in and distribute name tags* *5 mins before start, lead facilitator in each location to welcome participants*	-
1.1 Welcome, introduction to the research the session	Lead facilitator welcomes everyone. • Who's in the room • Purpose of the session	10
1.2 Warm-up and digital footprints Aim: Help participants to understand the breadth of smart data that exists and become aware of their own behaviours.	Participants introduce themself including: • Name • Where and who they live with • What they would usually be doing on a Saturday / Sunday Facilitator to ask if there are any questions about the research. *In this session, we are going to be discussing smart data, thinking about the types of smart data that might be collected about us, the rules and regulations in place around this, and how smart data can be used in publicly funded research. You might never have heard the term before or know what it means and that is ok! We'll be working together on a range of different activities today*	30

	which will help you to learn what you need to know about this topic. There are no right or wrong answers, so please feel free to share your thoughts! Facilitator sets a timer of 30 seconds, for participants to write down 6 pieces of information about themselves. Facilitator to collect all post-its and go through them one by one, asking participants to try to identify one another using them. Facilitator to explain that data is just information, and that some types or sets of data might be more sensitive or personally identifiable than others — while other forms of data might not be sensitive at all. *In today's discussions, we're going to be talking about something called smart data. We'll be hearing an explanation of what smart data is soon, but before we do, I want us to think about the data you share in your day-to-day life.* *When we are talking about smart data in this context we are talking about information that is recorded by digital technologies or online services. This could be information you provide deliberately e.g. by filling out a form, or information that is collected about you in the background (e.g. location data or app activity).* Facilitator to give participants 5 minutes to fill out their digital footprint maps, thinking about 4/5 things they did with digital technologies or online in a week, whether at home or elsewhere. Facilitator to encourage them to think about things they do using phones, computers or any other smart device they interact with, and to list the data they think was collected, specifying whether it was actively or incidentally shared, and who they think has access to it. If needed, facilitator to prompt on different categories of data: name, address, phone number, age, gender, ethnicity, health information, preferences, purchase history, shopping habits, browsing habits etc. **Can you tell me which data you think is collected from you over the course of a week?** • What type of data do you share?	

	• How it is shared (i.e. is it actively shared or incidentally shared)? • Who is it shared with (public vs. private organisations)? **Thinking about each piece of data and who accessed it, how do you think that organisation is using the data?** • What are the benefits for them of having access to the data? • Have you thought about how this organisation is using your data before? • Who else might have access to the data? • Have you seen any information from this organisation about what this data is used for? At 10:30, Lead facilitator to introduce speaker in plenary, via Zoom, to all locations. Q&A from participants. **Keeping in mind what you just learned about what smart data is and how it's collected, are there any other examples of data that you think could be included?** • How do you think this data is being collected? • How do you feel about this data being collected in this way?	
1.3 Physical ranking exercise	Participants rank some of the most commonly used data examples on a scale – that goes from "highly sensitive" to "not at all sensitive". *[Likely examples include: banking data, shopping data, location data, health data etc.]* *Lead facilitator to probe on reasons for ranking throughout, asking:* • *How sensitive it is* • *What makes this data more / less sensitive* Lead facilitator to ask if there are any circumstances where participants think they might be more or less comfortable with sharing data based on: • Who it is being shared with • What it is being used for	10

	• How it is being collected/given	
REFRESHMENT BREAK	10:50 – 11am	10
1.4 What is SDR? **Aim:** Introduce participants to smart data research.	Lead facilitator in each location to introduce SDR speaker to discuss how this data could be used in research. Presentation in plenary from SDR UK rep (summary on page 4-5 of participant handbook). *Moderator to lead location specific Q&A.*	**10**
1.5 Key concepts **Aim:** Introducing the key concepts that will be referred to throughout the research, as well as the 'baseline conditions' of regulation	*Lead facilitator in each location to play animations in plenary, via Zoom, to all locations. Participants to be given infographics at tables covering the information shared via the animation to follow along and identify any areas of confusion to be discussed at tables.* *Animation in plenary:* • *What is personal data* • *GDPR and ICO* • *Five safes and de-identifying data* *At breakout tables, 5 minutes of discussion to collect questions and ensure participants understand concepts explained. Space on pages 6-8 of the handbook for jotting down any questions, and a summary of the animation.* **From the animation we've just seen, what word would you use to sum up how you feel about the content?** • Why did you choose that word? • What, if anything, is surprising about what you have heard? **Do you have any questions or concerns about what you have heard?** **Based on what you've learnt so far, how do you feel about your data being collected and used now? Why?** *Lead facilitator in each location to probe on areas of consensus / disagreement, and any changes in opinion based on understanding of key concepts.*	20

1.6 The power of research **Aim:** Help participants understand the impact of social and economic research in different domains.	*Expert presentation in plenary, via Zoom, to all locations:* • Paul Monks, Chief Scientific Advisor at the Department for Energy Security and Net Zero (page 9 of the participant handbook for a summary and space for questions). • Dr Rachel Oldroyd from the Consumer Data Research Centre at the University of Leeds. (page 10 participant handbook). *Moderator to lead short Q&A* **How do you feel about your data being used for research? Why do you feel that way?** • Were you aware your data is used in this way? **Does knowing the impact research might have change how you feel about your data being used? Why / why not?**	15
1.7 What does smart data research look like in practice? **Aim:** To help bring to life the possibilities and opportunities for smart data research.	*Back in break-out groups, introduction of case studies, outlining SDR application:* • *Boots advantage card data and ovarian cancer (health) (page 11 participant handbook)* • *Banking data and what it tells us about the financial health of Older Workers and Later Lives (finance/social exclusion) (page 12 participant handbook)* • *Driving the smart use of consumer data in UK supermarkets (business perspective) (page 13 participant handbook)* • *Strava data and what it can tell us about town planning impacts on infrastructure (location data) (page 14 participant handbook)* *Each group to discuss 2 case studies:* **What are your initial responses to this example?** **In this example, who is sharing the data?** • Why are they sharing it? **Who is using the data?** • Why are they using it? What are they trying to achieve?	30

	• What impact could using the data in this way have? • On individuals? • On society as a whole? **Who benefits from data being used in this way? If so, who and how?** **Could anyone be disadvantaged by using data in this way? If so, who and how?** • Is this unavoidable? • What do you think could be done to prevent anyone from being disadvantaged? • Is there anyone who might miss out? **How would you feel about your data being used in this way?** *Moderator to probe on anything exciting or worrying about this example.*	
LUNCH	12:15– 12:45pm	30
1.8 Spontaneous hopes and fears **Aim**: To understand their spontaneous hopes and fears for smart data research and reactions to information shared so far, as well as where they see the biggest potential.	*Table facilitators to spend 20 minutes leading flipchart brainstorm session exploring spontaneous hopes and fears for smart data research, and issues that smart data could solve, both at a personal and wider societal level.* *[Table facilitator to talk participants through 1-2 line summary of all expert presentations, found on page 15 in workbooks].* **Thinking about what you have learned and what we have discussed today, what are your biggest hopes for smart data research?** What impact do you hope it will have: • On society? • On you personally? • Is there anything that particularly interests or excites you about smart data research? **Thinking about what you have learned and what we have discussed today, what are your biggest fears when it comes to smart data research?** • Are there any worries you have that had not occurred to you before today? • What might be the consequences if these worries came true?	50

	• Can you think of anything that could be done to address or prevent these worries? *Table facilitator to run physical ranking exercise [e.g. by asking participant to place a post-it note on a flipchart scale] to understand levels of comfort with data being used for smart data research, asking participants to place themselves on a scale that goes from not comfortable at all to extremely comfortable, based on their current comfort with their data being used for SDR. Probe throughout and encourage participants to move around if/as their views shift during discussion.* **How comfortable do you currently feel with your data being used for smart data research?** • Are there situations where you'd be happy or excited that your data was being used? • Are there any situations where you'd be worried or uncomfortable about your data being used? • What could be done to make you feel comfortable or confident knowing your data was being used in this way? • Can you think of any reasons why someone else might be uncomfortable? We're now going to think about how different people might feel about smart data research, and the impact it might have in a range of circumstances. We're going to meet all the residents of Wisteria Lane and their experiences with smart data and understand their digital footprints (page 16 participant handbook). *Facilitator to present different households and spend 15 minutes revisiting hopes and fears from their perspective and annotating flip charts with any additions. Moderator should focus on one or two residents who might particularly challenge currently held perceptions, based on the table's previous conversation, but allow participants to raise discussions of other personas if they are interested in them:* **Thinking about the people you've just met, what hopes do you think they would have for smart data research?** • Why might they feel this way?	

	• What issues could smart data research solve for the people on this street? **What fears do you think they might have about smart data research?** • Why might they feel this way? • Are these fears different to yours? Why / in what way? *Facilitator to repeat scale placement exercise for one or two personas, asking participants to work together to place them on a scale of not comfortable at all to extremely comfortable.* **How comfortable do you think they currently feel with their data being used for smart data research?** • Are there situations where they'd be happy or excited that their data was being used? • Are there any situations where they'd be worried or uncomfortable about their data being used? • What could be done to make them feel comfortable or confident knowing their data was being used in this way? *[If time] Facilitator to ask participants to re-rank hopes, fears, and issues based on both discussions, then to identify the top 3 of each to feed back in plenary.*	
1.9 Understanding SDR eco system **Aim:** Introduce participants to key players in SDR and how the process will work	*Lead facilitator in each location to re-introduce SDR UK speaker in plenary to discuss what smart data research looks like in the UK, including who is involved.* We are now going to tell you more about the role SDR UK (the organisation who've asked us to meet today) plays in smart data research and the key questions they want you to help them with over the course of this engagement (participant handbook page 17). Presentation in plenary from SDR UK rep introducing their role in smart data eco system: • Who are the key players? (i.e. private companies, SDR UK, data services, researchers and policy makers/media/the public) • What is the role of SDR UK?	10

	• What is the aim of this engagement? What is and is not in scope? Share key questions SDR UK has for participants. What should they do to ensure: • Research is for public good • Arrangements with private companies are fair • Data sharing is safe • The public are engaged in SDR • Q&A	
1.10 Wrap-up and close **Aim:** Ensuring everybody understands the expectations and how to take part in online sessions, process for incentive payments.	Lead facilitator in each location to give a short summary of the hopes and fears discussed. [Lead facilitator to come to lead facilitator in location in turn to ask for a summary, pulling out any areas of similarity or difference with the location before]. Thanks everyone – we'll be collating all your hopes and fears into a single word cloud to show you at the start of the next session. This word cloud will help us see some of the key themes that have come out of your discussions. Lead facilitator to thank everyone for your time today and reminder of the next meeting. Facilitator to remind participants to complete short survey about how they found participating in the session.	15

Stimulus

The materials used in workshop 1 include:

- Running slides, participant workbook, case studies and animations. These are available upon request.
- Some videos were shown to participants which are not in the public domain including an animation explaining personal data (script can be seen below).

Presentation from Joe Cuddeford, Director, Smart Data Research UK: What is smart data and using smart data for research.

The purpose of this presentation was to introduce the concept of smart data to participants including:

- What smart data is,

- Examples of smart data,
- How smart data is used.

Presentation from Paul Monks, Chief Scientific Advisor, Department of Energy Security and Net Zero: How can public and social research inform policy.

The purpose of this presentation was to explain to participants the ways in which smart data is used including:

- The power of smart data,
- Examples of how public and social data can be used to inform policy (including carbon production research),
- The future of smart data.

Presentation from Rachel Oldroyd, Lecturer, University of Leeds: Priority Places for Food Index.

The purpose of this presentation was to show participants one way in which smart data is used including the Priority Places for Food Index:

- The seven domains of the Priority Places for food index,
- The Priority Places for Food Index Dashboard and how it is used by policymakers, charities, retailers, nutritionists and consumer groups to tackle food insecurity,
- Policy activities.

Workshop 1 Animation script

Part 1: Personal data

Personal data is any data that relates to a specific person. It could be information about you, like your name, or address, date of birth, or medical diagnoses. It could also be about your behaviour, like your shopping history, or your banking transactions.

As you have heard, when we talk about "smart data", we mean different sources of data in the digital world: sometimes this is personal data, and sometimes it is not.

For example, environmental data from satellites is used by weather forecasters, and web service data about how well the service is functioning is used by cybersecurity teams. Both of these data sources could be referred to as smart data, but neither of these data sources will typically need to contain personal data.

So, not all smart data is personal data, but when it is first generated by people, it probably **is**.

Part 2: Existing data rules

In the UK we have strict laws that govern the way personal data can be collected and used.

The General Data Protection Regulation (GDPR) is a law that controls how personal information is used by organisations, businesses or the government. An organisation called the Information Commissioner's Office (ICO) makes sure the rules are followed and can issue large fines if an organisation fails to comply.

GDPR recognises the importance of scientific research, so enables researchers to process personal data as long as this is necessary for research purposes and appropriate safeguards are in place.

Part 3: Safeguards

You're probably wondering how your privacy is protected when your data is used for research. Where does all the personal data go, and how is it kept safe?

This is done by following a set of principles called the Five Safes Framework.

- *Safe Data*
- *Safe Projects*
- *Safe People*
- *Safe Settings*
- *Safe Outputs.*

Today we'll be focusing on the first of these: Safe Data.

Data is usually stored together in a dataset, which means it's organised and stored in a digital file.

Before the data is used for research, information about your identity is removed from the dataset.

This is sometimes called 'de-identification'. Information that could easily be used to identify you, such as your name, is removed.

For example, Deborah is a researcher at a university, researching how to reduce rates of obesity. She analyses data from supermarket loyalty cards, to see what effect buy-one-get-one-free offers have on buying unhealthy food.

To answer her research questions, she doesn't need to know the name of each individual customer who bought a chocolate bar, and she doesn't need to contact them individually. She can use de-identified data to find the answer.

To support her analysis about the effects of shopping on nutrition across society, she might want to know more information about the people who are buying the products – for example, are they men or women? Which age group are they?

This would be personal data, but she is still not needing to know about each individual person. She can look at groups of customers instead.

De-identified data may still be personal data. It might be possible to identify an individual person within a de-identified dataset.

So, a range of other safeguards from the Five Safes Framework are used at the same time, to ensure that the data is used safely. In later sessions, we will talk more about the Five Safes Framework and see what it looks like in action.

Participant workbook

Page 1: Welcome

Thank you for taking part in the project!

We are Thinks Insight & Strategy, an independent research company running this research on behalf of Sciencewise and Smart Data Research UK (SDR UK). In the sessions, we would like to find out more about your views on smart data and its use.If you have any questions during the research, please contact **sdrengagement@thinksinsight.com**

Thanks again for agreeing to participate and we really look forward to working with you!

Page 2: What is smart data?

Joe Cuddeford, Director of Smart Data Research UK

'Smart data' is a general term we use to cover a wide range of data generated as a byproduct of our digital lives.

- Shopping data: Through loyalty cards and accounts, data is captured at checkout. It goes into a database where the retailer can run analysis.
- Internet, apps, streaming: Social media, apps, streaming platforms, online games, search services usually track some information about your use of the service via cookies.
- Banking data: Contactless payments and online banking can give insight into our spending habits. Data is often shared between banks to give you a full picture.
- Transport data: There are sensors monitoring the condition of the road or railways, GPS devices mapping our journeys with pinpoint accuracy, and we tap in and out of public transport.
- Remote images: Satellite images can give us data about out the planet's surface – what's happening to our rivers, forests and cities.
- Smart devices: We have wearable devices such as smart watches monitor your heart rate, movement, location, but also use smart appliances in our homes. These devices record data about their use.

Page 3: What is smart data research?

Representative from Smart Data Research UK

There are many different possibilities for researchers using smart data. Here are four examples.

- How our shopping habits impact health: By looking at data from loyalty card schemes, researchers can shed light on people's diets or identify parts of the country where healthy food is less available.
- Understanding regional differences: By combining various types of smart data, researchers can study regional inequalities. That can help us to understand how opportunities differ in different parts of the country (such as higher unemployment, or where public services need to improve). This can help the government design projects to improve places and services that are struggling.
- Supporting urban development: We can use images from satellites as well as other smart data to make better planning decisions, so that houses are built where they're needed, while protecting the environment. Or to plan improvements to towns so that people have access to the transport services they need.
- Digital wellbeing: Smart data can also unlock insights into our digital lives. For instance, researchers want to know more about the effect that social media has on young people's mental health. Or study how disinformation and hate speech spreads on social media.

Page 4: How is smart data shared with researchers

- Personal data is protected under strict laws.
- The law allows personal data to be used for research, but it needs to be handled very carefully.
- It can still be difficult to persuade private companies to share data with researchers.

Page 5: Animation: What is personal data

You will now be shown a short animation about personal data. If you have any questions, you can write them below.

- Personal data is any data that relates to a specific person. It could be information about you, like your name, or address, date of birth, or medical diagnoses. It could also be about your behaviour, like your shopping history, or your banking transactions.
- "Smart data" can be from different sources in the digital world: sometimes this is personal data, and sometimes it is not, for example, satellite images.
- So, not all smart data is personal data, but when it is first generated by people, it probably is.

Page 6: Animation: Existing data rules

You will now be shown a short animation about the existing rules around smart data. If you have any questions, you can write them below.

- In the UK we have strict laws that govern the way personal data can be collected and used. This includes the General Data Protection Regulation (GDPR).
- The Information Commissioner's Office (ICO) makes sure the rules are followed and can issue large fines if an organisation fails to comply.
- GDPR recognises the importance of scientific research, so enables researchers to process personal data as long as this is necessary for research purposes and appropriate safeguards are in place.

Page 7: Safeguards

You will now be shown a short animation about the safeguards around smart data. If you have any questions, you can write them below.

- Five Safes Framework: *Safe Data, Safe Projects, Safe People*, *Safe Settings, Safe Outputs.*
- Safe Data: Before data is used for research, information about your identity is removed from the dataset – this is called de-identification.
- De-identified data may still be personal data. It might be possible to identify an individual person within a de-identified dataset. Other safeguards (from the Five Safes Framework need to be in place.

Page 8: The power of research

Paul Monks, Chief Scientific Advisor at the Department for Energy Security and Net Zero. If you have any questions, you can write them below.

- Chief Scientists make sure that government uses the best science available to make policy.
- Government uses data in lots of ways. From measuring carbon emissions to understanding the spread of Covid at sporting events.
- Data is powerful because it allows us to understand people and how they respond to policy.

Page 9: The power of research

Dr Rachel Oldroyd, Consumer Data Research Centre at The University of Leeds

Project: Priority Places for Food Index

- The Priority Places for Food Index (PPFI) identifies neighbourhoods in the UK most at risk of food insecurity
- People living in priority areas may be vulnerable to increases in the cost of living and may lack access to affordable, healthy, and reliable sources of food.
- The Priority Places for Food Index dashboard is being used by policy makers, charities, retailers, nutritionists, and community groups in their efforts to tackle causes of food insecurity

Pages 10-14: Case studies (case studies can be seen below)

Page 15: Summary of presentations

- ‘Smart data’ is a general term we use to cover a wide range of data generated as a byproduct of our digital lives.
- There are many different possibilities for researchers using smart data, for example, using shopping data to understand health challenges, understanding regional differences, and supporting urban development.
- Not all smart data is personal data, but when it is first generated by people, it probably is.
- In the UK we have strict laws that govern the way personal data can be collected and used. This includes the General Data Protection Regulation (GDPR).
- GDPR recognises the importance of scientific research, so enables researchers to process personal data as long as this is necessary for research purposes and appropriate safeguards are in place.
- Safeguards include de-identification. Before data is used for research, information about your identity is removed from the dataset.
- Government uses data in lots of ways. Data is powerful because it allows us to understand people and how they respond to policy.

Page 16: On our street (case studies for ‘On our street’ can be found in workshop 5 materials)

Page 17: The SDR eco-system

- The public are who we want to benefit from smart data research. SDRUK need to engage with them to make sure they are using your data well and meeting your expectations.
- The government develops laws and delivers public services. They want to use data and research to make better decisions that affect us all.
- Scientists and researchers have different specialisms. They have lots of different research questions and want to access smart data to answer them. Some research could lead to breakthroughs that solve big social challenges - but not all research will lead to solutions.
- Smart Data Research UK Programme. Data services are teams of scientists and other researchers at UK universities. They work with data controllers to bring data into a research environment and provide secure settings for researchers to access sensitive data. SDRUK is a central hub that provides funding, coordinates efforts, and sets the direction.
- SDRUK are funded by UK Research and Innovation (UKRI), which is funded by the government. They are publicly funded, from taxpayers, and not here to make a profit.
- Data controllers include many private companies, but it can also include publicly funded organisations and charities. They use data for their own purposes and hold the keys to providing access for researchers.

Page 18: what we are doing together

Workshop 1:

- Get to know one another
- Find out about SDR UK and discuss how smart data is used

Workshops 2-4:

- Discuss the challenges with specialists and each other

Workshop 5:

- Tell us how you think SDR UK should decide what matters most

Page 19: Thank you!

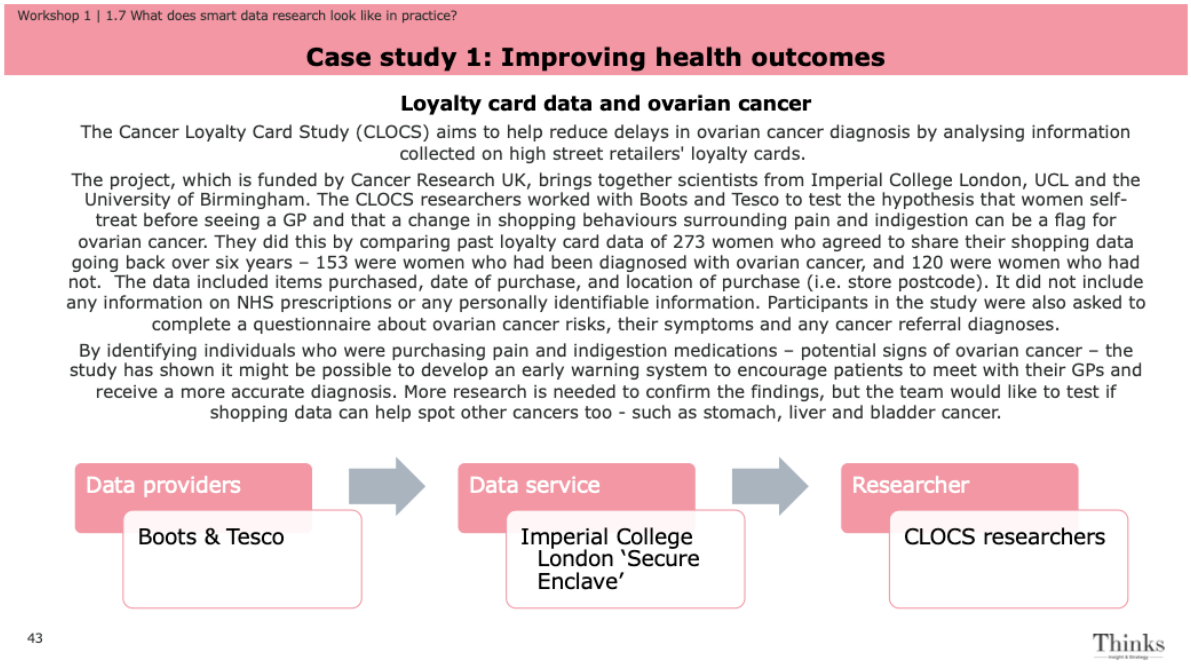
Case study 1: Improving health outcomes

Loyalty card data and ovarian cancer

The Cancer Loyalty Card Study (CLOCS) aims to help reduce delays in ovarian cancer diagnosis by analysing information collected on high street retailers' loyalty cards.

The project, which is funded by Cancer Research UK, brings together scientists from Imperial College London, UCL and the University of Birmingham. The CLOCS researchers worked with Boots and Tesco to test the hypothesis that women self-treat before seeing a GP and that a change in shopping behaviours surrounding pain and indigestion can be a flag for ovarian cancer. They did this by comparing past loyalty card data of 273 women who agreed to share their shopping data going back over six years – 153 were women who had been diagnosed with ovarian cancer, and 120 were women who had not. The data included items purchased, date of purchase, and location of purchase (i.e. store postcode). It did not include any information on NHS prescriptions or any personally identifiable information. Participants in the study were also asked to complete a questionnaire about ovarian cancer risks, their symptoms and any cancer referral diagnoses.

By identifying individuals who were purchasing pain and indigestion medications – potential signs of ovarian cancer – the study has shown it might be possible to develop an early warning system to encourage patients to meet with their GPs and receive a more accurate diagnosis. More research is needed to confirm the findings, but the team would like to test if shopping data can help spot other cancers too - such as stomach, liver and bladder cancer.



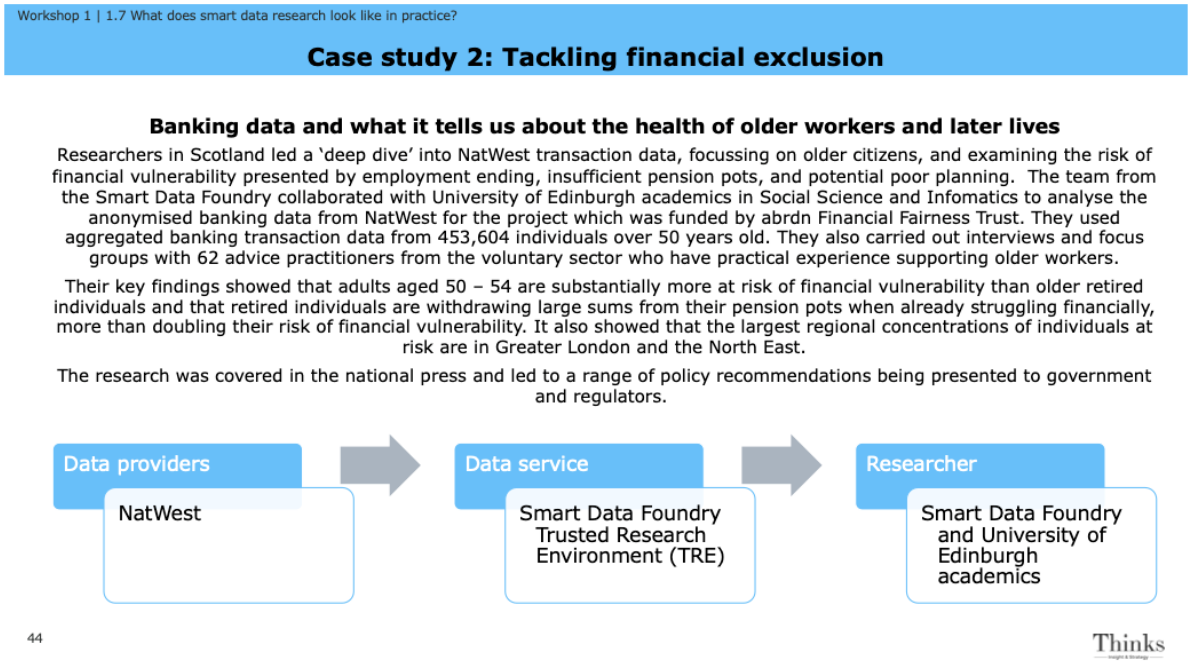
Case study 2: Tackling financial exclusion

Banking data and what it tells us about the health of older workers and later lives

Researchers in Scotland led a 'deep dive' into NatWest transaction data, focussing on older citizens, and examining the risk of financial vulnerability presented by employment ending, insufficient pension pots, and potential poor planning. The team from the Smart Data Foundry collaborated with University of Edinburgh academics in Social Science and Infomatics to analyse the anonymised banking data from NatWest for the project which was funded by abrdn Financial Fairness Trust. They used aggregated banking transaction data from 453,604 individuals over 50 years old. They also carried out interviews and focus groups with 62 advice practitioners from the voluntary sector who have practical experience supporting older workers.

Their key findings showed that adults aged 50 – 54 are substantially more at risk of financial vulnerability than older retired individuals and that retired individuals are withdrawing large sums from their pension pots when already struggling financially, more than doubling their risk of financial vulnerability. It also showed that the largest regional concentrations of individuals at risk are in Greater London and the North East.

The research was covered in the national press and led to a range of policy recommendations being presented to government and regulators.



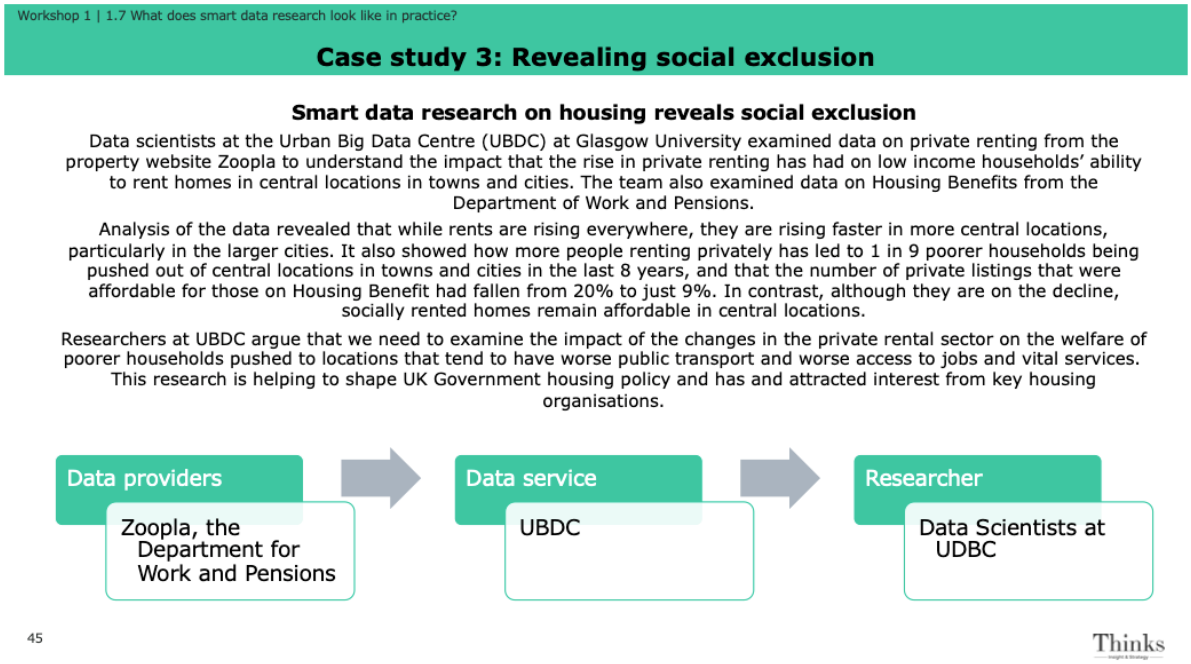
Case study 3: Revealing social exclusion

Smart data research on housing reveals social exclusion

Data scientists at the Urban Big Data Centre (UBDC) at Glasgow University examined data on private renting from the property website Zoopla to understand the impact that the rise in private renting has had on low income households' ability to rent homes in central locations in towns and cities. The team also examined data on Housing Benefits from the Department of Work and Pensions.

Analysis of the data revealed that while rents are rising everywhere, they are rising faster in more central locations, particularly in the larger cities. It also showed how more people renting privately has led to 1 in 9 poorer households being pushed out of central locations in towns and cities in the last 8 years, and that the number of private listings that were affordable for those on Housing Benefit had fallen from 20% to just 9%. In contrast, although they are on the decline, socially rented homes remain affordable in central locations.

Researchers at UBDC argue that we need to examine the impact of the changes in the private rental sector on the welfare of poorer households pushed to locations that tend to have worse public transport and worse access to jobs and vital services. This research is helping to shape UK Government housing policy and has and attracted interest from key housing organisations.



Case study 4: Using location data to plan better infrastructure

Strava data and what it can tell us about town planning impacts on infrastructure

At the Urban Big Data Centre (UBDC) at Glasgow University, researchers have worked extensively with data from Strava, a fitness app that allows users to track a variety of sporting activities, including logging and sharing their cycle journeys.

The data from the app enables town planners to understand which routes people use, and which destinations are busiest. This can help academics and policymakers track how popular cycling is over time, evaluate impacts of infrastructure, and plan new interventions to encourage active travel.

For example, many local authorities have invested a lot of time and money in safe cycling infrastructure to improve cycling environments. But it is not clear whether these investments [are] actually encouraging people to cycle more in cities where there is a high level of rain. By analysing Strava data in Glasgow, researchers found that providing safe cycle paths could encourage people to cycle more, especially on dry days, but cycling decreased on rainy days. This suggests that when planning for rainy locations, other policies may be needed to increase cyclists' resilience to bad weather (e.g., providing shower facilities at workplaces, incentives to cycle, etc.).

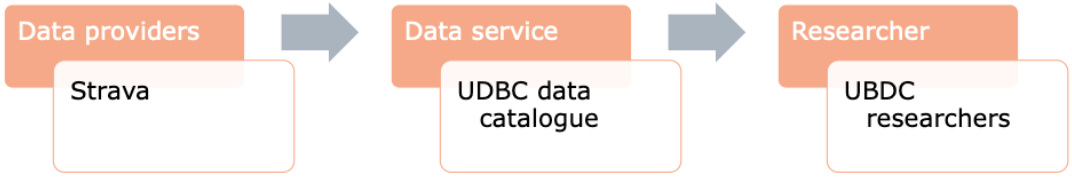
Case study 4: Using location data to plan better infrastructure

Strava data and what it can tell us about town planning impacts on infrastructure

At the Urban Big Data Centre (UBDC) at Glasgow University, researchers have worked extensively with data from Strava, a fitness app that allows users to track a variety of sporting activities, including logging and sharing their cycle journeys.

The data from the app enables town planners to understand which routes people use, and which destinations are busiest. This can help academics and policymakers track how popular cycling is over time, evaluate impacts of infrastructure, and plan new interventions to encourage active travel.

For example, many local authorities have invested a lot of time and money in safe cycling infrastructure to improve cycling environments. But it is not clear whether these investments [are] actually encouraging people to cycle more in cities where there is a high level of rain. By analysing Strava data in Glasgow, researchers found that providing safe cycle paths could encourage people to cycle more, especially on dry days, but cycling decreased on rainy days. This suggests that when planning for rainy locations, other policies may be needed to increase cyclists' resilience to bad weather (e.g., providing shower facilities at workplaces, incentives to cycle, etc.).



Workshop 2: Discussion Guide and Materials

Workshop objectives:

The specific objectives for this workshop are to:

- Understand spontaneous views of what defines "research for the public good" and priorities for future research.
- Introduce and interrogate SDR UK's definition of public good and compare and contrast with participant definition.
- Introduce key potential harms related to public good, including a speaker on exclusion
- Introduce principles from other data services on 'why' researchers can access data.
- Discuss rules participants would put in place to ensure SDR works for the public good?

Workshop 2 guide (online) | September 2024		
Section and aim	**Key questions and probes**	**Time**
Arrival	**Participants arriving** • *Participants will be asked to arrive from 5:45pm* • *Sign people in and check user names* • *Address tech issues*	-

2.1 Welcome and introduction to the session	*Lead facilitator to welcome all participants*:
	• *Recap purpose of dialogue*
	• *Introduce people on the call (Thinks team, SDR UK / Sciencewise attendees and experts)*
	• *Reiterate ground rules for participating in the sessions*
	Lead facilitator to introduce the session:
	• *Purpose of the session*
	• *Agenda for the day*
	• *Recap of key points from Workshop 1, including:*
	  ○ *What is smart data?*
	  ○ *What is smart data research?*
	  ○ *How does data de-identification work? [new slide]*
		5

2.2 Warm-up and headline recap **Aim**: Get participants' feedback on the ongoing analysis process.	*Participants to be split into breakout groups.* *Facilitator to introduce self and ask participants to recap introductions:* • *Name* • *One exciting thing that's happened since we last met* In today's session we will continue to discuss smart data research, looking at examples of how it can be used in real life, and thinking about some of the potential harms involved. But first, we're going to continue to reflect a bit on the first face-to-face session. *Facilitator to share screen with headline findings, noting down areas of agreement / disagreement for each as they are raised.* **What do you think about this?** • To what extent, if at all, do you agree with this? • What, if anything, would you add or change about this? • Is there anything big that has been missed out? • Is there anything you're surprised to see?	10
2.3 What is research for the public good? Understanding participant's spontaneous views on what defines research for the public good	Now that we have recapped what we learnt in Session 1, I would like us to discuss research for the public good and what this means. **When you hear the term 'research for the public good', what does it mean to you?** • What makes research for 'the public good'? • What types of benefits would you expect from smart data research that claims to be in the public good? *Facilitator to work with participants to establish hierarchy of public good on slide, ranking different considerations as they are explored.* **Do you think there are different levels of public good?** • If so, what are they? If needed, facilitator to probe on: • Number of people impacted • Scale of impact	30

	• Equality of impact • Issue addressed • Impact on specific organisations/bodies **Who are the 'public' in public good?** *Moderator to probe present vs. future, UK vs. international, individuals in their personal life vs. professional* **How do we define 'good' in the context 'public good'?** *Moderator to probe on the importance of actual vs potential impact, who decides what constitutes as for the public good? Can risky projects still be in the public good?* **What are the most important outcomes you think 'research for the public good' should achieve?** **Who delivers public good?** • What types of organisations might deliver public good? *Explore spontaneous responses before probing: government, private companies, charities, others?* We're now going to think more about where you think smart data research has potential for the biggest impact on the public. **What areas do you think are the most important for smart data research to focus on for public good? Why?** • What makes these areas important? **Which areas do you think will have the most impact?** • Who do you think will be impacted and how? *Moderators to probe on specific groups, personal impact, and impact on society as a whole* **Is there anything you think might slow down or prevent smart data research on these issues from being successful (and having a positive impact on people's lives)?** • Are there any barriers to the use of smart data in this way? • What would prevent any impacts from being realised?	

	How, if at all, did this conversation change how you feel about the use of your data for smart data research? **Is there anything that made you more optimistic about smart data research?** *Reconvene in plenary to feedback to group:* In locations, lead facilitator to ask participants from each table to summarise and share what research for the public good means to them.	
2.4 How is SDR UK thinking about research for the public good? **Aim:** Introducing participants to SDR UK definition for the public good	*Expert presentation in plenary, via Zoom, to all locations:* SDR UK have been doing a lot of thinking about how smart data research can be used for the public good. They've grouped the types of research they want to encourage, enable and promote into 4 pillars. We're going to hear a bit more about these now: ***Presentation from SDR UK representative to explain the four thematic pillars and why they have been chosen:*** • *Productivity and prosperity for all* • *Health and wellbeing* • *Digital society* • *Sustainability* *Back in breakout tables for discussion:* We are now going to think about what research for public good in these areas could look like in practice. During this part of the session, each group will come up with a list of priorities for smart data research based on one thematic pillar, using a series of prompts to refine them. We'll all work into one document so we can come together at the end to discuss and build on where each group got to. *Facilitator to record ideas on first thematic value slide in shared deck, using digital footprint maps to prompt as needed:* **What are some issues relating to this pillar that you think research using smart data could investigate?** **What kind of smart data do you think could be used to address these issues?** • Where would this smart data come from?	25

	What benefit could this research have? On society as a whole? On specific types of people? *Notetaker to merge all groups' ideas onto a single slide to discuss and refine. Plenary discussion of participant examples of research for the public good:* **What are your initial thoughts about this list of priorities?** • Is there anything you find surprising? Why? **Is there anything missing from another group's list that you'd like to add?** • Why? **Looking at the list we've made, is there anything two or more of these issues have in common?** • What makes these issues important to address? **What impact should / would prioritising these issues for smart data research have?** *Moderator to prompt if needed:* • *On you personally?* • *On people you know?* • *On society as a whole?*	
BREAK	7:05– 7:15pm	10
2.5 Who might be left behind? **Aim**: To help participants understand the key ethical questions around who smart data research might exclude and what the impact would be	*Participants to return to plenary for expert testimonial.* We are now going to hear from Maxine Mackintosh who will tell us about how smart data research might exclude certain types of people if not managed correctly. *Expert to introduce themselves and their role/organisation, before covering:* • *Why might smart data research exclude people?* • *The types of people who might be left behind* • *The impact that this would have of these groups of people/society at large* • *Any mitigations that can help protect against this*	15

	• *[If they would like] the benefits of SDR for these groups if managed properly and conducted ethically* *Q&A* *In breakout groups:* **What word would you use to sum up how you feel about what you've just learned?** *Moderator to probe on reasons behind words chosen.* • Why did you choose that word? • What is surprising about what you have heard from the experts? **Having heard this, what impact do you think smart data research could have:** • *On society?* • *On you personally?* • *On other people you know?* **How would you feel about your data being used in this way?** **Do you have any more hopes about smart data research after hearing this?** **Do you have any more concerns about smart data research after hearing this?** **Do you still have any questions about what you've learned?** *Moderator to answer questions they can and collect the rest to feedback to SDR UK for potential clarification.*	
2.6 Potential harms of smart data research **Aim**: To bring to life the potential harms associated with smart data research if proper governance measures aren't in place.	*Back in break-out groups, introduction of case studies, outlining what potential harms of SDR could look like if realised.* • *Case study 1: [Fictional case study showcasing where data is used for a purpose other than intended]* • *Case Study 2: [Fictional case study showcasing where policy decisions that impacted a group were made without their data being used in the research that informed it]* • *Case study 3: [Fictional case study where a groups data is used in research and the findings are used to either disparage them or inform policy that disadvantages them]*	20

	• *Case study 4: [Fictional case study where data is shared and private company benefits upfront but research doesn't deliver for the public in the way that was expected/promised]* We're now going to look at a few example case studies that highlight some potential harms associated with smart data research. *Each group to look at 2 case studies in turn, covering all harms in each location. Facilitator to share screen and read out case study before asking:* **What are your initial responses to this example?** **In this example, who is sharing the data?** • Who is using the data? • What are they both trying to achieve? **What do you think about what has happened in this scenario?** **Who, if anyone, has benefitted in this scenario? How?** **Who, if anyone, has been disadvantaged in this scenario? How?** **How concerned, if at all, are you about this happening?** • How concerned, if at all, are you that this could happen to someone like you? **Can you think of any way that this harm could have been prevented?** • What could the private company / researchers / SDR UK have done differently?	
2.7 Principles for deciding why a researcher can access smart data **Aim:** Stretch participants' understanding of the rules that could be	As you know, SDR UK oversees the data services that provide data to researchers. The data services are responsible for ensuring that consumer data is safe and used only for ethical reasons, and our opinions can help feed into how SDR UK encourages and requires data services to do this. We are not the first people to be thinking about principles around data sharing. Data services follow rules (e.g. the 5 Safes, CDRC's access criteria, etc.) We are going to talk in more detail about most of these rules in workshop 4 in a few weeks. But today, as we are discussing the types of research we'd like	10

in place to ensure SDR works for public good	to see smart data used for, it makes sense to discuss the rules that other services have in place for 'why' researchers should be allowed to access data. We've pulled together some examples of the types of rules and conditions that other data services use. Some of these were developed via public engagement like this one, others are based on the views of experts. They also relate to different types of data, some apply to very sensitive data, others to open data or statistics (that's data anyone can access). **Let's review these examples for a few minutes.** *Prompt questions for each example and remind participants of smart data definition if any confusion.* **How relevant do these feel to smart data?** • Are some more relevant than others? • Are the some that don't apply at all? • Is there anything missing here that needs to be in place for smart data? • What are the absolute most important points for SDRUK to consider?	
2.8 What should SDR UK do? Laying the groundwork for final principles of public good	We are now going to spend some time thinking about the things SDR UK should encourage and discourage and if there are any principles you think should be in place around smart data research. For the rest of this session, we'd like to focus on one key question: If you were in charge, what rules would you put in place to ensure smart data research works for the public good? We will do this by coming up with our own principles around who should benefit and how. *Each group to start come up with a list of principles on a slide. <u>If needed the moderator can probe on</u>:* **What kinds of rules and behaviours should data services be implementing?** **What kinds of rules and behaviours should researchers be implementing?** **What should SDR UK be doing to encourage these?** **Thinking about some of the concerns you've mentioned or potential harms we've discussed throughout this research, what rules/principles would you like to see to address these?**	20

	What kind of rules would you want to see around: • Safeguards or protections that should be put in place to ensure that smart data is used for the public good • Ensuring smart data research is inclusive and considers the needs of all members of the public • The review process to make sure smart data research projects are working towards the public good • The criteria used to evaluate whether a smart data research project is successful in serving the public good • People who think about the ethics and principles of "good" research often use the term "research with integrity". Research that has integrity means conducting research with honesty, transparency, open communication, and respect, for a positive and accountable research environment. • What do you think this would mean in practice? • How does this align with the rules we've suggested? Is there anything we might want to add or change to support "research with integrity"? **If these rules were in place, would you feel confident about smart data being used for research?** What makes you say that? *Probe on what influences trust – e.g. who accesses data, what it is used for, how it is stored.* • What, if anything, is missing? • What are the absolute most important points for SDR UK to consider?	
2.9 Wrap-up and close **Aim:** To explain next steps.	*Lead facilitator to lead plenary session, with facilitator from each group feeding back their key out-takes, including any outstanding questions.* *Finally, lead facilitator to:* • *Thank participants for their time today* • *Remind participants of payment details* • *Confirm next steps*	5

Stimulus

The materials used in workshop 2 include:

- Running slides, expert speakers and case studies. These are available upon request.

Workshop two included a presentation from Dr. Maxine Mackintosh, AI and Health Consultant from the Alan Turing Institute.

The purpose of this presentation was to introduce the concept of data bias to participants including:

- Where bias in data-driven research comes from (i.e. the questions we pose, the data we use, how we design studies, the approaches we use, how we monitor impact and the broader context),
- What might the data bias be including examples of data sources such as smart watches, clinical trials and club cards.

Workshop two also included a presentation from Andy Morris, Chief Data Officer from Boots.

The purpose of this presentation was to discuss how private companies might be involved in Smart Data Research. Andy Morris outlined previous research that Boots has contributed towards including:

- The type of smart data collected at Boots,
- The societal benefits of the research Boots provided smart data for,
- Why Boots chose to contribute their smart data,
- A Q&A.

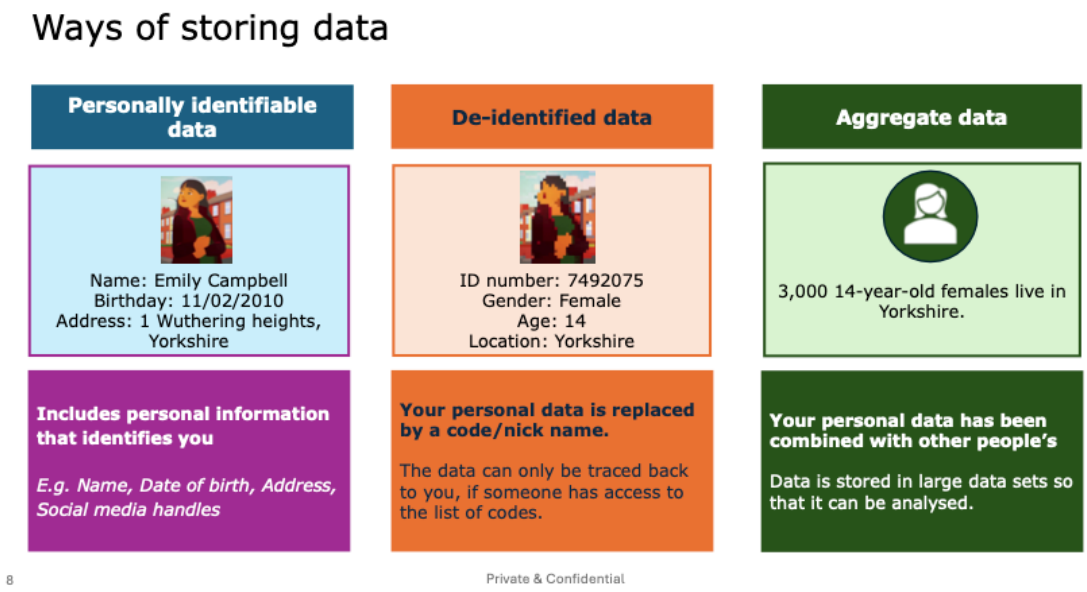
Information we shared with participants about ways of storing data

Personally identifiable data, includes personal information that identified you:

- Name
- Date of birth
- Address
- Social Media Handles

De-identified data, your personal data is replaced by a code/nick name. The data can only be traced back to you, if someone has access to the list of codes.

Aggregate data, your personal data has been combined with other people's. Data is stored in large data sets so that it can be analysed.



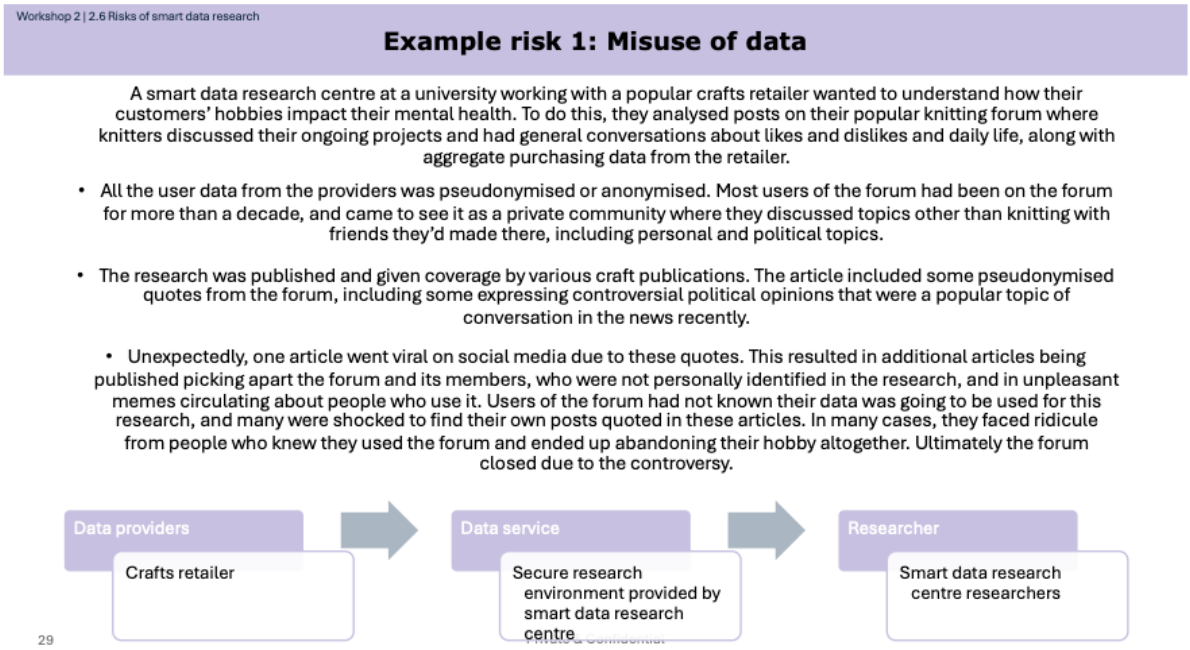
What we told participants about the smart data research system

- Data controllers include many private companies, but it can also include publicly funded organisations and charities. They use data for their own purposes and hold the keys to providing access for researchers.
- Smart Data Research UK Programme. Data services are teams of scientists and other researchers at UK universities. They work with data controllers to bring data into a research environment and provide secure settings for researchers to access sensitive data. SDRUK is a central hub that provides funding, coordinates efforts, and sets the direction.
- SDRUK are funded by **UK Research and Innovation (UKRI)**, which is funded by the government. They are publicly funded, from taxpayers, and not here to make a profit.
- **Scientists and researchers** have different specialisms. They have lots of different research questions and want to access smart data to answer them. Some research could lead to breakthroughs that solve big social challenges - but not all research will lead to solutions.
- **The government** develops laws and delivers public services. They want to use data and research to make better decisions that affect us all.
- **The public** are who we want to benefit from smart data research. SDRUK need to engage with them to make sure they are using your data well and meeting your expectations.

Case studies

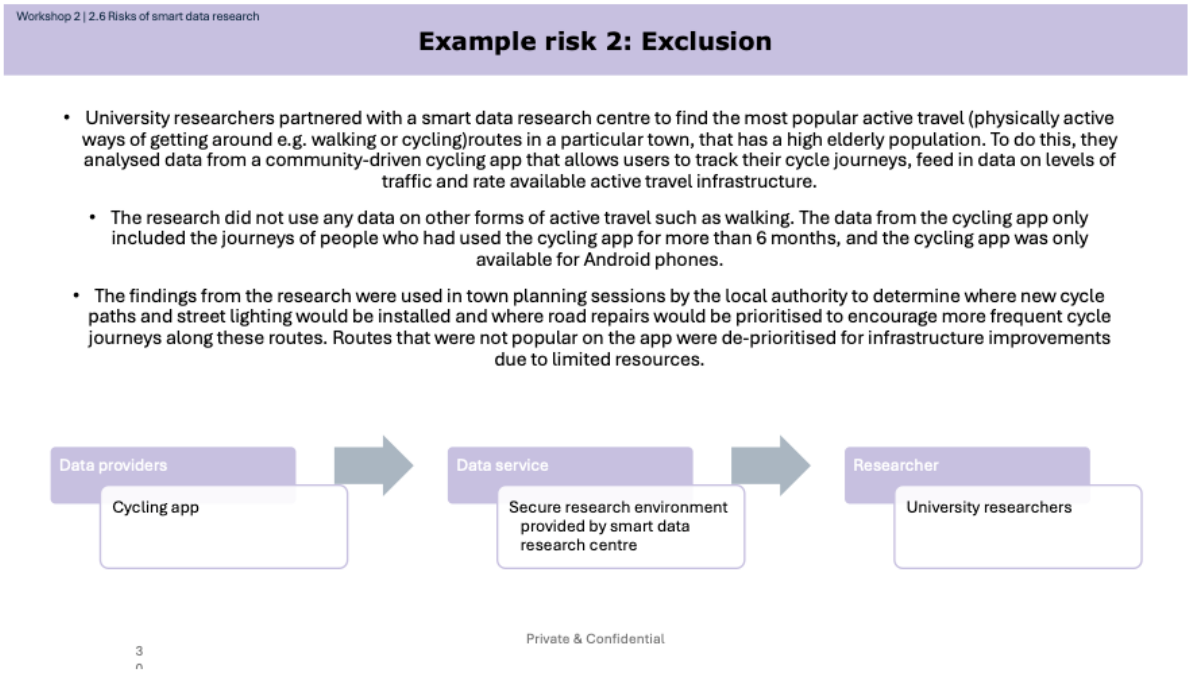
Example 1: The misuse of data

- A smart data research centre at a university working with a popular crafts retailer wanted to understand how their customers' hobbies impact their mental health. To do this, they analysed posts on their popular knitting forum where knitters discussed their ongoing projects and had general conversations about likes and dislikes and daily life, along with aggregate purchasing data from the retailer.
- All the user data from the providers was pseudonymised or anonymised. Most users of the forum had been on the forum for more than a decade, and came to see it as a private community where they discussed topics other than knitting with friends they'd made there, including personal and political topics.
- The research was published and given coverage by various craft publications. The article included some pseudonymised quotes from the forum, including some expressing controversial political opinions that were a popular topic of conversation in the news recently.
- Unexpectedly, one article went viral on social media due to these quotes. This resulted in additional articles being published picking apart the forum and its members, who were not personally identified in the research, and in unpleasant memes circulating about people who use it. Users of the forum had not known their data was going to be used for this research, and many were shocked to find their own posts quoted in these articles. In many cases, they faced ridicule from people who knew they used the forum and ended up abandoning their hobby altogether. Ultimately the forum closed due to the controversy.



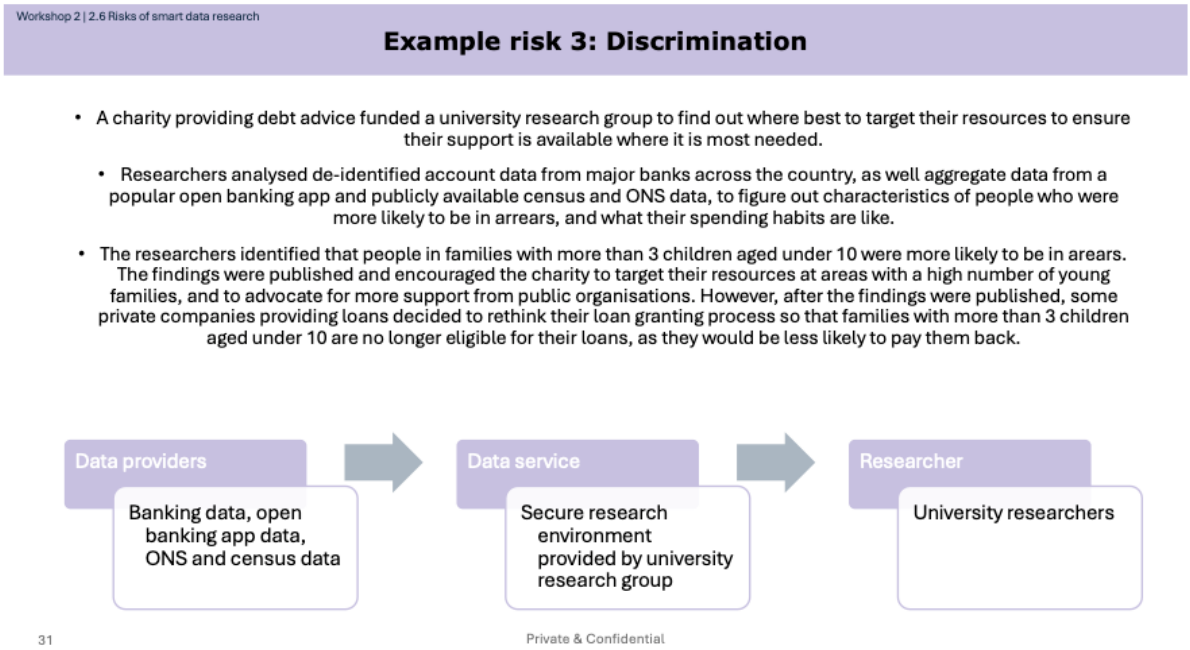
Example 2: Exclusion

- University researchers partnered with a smart data research centre to find the most popular active travel (physically active ways of getting around e.g. walking or cycling)routes in a particular town, that has a high elderly population. To do this, they analysed data from a community-driven cycling app that allows users to track their cycle journeys, feed in data on levels of traffic and rate available active travel infrastructure.
- The research did not use any data on other forms of active travel such as walking. The data from the cycling app only included the journeys of people who had used the cycling app for more than 6 months, and the cycling app was only available for Android phones.
- The findings from the research were used in town planning sessions by the local authority to determine where new cycle paths and street lighting would be installed and where road repairs would be prioritised to encourage more frequent cycle journeys along these routes. Routes that were not popular on the app were de-prioritised for infrastructure improvements due to limited resources.



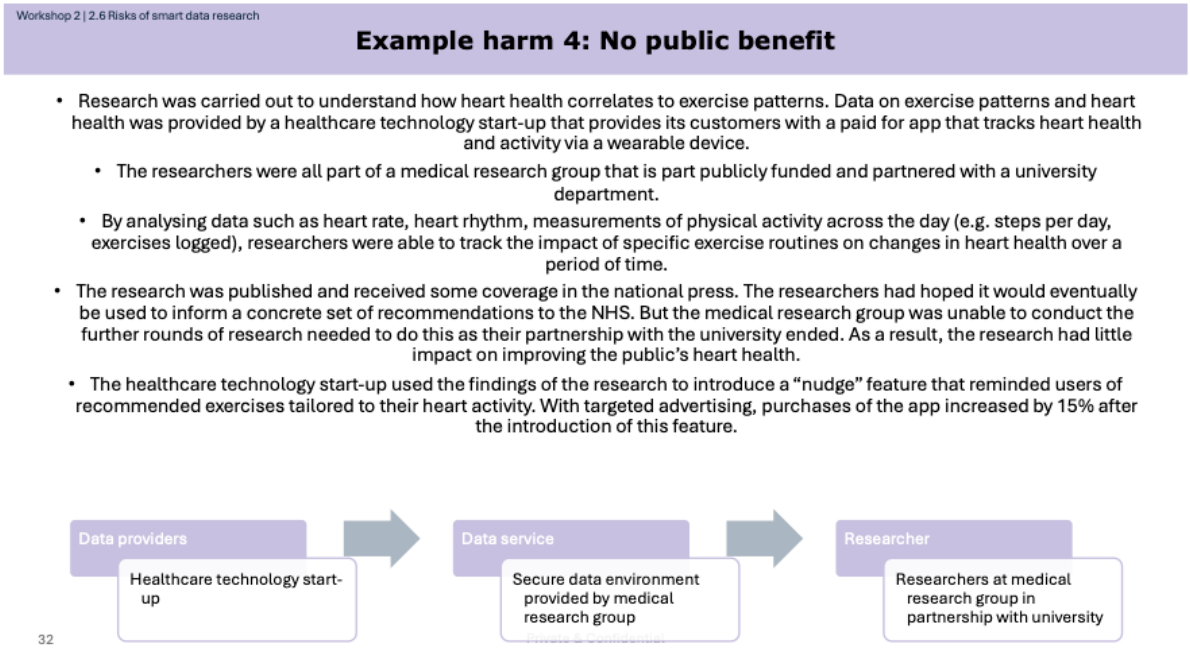
Example 3: Discrimination

- A charity providing debt advice funded a university research group to find out where best to target their resources to ensure their support is available where it is most needed.
- Researchers analysed de-identified account data from major banks across the country, as well aggregate data from a popular open banking app and publicly available census and ONS data, to figure out characteristics of people who were more likely to be in arrears, and what their spending habits are like.
- The researchers identified that people in families with more than 3 children aged under 10 were more likely to be in arears. The findings were published and encouraged the charity to target their resources at areas with a high number of young families, and to advocate for more support from public organisations. However, after the findings were published, some private companies providing loans decided to rethink their loan granting process so that families with more than 3 children aged under 10 are no longer eligible for their loans, as they would be less likely to pay them back.



Example 4: No public benefit

- Research was carried out to understand how heart health correlates to exercise patterns. Data on exercise patterns and heart health was provided by a healthcare technology start-up that provides its customers with a paid for app that tracks heart health and activity via a wearable device.
- The researchers were all part of a medical research group that is part publicly funded and partnered with a university department.
- By analysing data such as heart rate, heart rhythm, measurements of physical activity across the day (e.g. steps per day, exercises logged), researchers were able to track the impact of specific exercise routines on changes in heart health over a period of time.
- The research was published and received some coverage in the national press. The researchers had hoped it would eventually be used to inform a concrete set of recommendations to the NHS. But the medical research group was unable to conduct the further rounds of research needed to do this as their partnership with the university ended. As a result, the research had little impact on improving the public's heart health.
- The healthcare technology start-up used the findings of the research to introduce a "nudge" feature that reminded users of recommended exercises tailored to their heart activity. With targeted advertising, purchases of the app increased by 15% after the introduction of this feature.



Why should researchers be able to use smart data

Examples of existing rules for other data projects:

- For scientific advancement, e.g. it develops new knowledge, understanding or techniques
- It will generate knowledge that can benefit society
- If they have a clear plan to publish their findings and methodology in a way that allows for scrutiny and further research
- If they plan to share the data and findings in a designated data archive (e.g. the UK data service), where other approved researchers can access and learn from the information.
- If the project seems likely to work, e.g. it uses an appropriate methodology have the right credentials to implement it
- If the data will be used for the same intended purpose as when it was initially shared (i.e. a new research team using the data for the same purpose as the team who initially got permission to use the data)
- If the researcher has approval from an ethics panel at their university to access this data

Workshop 3: Discussion Guide and Materials

Workshop objectives:

The specific objectives for this workshop are to:

- Explore spontaneous views on drivers and barriers for private companies to become involved in SDR.
- Hear from speaker from the private sector about drivers and barriers to become involved and insight into possible commercial arrangements that could be in place.
- Exploring potential harms related to private company sharing of smart data
- Discuss rules participants would put in place to ensure commercial relationships are fair

Workshop 3 guide (online) | July 2024		
Section and aim	**Key questions and probes**	**Time**
Arrival	**Participants arriving** *Participants will be asked to arrive from 5:45pm* *Sign people in and check user names* *Address tech issues*	-
3.1 Welcome, introduction to the dialogue and introduction to the session	*Lead facilitator to welcome all participants*: • *Recap purpose of dialogue* • *Introduce people on the call (Thinks team, SDR UK / Sciencewise attendees and experts)* • *Reiterate ground rules for participating in the sessions* • *Lead facilitator to introduce the session:* • *Purpose of the session* • *Agenda for the day* • *Recap of SDR UK organisational structure and purpose* *Lead facilitator to hand over to the break-out facilitators for introductions and in-depth discussion*	5
3.2 Warm-up and headline recap **Aim**: Get participants' feedback on the ongoing	*Participants to be split into breakout groups.* *Facilitator to introduce self and ask participants to recap introductions:* • *Name* • *One exciting thing that's happened since we last met*	10

analysis process.	The purpose of today is to delve a bit deeper into the different actors involved in smart data research, their motivations for being involved and the benefits they might be looking to gain from participating. This will help us think about the balance between the benefits for private company of sharing smart data and the public good. Before we get started, I would like to go through some of the key concepts that we have explored so far to check how you feel about them and whether there's anything you'd like to add. *Facilitator to share screen with headline findings, noting down areas of agreement / disagreement for each as they are raised.* **What do you think about this?** • To what extent, if at all, do you agree with this? • What, if anything, would you add or change about this? • Is there anything big that has been missed out? • Is there anything you're surprised to see?	
3.3 Spontaneous exploration: private benefits **Aim:** Understand participants' views on the extent to which they feel that private interests and public good can go hand in hand.	I want talk about some of your own expectations for private involvement in smart data research, like the sorts of research projects we've looked at in workshop 1. *Facilitator to share short summaries of the case studies explored in workshop 1, with each group revisiting the same case studies they discussed in that session.* **Why do you think private companies might want to share the data they collect with researchers?** • What benefit do you think they will get from this kind of arrangement? • What worries you about these arrangements? **What conditions do you expect private companies might set on researchers before they share their smart data?** *If needed, facilitator to provide examples e.g. type of research data is used for, financial compensation for data, rules around sharing of findings and publicising research.*	20

	• Do you think these are fair?	
3.4 Information sharing: bringing to life the benefits for private companies and public of SDR **Aim:** Explore the specific motivations and benefits of SDR for private companies and researchers and the impact this has on public good.	*Participants to return to plenary for expert testimonial* *One (or two) speakers to represent private interests, covering considerations for companies and potential benefits of SDR for them and potential trade-offs.* *Q&A* *In breakout groups (speakers to join breakout rooms to answer questions):* **I hope those speakers were interesting and informative. Can everyone please share one word in the chat section summarising how they feel after that discussion?** • Why did you choose that word? • What is surprising about what you have heard from the experts? **Reflecting on what you heard, how do you feel now about the smart data you share with private companies being used for research for the public good?** **How would you feel about researchers paying for data from private companies for research that benefits the public good?** • Do you think it is / isn't reasonable? • Do you feel the same about all types of data or not? • Is paying for some types of data more / less acceptable than paying for others? Why / why not? • What other things do you think researchers or data services can offer to companies in return for the data they hold? • Is there any situation in which you think private companies should have to share smart data with researchers? **How, if at all, did what you heard impact some of your hopes and fears for private/public collaboration when it comes to smart data for research?** **Do you still have any questions about what you've learned?** *Moderator to answer questions*	35

	they can and collect the rest to feedback to SDR UK for potential clarification.	
BREAK	7:20– 7:25pm	5
3.5 Key risks associated with using data from private companies **Aim:** Explore the ways in which private company sharing of data could go wrong and the potential harm this could cause the public and society.	Welcome back everyone. Before the break, you heard a bit more about the smart data you share with private companies being used for research for the public good. We will now split into our breakout groups to talk about the risks (and potential harms) associated with using data provided by private companies. Once you have discussed this in your groups, we will come back together and hear the main concerns from each group. *In breakout groups:* **What are your main concerns around your smart data being shared by private companies?** *Moderators to listen out for any spontaneous mentions of consent.* What could be done to alleviate these concerns? Moderator to share 3 risks on screen Here we have listed 3 risks of private companies sharing smart data: • The public don't feel that they have consented to their data being used for research • Private companies don't get involved because it is seen to be too complicated, missing an opportunity to conduct research that could have a meaningful impact on public good • The expense of getting data from a private company outweighs the potential public benefit of the research. **From a first glance, which of these potential risks are you the most concerned about? Why?** *If not previously mentioned:* Did any of these come to mind earlier? Thinking about consent… **What are your thoughts about members of the public feeling they haven't consented to their data being used?**	45

	• What might this look like in practice? • What harm, if any, might this cause? To whom? • How big of a risk do you think this is? • Do you have any concerns about this? How much of an issue is this for you personally? • What, if anything, do you think could be done to minimise this risk? • What assurances or information or about smart data research do you think would minimise this risk? Thinking about lack of private company involvement… **What are your thoughts about private companies deciding not to get involved in smart data research?** • What might this look like in practice? • What harm, if any, might this cause? To whom? • How big of a risk do you think it is? • Do you have any concerns about this? How much of an issue is this for you personally? • What, if anything, do you think could be done to minimise this risk? • What incentives or assurances do you think could help private companies share their data? Thinking about the risk that the benefit to the public of research doesn't justify the expense… **What are your thoughts about this risk?** • What might this look like in practice? • What harm, if any, might this cause? To whom? • How big of a risk do you think it is? • Do you have any concerns about this? How much of an issue is this for you personally? • What, if anything, do you think could be done to minimise this risk? • What factors should be considered when deciding whether the cost of obtaining data is too high?	

	Thinking about everything we have spoken about… **What do you see at the biggest risk when it comes to private companies providing data for smart data research?** *Moderator to note down to feedback to the plenary session* • What makes this the biggest risk? • What steps can be taken to mitigate these risks? *Participants return to plenary for the moderator from each breakout group to feedback their groups' principal concern relating to private companies sharing data*	
3.6 What should SDR UK do? **Aim:** Laying the groundwork for final principles on relationships with private companies in workshop 5	Now that we have heard everyone's concerns around the potential risks of private companies sharing smart data, we are now going to spend some time bringing together everything we've discussed so far about how we can ensure partnerships with private companies involved in smart data research are fair. Specifically, we are going to think about the things SDR UK should encourage and discourage and any principles you think should be in place around public-private collaboration. For the rest of this session, we'd like to focus on one key question: • If you were in charge, what rules would you put in place to ensure commercial relationships are fair? We will do this by coming up with our own principles around how data is shared between private companies and smart data researchers, and how the risks we've discussed can be addressed. *Break-out group facilitators to share screen. Each group to start come up with a list of principles on a slide. If needed the moderator can probe on:* *In breakout groups:* • **What kinds of rules and behaviours should private companies be implementing?** • **What kinds of rules and behaviours should data services be implementing?** • **What kinds of rules and behaviours should researchers be implementing?**	20

	• **What should SDR UK be doing to encourage these?** • **Thinking about some of the concerns you've mentioned or potential harms we've discussed throughout this research, what rules/principles would you like to see to address these?** **What kind of rules would you want to see to:** • Balance the motivations of private companies with the public good **Looking at our set of rules, how can we manage the risk that they are too burdensome for private companies resulting in the "missed use" of data?** *Moderator then works with participants to add/amend their initial set of principles.* **If these rules were in place, would you feel confident about smart data being used for research?** • What makes you say that? *Probe on what influences trust – e.g. who accesses data, what it is used for, how it is stored.* • What, if anything, is missing? • What are the absolute most important points for SDR UK to consider?	
3.7 Wrap-up and close **Aim:** Explain next steps.	*Lead facilitator to lead plenary session:* **What are your hopes and fears for smart data research now? Have they changed in anyway?** • Why would you say that? • What information are you missing or hope to learn more about? *Finally, lead facilitator to:* • *Thank participants for their time today* • *Remind participants of payment details* • *Confirm next steps*	5

Stimulus

The materials used in workshop 3 include:

- Running slides, expert speakers, a short quiz about smart data research, and case studies. These are available upon request.

Information shared with participants to recap on smart data research

Smart data

Smart data is… Data that is generated as a byproduct of our digital lives – so when we interact with products, services and devices online

Smart data is not… Data that is gathered specifically for use in a research project

Smart data research

Smart data research is… A research project that analyses 'smart data' – so the data that is generated when we interact with products, services and devices online

Smart data research is not… A research project that analyses any type of data

Smart data research involves…

A private company* *SDR UK calls them data controllers:* They collect smart data as part of their business activities. They decide to share smart data so it can be used by researchers working for the public good.

A data service: They work with private companies to take the smart data and turn it into something researchers can use (a data set) and set the rules of who can access it , when and how.

Scientists and researchers: They access the smart data and use it to understand the big social challenges we are facing

SDR UK: SDR UK are here to help this process by funding teams of scientists and researchers (data services), providing advice, coordinating the various organisations involved, and setting the priorities for smart data research in the UK.

What is smart data?

A) Data that is gathered specifically for use in a research project
B) Data that is exclusively generated from smart meters
C) Data that is generated as a byproduct of our digital lives – so when we interact with digital products, services and devices **(correct answer)**

What is smart data research?

A) A research project that analyses any type of data
B) A research project that analyses 'smart data' – so the data that is generated when we interact with products, services and devices online **(correct answer)**
C) A research project conducted by people whose mums say they are "very smart indeed"

Who are SDR UK?

A) SDR UK is here to help the smart data research process by funding data services, providing advice, coordinating the various organisations

involved, and setting the priorities for smart data research in the UK. **(correct answer)**

B) SDR UK conduct smart data research themselves
C) SDR UK is an organisation where the smartest people in the UK come together to solve big problems

Recap: What we have learnt so far

1. When thinking about the "public good" many of you focused not only on what would help the most people, but also what would make the biggest difference to help address disadvantage and inequality.
2. SDR UK's four pillars – productivity and prosperity, health and wellbeing, digital society, sustainability – feel like important areas to focus on.
3. Most of you had not considered the risks of data bias before. Hearing from Maxine Mackintosh was reassuring – because someone is working on the problem – but also raised concerns about the accuracy and reliability of data.
4. Existing data access rules feel a little technical and it is not yet clear how they might work for smart data research. Your priorities are transparency, accountability, and ensuring that research is motivated by the public good.

Case study: Loyalty card data and ovarian cancer

Researchers worked with Boots and Tesco to test the hypothesis that women self-treat before seeing a GP and that a change in shopping behaviours surrounding pain and indigestion can be a flag for ovarian cancer.

They did this by comparing past loyalty card data of women who agreed to share their shopping data. Participants in the study were also asked to complete a questionnaire about ovarian cancer risks, their symptoms and any cancer referral diagnoses.

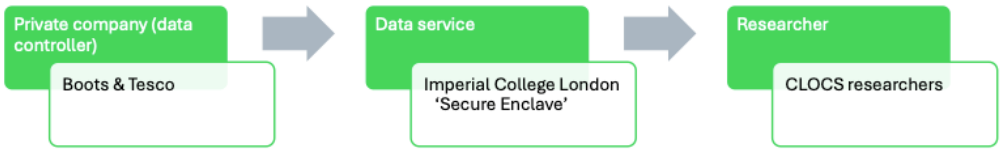
By identifying individuals who were purchasing pain and indigestion medications – potential signs of ovarian cancer – the study has shown it might be possible to develop an early warning system to encourage patients to meet with their GPs and receive more accurate diagnosis.

Case study: Loyalty card data and ovarian cancer

Researchers worked with Boots and Tesco to test the hypothesis that women self-treat before seeing a GP and that a change in shopping behaviours surrounding pain and indigestion can be a flag for ovarian cancer.

They did this by comparing past loyalty card data of women who agreed to share their shopping data. Participants in the study were also asked to complete a questionnaire about ovarian cancer risks, their symptoms and any cancer referral diagnoses.

By identifying individuals who were purchasing pain and indigestion medications – potential signs of ovarian cancer – the study has shown it might be possible to develop an early warning system to encourage patients to meet with their GPs and receive more accurate diagnosis.



Private & Confidential 14

Case study: Banking data and what it tells us about later life

Researchers in Scotland led a 'deep dive' into NatWest transaction data, focussing on older citizens, and examining the risk of financial vulnerability caused by employment ending, insufficient pension pots, and potential poor planning.

They used aggregated banking transaction data from almost 50,000 Nat West customers over the age of XX. They also carried out interviews and focus groups with 62 with experts at providing advice to older workers.

Their key findings showed that:

- Adults aged 50–54 are substantially more at risk of financial vulnerability than older retired individuals
- Retired individuals are withdrawing large sums from their pension pots when already struggling financially, more than doubling their risk of financial vulnerability.
- The largest groups of individuals at risk are in Greater London and the Northeast.

Case study: Banking data and what it tells us about later life

Researchers in Scotland led a 'deep dive' into NatWest transaction data, focussing on older citizens, and examining the risk of financial vulnerability caused by employment ending, insufficient pension pots, and potential poor planning.

They used aggregated banking transaction data from almost 50,000 Nat West customers over the age of XX. They also carried out interviews and focus groups with 62 with experts at providing advice to older workers.

Their key findings showed that

- Adults aged 50–54 are substantially more at risk of financial vulnerability than older retired individuals
- Retired individuals are withdrawing large sums from their pension pots when already struggling financially, more than doubling their risk of financial vulnerability.
- The largest groups of individuals at risk are in Greater London and the Northeast.

15

Case study: Housing and social exclusion

Researchers examined data on private renting from the property website Zoopla to understand the impact that the rise in private renting has had on low-income households' ability to rent homes in central locations in towns and cities. The team also examined data on Housing Benefits from the Department of Work and Pensions.

Analysis of the data revealed that:

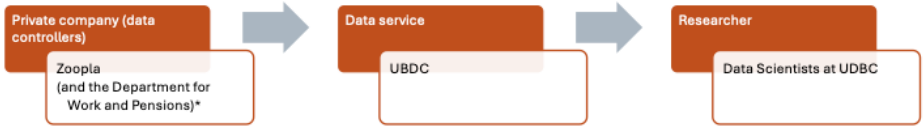
- While rents are rising everywhere, they are rising faster in more central locations, particularly in the larger cities.
- Increase in private renting has led to 1 in 9 poorer households being pushed out of central locations in towns and cities in the last 8 years.
- The number of private listings that were affordable for those on Housing Benefit had fallen from 20% to just 9%.
- In contrast, although they are on the decline, socially rented homes remain affordable in central locations.

Case study: Housing and social exclusion

Researchers examined data on private renting from the property website Zoopla to understand the impact that the rise in private renting has had on low-income households' ability to rent homes in central locations in towns and cities. The team also examined data on Housing Benefits from the Department of Work and Pensions.

Analysis of the data revealed that:

- While rents are rising everywhere, they are rising faster in more central locations, particularly in the larger cities.
- Increase in private renting has led to 1 in 9 poorer households being pushed out of central locations in towns and cities in the last 8 years.
- The number of private listings that were affordable for those on Housing Benefit had fallen from 20% to just 9%.
- In contrast, although they are on the decline, socially rented homes remain affordable in central locations.



Private & Confidential 16

Case study: Using location data to plan better infrastructure

Researchers have worked extensively with data from Strava, a fitness app that allows users to track a variety of sporting activities, including logging and sharing their cycle journeys. The data from the app enables town planners to understand which routes people use, and which destinations are busiest.

By analysing Strava data in Glasgow, researchers found that:

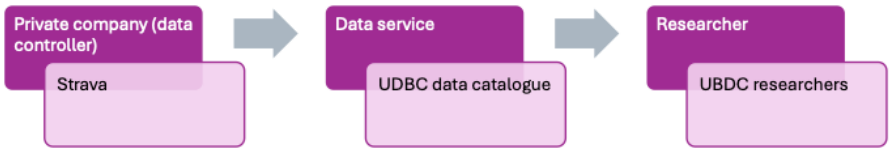
- Providing safe cycle paths could encourage people to cycle more, especially on dry days, but cycling decreased on rainy days.
- This suggests that when planning for rainy locations, other policies may be needed to increase cyclists' resilience to bad weather (e.g., providing shower facilities at workplaces, incentives to cycle, etc.).

Case study: Using location data to plan better infrastructure

Researchers have worked extensively with data from Strava, a fitness app that allows users to track a variety of sporting activities, including logging and sharing their cycle journeys. The data from the app enables town planners to understand which routes people use, and which destinations are busiest.

By analysing Strava data in Glasgow, researchers found that:

- Providing safe cycle paths could encourage people to cycle more, especially on dry days, but cycling decreased on rainy days.
- This suggests that when planning for rainy locations, other policies may be needed to increase cyclists' resilience to bad weather (e.g., providing shower facilities at workplaces, incentives to cycle, etc.).



Private & Confidential 17

Examples of risks of private companies sharing smart data that we shared with participants:

The public don't feel that they have consented to their data being used for research. An example we made up…

- Banking data from 'Safe Harbour Bank' is used by researchers to highlight the negative financial impact of using online gambling websites
- Researchers only viewed the data in aggregate in a large data set
- But some Safe Harbour Bank customers still felt they hadn't consented to their spending data being used in this way

Private companies don't get involved because it is seen to be too complicated, missing an opportunity to enable research that could have a meaningful positive impact on the public. An example we made up…

- 'Swift Ride', a taxi-hailing app, have data about the most used taxi routes for women late at night.
- Researchers in Nottingham could use this data to highlight the types of places in town where additional streetlamps and designated pick-up points could help women get into taxis safely.
- But Swift Ride thinks that it is too complicated and costly to share this data so decide not to.

The financial cost of getting data from a private company outweighs the potential public benefit of the research. An example we made up…

- An energy company called the 'WarmHomeAlliance', shares its customers' smart meter data.
- It charges £50,000 to cover its costs.
- Researchers use the data to study changes in weather, and what that means for energy use and affordability around the UK.
- But to get a true picture of the problem, this data would need to be combined with data from other companies. This proves too difficult to achieve and so the research has no actionable findings.

Workshop 4: Discussion Guide and Materials

Workshop objectives:

The specific objectives for this workshop are to:

- Remind participants of the key players involved in SDR and how smart data will be shared.
- Explore principles around data sharing and governance from other data services and trusted research environments and discuss how they would apply to SDR.
- Discuss rules / principles participants would put in place to ensure data is handled safely and securely

Workshop 4 guide | July 2024		
Section and aim	**Key questions and probes**	**Time**
Arrival	**Participants arriving** • *Participants will be asked to log in from 5:45pm* • *Sign people in and check user names* • *Address tech issues*	-
4.1 Welcome and introduction to the session	*Lead facilitator to welcome all participants*: • *Recap purpose of dialogue* • *Introduce people on the call (Thinks team, SDR UK / Sciencewise attendees and experts)* • *Reiterate ground rules for participating in the sessions* *Lead facilitator to introduce the session:* • *Purpose of the session* • *Agenda for the day* • *Lead facilitator to run through polls:* • *Recapping key data concepts and definitions, clarifying areas of confusion* • *Outlining headlines so far, probing on agreement / disagreement*	15
4.2 Warm-up and headline recap **Aim**: Get participants'	*Participants to be split into breakout groups.* *Facilitator to introduce self and ask participants to recap introductions:* • *Name*	10

feedback on the ongoing analysis process.	• *One exciting thing that's happened since we last met* In today's session we will continue to discuss smart data research, looking at how data should be handled to minimise risk and avoid any harms. But first, we're going to continue to reflect a bit on the last online workshop. I'm going to share some of the lessons we've learnt in all our different locations about smart data research so far to hear how you feel about these and make sure you feel we haven't missed anything! *Facilitator to share screen with headline findings, noting down areas of agreement / disagreement for each as they are raised.* **What do you think about this?** • To what extent, if at all, do you agree with this? • What, if anything, would you add or change about this? • Is there anything big that has been missed out? • Is there anything you're surprised to see?	
4.3 Reminder of SDR eco system and key terms **Aim:** Remind participants of the key players involved in SDR and how data is shared.	*Participants to return to plenary for SDR UK presentation.* In the first workshop, we heard from SDR UK about how private companies, data services, researchers, and SDR UK all interact with one another. We learned about how data flows between these different organisations/people. Before we discuss the role of data services in more detail, it might be helpful to recap the key things we learned a few weeks ago. *In plenary, lead facilitator to recap presentation / replay SDR UK presentation explaining:* • *Who the key players in SDR are (i.e. the public, private companies, SDR UK, researchers and policy makers/media/the public)* • *The role of SDR UK in SDR* • *How does SDR UK work with Data Services? How does SDR UK oversee them and manage risks?* • *What's the ecosystem between individual researchers and companies and where does SDR UK fit in?*	15

	• *How do Data Services get private companies to share their data? How do they manage access to this data?* • *Q&A to recap* • *Poll recapping key ideas to ensure participants' understanding*	
4.4 Why does privacy matter? **Aim:** To help participants better understand the importance of privacy and data protection.	*Participants to remain in plenary for expert testimonial* One of the considerations all the organisations and individuals involved in smart data research have is ensuring data used for smart data research is securely handled. We're going to hear from Cassie Smith about data protection and privacy, and then spend some time discussing it in more detail. • *Data protection and privacy expert to introduce themselves and their role/organisation, before covering:* • *What privacy means in a research context* • *Why it is important* • *What are the privacy risks and potential harms associated with SDR, including being identified from the data or not wanting your sensitive data to be used for this purpose, and what are the potential harms if these risks are realised* • *Q&A* *In breakout groups:* • **How do you feel about what you've just heard?** • **Has this changed how you feel about the security of your data or not?** • What makes you say that? • What else, if anything, would you need to know to trust your data is being securely handled? • **Has this changed how you feel about your data being used for smart data research or not?** • **What, if any, concerns do you have about smart data research after hearing this?** One of the risks was that individuals might end up being identified when datasets are linked up, or if data used for smart data research isn't handled securely. Thinking about this risk: • **What do you think about this risk?**	35

	• How likely do you think it is to happen? • How concerned, if at all, are you about this risk? • **What, if anything, do you think the impact of re-identification would be?** *Facilitator to probe to understand what the perceived harms of re-identification are, and whether participants have specific concerns beyond loss of privacy.* • **Who, if anyone, would be impacted by this?** • **Can you think of any way that this risk could be minimised?** • What could the private company / researchers / SDR UK do to reassure you that this risk has been minimised? • What do you think the public need to know in order to be reassured about this risk?	
BREAK	7:15 – 7:25 pm	10
4.5 Engaging with principles for data sharing **Aim:** Introduce principles other data services follow for how data should be handled.	*In plenary:* Welcome back! We're going to spend some time now discussing how data is accessed by both data services and researchers. The concerns we have discussed, including about data being securely handled, are not specific to smart data research. Researchers and universities who use data have already considered these risks and potential harms, and so have systems in place to make sure data is securely handled. This includes trusted research environments (TREs, also known as secure data environments), which are highly secure computing environments that contain de-identified data. These TREs follow rules (e.g. the 5 Safes framework), and researchers and their projects must go through a rigorous application process to access and use this data. Presentation from researcher about how data used for research is managed in practice: • *How data is protected (5 safes) and structures in place to ensure data is being accessed and used correctly* • *What is a Trusted Research Environment?* When, why and how do researchers access TREs?	20

	• *The risk that an overly restrictive approach leads to the "missed use" of data.* *In breakout groups:* **How do you feel about what you've just heard?** **Has this changed how you feel about your data being used for smart data research or not?** **What, if any, concerns do you have about smart data research after hearing this?** • What makes you say that? • What else, if anything, would you need to know to trust your data is being securely handled? **Is there anything missing here that needs to be in place for smart data specifically?** By this I mean is there anything specific to the types of data that might be accessed? Or to the types of research/researcher? As you know, SDR UK oversees the data services that provide data to researchers. The data services are responsible for ensuring that consumer data is safe. Data services follow rules (e.g. the 5 Safes, CDRC's access criteria, etc.) We're going to hear more now about examples of the types of rules and conditions that other data services use. Some of these were developed via public engagement like this one, others are based on the views of experts. They also relate to different types of data, some apply to very sensitive data, others to open data or statistics (that's data anyone can access). Presentation from Thinks about data services: Introducing principles from other data services around: • Who should be able to access data and why • The types of data they should be able to access • How they would be able to use it • Q&A *In breakout groups:* **How do you feel about what you've just heard?** **Has this changed how you feel about your data being used for smart data research or not?** **What, if any, concerns do you have about smart data research after hearing this?** • What makes you say that?	

	• What else, if anything, would you need to know to trust your data is being securely handled? **Is there anything missing here that needs to be in place for smart data?**	
4.6 What should SDR UK do? **Aim:** Laying the groundwork for final principles on data handling in workshop 5.	For the rest of this session, we'd like to focus on one key question: If you were in charge, what rules would you put in place to ensure smart data is handled safely? We will do this by coming up with our own principles around the areas that researchers and data services consider when thinking about how smart data is shared. *If needed the moderator can probe on:* **What kinds of rules and behaviours should data services be implementing?** **What kinds of rules and behaviours should researchers be implementing?** **What should SDR UK be doing to encourage these?** **Thinking about some of the concerns you've mentioned or potential harms we've discussed throughout this research, what rules/principles would you like to see to address these?** **What kind of rules would you want to see around:** • Who would be able to access smart data? • What kind of data they would be able to access? • How they would be able to use it? **Looking at our set of rules, how can we manage the risk that they are too burdensome for researchers resulting in the "missed use" of data?**. **If these rules were in place, would you feel confident about smart data being used for research?** What makes you say that? *Probe on what influences trust – e.g. who accesses data, what it is used for, how it is stored.* • What, if anything, is missing? • What are the absolute most important points for SDR UK to consider?	30

4.7 Wrap-up and close **Aim:** To explain next steps.	*Lead facilitator to lead plenary session, summing up the principles decided and gathering feedback on them as a whole.* *Finally, lead facilitator to:* • *Thank participants for their time today* • *Remind participants of payment details* • *Confirm next steps*	15

Stimulus

The materials used in workshop 4 include:

- Running slides, expert speakers, and a case study on reidentification. These are available upon request.

Workshop four included a presentation from Cassie Smith, Head of legal, Trust and Ethics at Health Data Research UK (HDR UK).

The purpose of this presentation was to outline the common concerns regarding using the public's data for smart data research including the loss of privacy and the risk of cyber-attacks.

However, Cassie explained to participants that is important for the public to trust and allow private companies and researchers to access and use their data as opting out may lead to bias and incomplete datasets. To reassure participants Cassie outlined the steps companies take to protect the public's data including:

- Deidentification
- Storing the data in a secure environment
- Ensuring data is used by the correct researchers for the correct purpose i.e. the public good

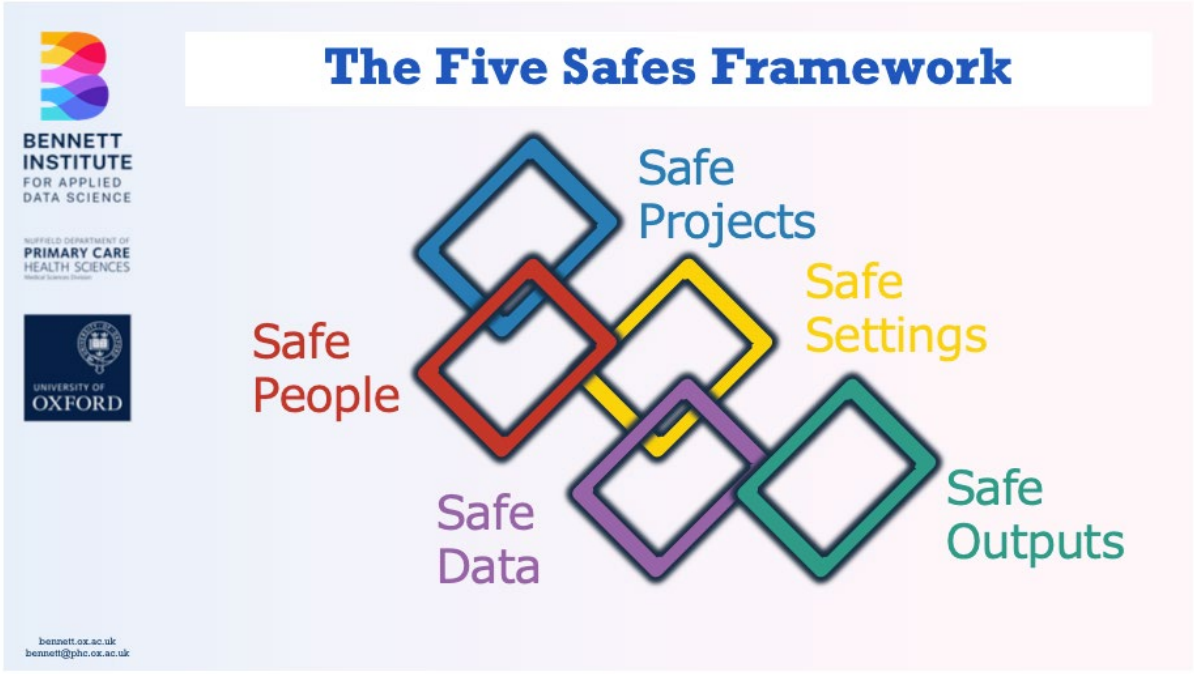
Cassie also explained that private companies have a legal obligation to adhere to privacy laws and if they do not follow these laws they are penalised.

Workshop four also included a presentation from Pete Stokes, Director of Platform Development for Bennett Institute for Applied Data Science.

The purpose of this presentation was to discuss the Five Safes in data research, to address data security concerns from the participants and inform them the rules that are followed when smart data is used for research. The Five Safes include:

- **Safe projects:** Proposal must have support of data owner(s). Research must "serve the public good". Results must be published. Project proposals should be reviewed by a recognised Ethics Panel. Project proposals must have a clear scope, be feasible and time-bound, have the right data and variables, and use appropriate methodology.

- **Safe settings:** Trusted Research Environments protect data using a range of controls including secure technology, physical security and security procedures and protocols.
- **Safe outputs:** A Safe Output is non-disclosive & ensures the confidentiality of data subjects. Most TRES follow a "principles-based" approach to check outputs. This requires collaboration with researchers and does not introduce unnecessary barriers to research.
- **Safe people:** Researchers must demonstrate their competence to use data, through academic qualifications and/or research experience.
- They must also commit to transparency requirements of legislation. Researchers are trained in how to use secure environments and to ensure outputs are not disclosive. After training, researchers complete a short assessment, to confirm that they understand their obligations and will be able to use data as intended. Ongoing assessment of researchers in the environment, reviewing how they interact with support teams, and how they conduct research.
- **Safe data:** Clean, prepare and de-identify the data



BENNETT INSTITUTE FOR APPLIED DATA SCIENCE

NUFFIELD DEPARTMENT OF PRIMARY CARE HEALTH SCIENCES

UNIVERSITY OF OXFORD

The Five Safes – Safe People

Application Process

Researchers must demonstrate their competence to use data, through academic qualifications and/or research experience.

They must also commit to transparency requirements of legislation.

Training

Researchers are trained in how to use secure environments and to ensure outputs are not disclosive.

Assessment

After training, researchers complete a short assessment, to confirm that they understand their obligations and will be able to use data as intended

Ongoing review

Ongoing assessment of researchers in the environment, reviewing how they interact with support teams, and how they conduct research.

bennett.ox.ac.uk
bennett@phc.ox.ac.uk

BENNETT INSTITUTE FOR APPLIED DATA SCIENCE

NUFFIELD DEPARTMENT OF PRIMARY CARE HEALTH SCIENCES

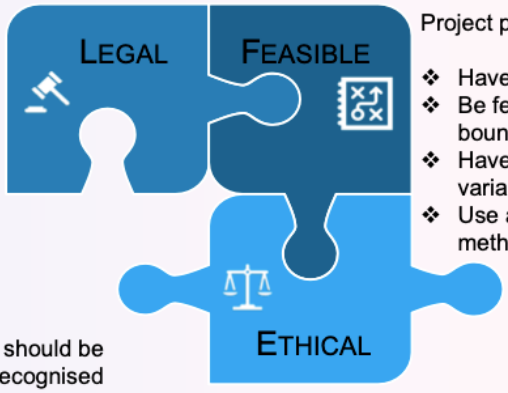
UNIVERSITY OF OXFORD

The Five Safes – Safe Projects

Proposal must have support of data owner(s)

Research must "serve the public good"

Results must be published



Project proposals must:

- Have a clear scope
- Be feasible and time-bound
- Have the right data and variables
- Use appropriate methodology

Project proposals should be reviewed by a recognised Ethics Panel

bennett.ox.ac.uk
bennett@phc.ox.ac.uk

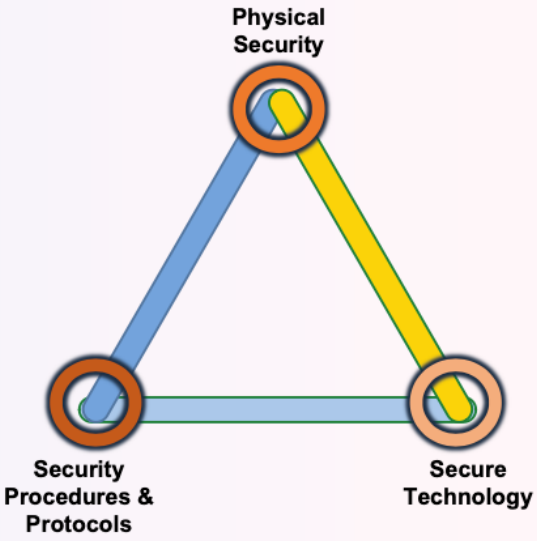
BENNETT INSTITUTE FOR APPLIED DATA SCIENCE

NUFFIELD DEPARTMENT OF PRIMARY CARE HEALTH SCIENCES

UNIVERSITY OF OXFORD

The Five Safes – Safe Settings

Trusted Research Environments protect data using a range of controls.



bennett.ox.ac.uk
bennett@phc.ox.ac.uk

BENNETT INSTITUTE FOR APPLIED DATA SCIENCE

NUFFIELD DEPARTMENT OF PRIMARY CARE HEALTH SCIENCES

UNIVERSITY OF OXFORD

The Five Safes – Safe Settings

Secure Technology

No internet connectivity and restrictions on data recording methods (e.g. phones, smart wear, notepads)

Restrictions on remote working (e.g. IP locking)

Scrutiny of software tools made available

Hardware, firmware, and software meeting high security specifications

Monitoring of researchers' interactions (e.g. keylogging, screen captures)

Physical Security

Audited access to safe rooms

CCTV monitoring

Coordination in allocating workstations to researchers

Security Procedures & Protocols

Processes audited against ISO standards

Regular audits of all security protocols

Daily audits of researchers' interactions

Clear records of breaches of procedures

bennett.ox.ac.uk
bennett@phc.ox.ac.uk

BENNETT INSTITUTE FOR APPLIED DATA SCIENCE

NUFFIELD DEPARTMENT OF PRIMARY CARE HEALTH SCIENCES

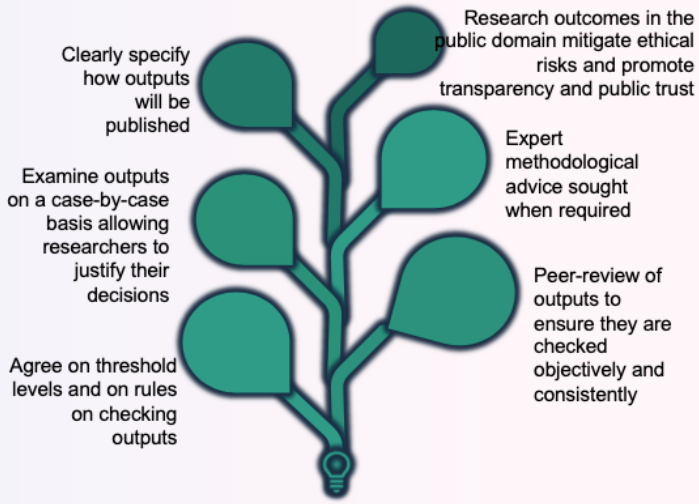
UNIVERSITY OF OXFORD

The Five Safes – Safe Outputs

A Safe Output is non-disclosive & **ensures the confidentiality** of data subjects.

Most TRES follow a "principles-based" approach to check outputs.

This requires **collaboration with researchers** and does not introduce unnecessary barriers to research.



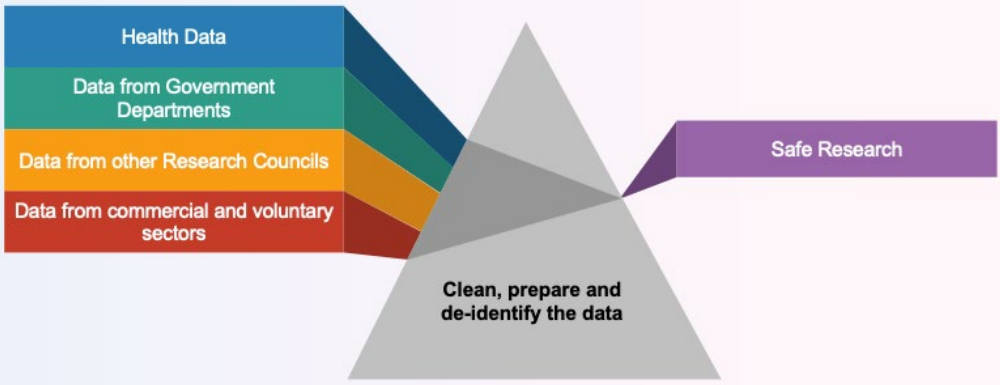
bennett.ox.ac.uk
bennett@phc.ox.ac.uk

BENNETT INSTITUTE FOR APPLIED DATA SCIENCE

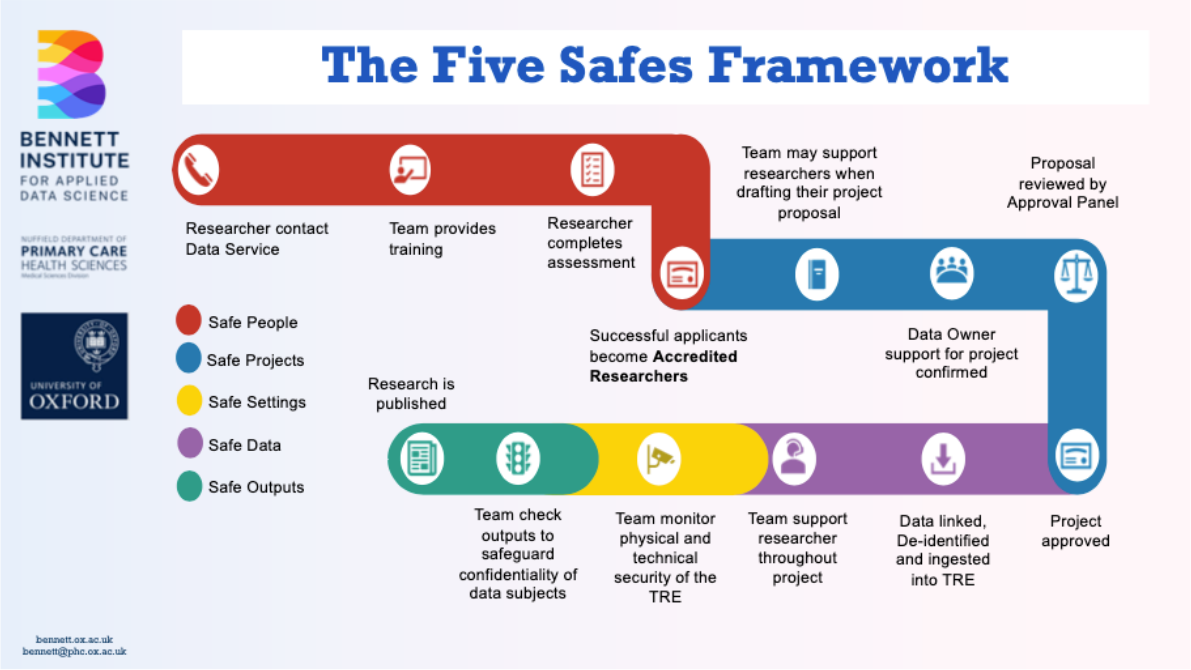
NUFFIELD DEPARTMENT OF PRIMARY CARE HEALTH SCIENCES

UNIVERSITY OF OXFORD

The Five Safes – Safe Data



bennett.ox.ac.uk
bennett@phc.ox.ac.uk



Example of a Data service

The Transport Smart Data Service (A team of researchers and scientists funded by SDR UK)

Mission: Work with transport companies to make their data safely accessible to researchers

Research fields this might appeal to:

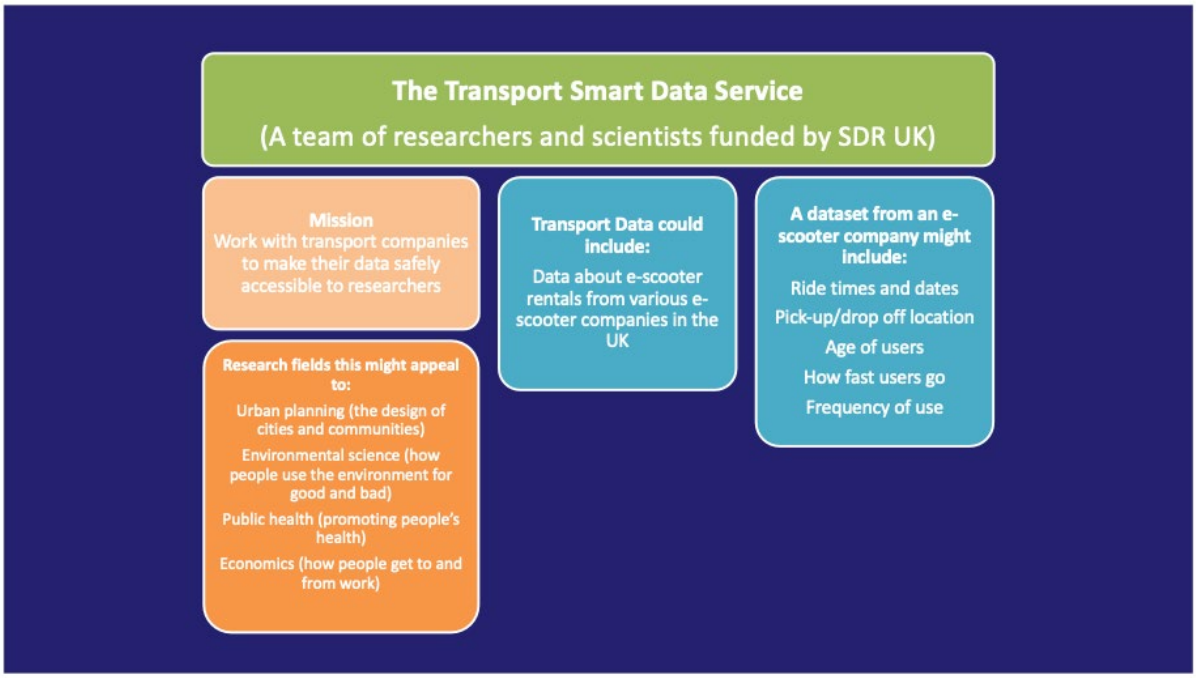
- Urban planning (the design of cities and communities)
- Environmental science (how people use the environment for good and bad)
- Public health (promoting people's health)
- Economics (how people get to and from work)

Transport data could include:

- Data about e-scooter rentals from various e-scooter companies in the UK

A dataset from an e-scooter company might include:

- Ride times and dates
- Pick-up / drop off location
- Age of users
- How fast users go
- Frequency of use


The Transport Smart Data Service
(A team of researchers and scientists funded by SDR UK)
Mission
Work with transport companies to make their data safely accessible to researchers
Research fields this might appeal to:
Urban planning (the design of cities and communities)
Environmental science (how people use the environment for good and bad)
Public health (promoting people's health)
Economics (how people get to and from work)
Transport Data could include:
Data about e-scooter rentals from various e-scooter companies in the UK
A dataset from an e-scooter company might include:
Ride times and dates
Pick-up/drop off location
Age of users
How fast users go
Frequency of use

Workshop 5: Discussion Guide and Materials

Workshop objectives:

The specific objectives for this workshop are to:

- Bring everything together after workshops 1-4 and take stock on hopes and priorities for smart data research and SDR UK
- Formulate recommendations for SDR UK across the themes of public good, public-private collaboration, data handling and security, and public engagement
- Understand people's expectations for public engagement in smart data research going forward

Workshop 5 guide (F2F) | 2nd October 2024		
Section and aim	**Key questions and probes**	**Time**
Arrival	**Participants arriving** *Sign people in* *Participants settle in at their tables*	-
5.1 Welcome and introduction to the session	*Lead facilitator to welcome participants to their final session, remind them of objectives, recap journey through the dialogue so far:* • *Introduce people across locations (Thinks team, SDR UK / Sciencewise attendees and experts)* • *Reiterate ground rules for participating in the sessions* • *Recap from Mark in Inverness to remind participants of the dialogue objectives and talk about how the findings will be used* *Lead facilitator in each location to introduce the session:* • *Purpose of the session* • *Agenda for the day* • *Public good* • *Relationship with the private sector* • *Ensuring data is handled safely* • *Public involvement*	10:00-10:20

	Lead facilitator in each location to share slides outlining: • *Recap of what SDR UK is and does (introduce updated diagram outlining what SDR does at each stage of the research process and in relation to each actor)* • *Give a sense of the levers at its disposal (influence, funding, etc.) and the things it can't/won't and can/might do* • See also participant workbook pages 2 and 3 • We also have some answers to your questions for Pete Stokes to share with you. *Lead facilitator in each location to share answers to questions for Pete Stokes* *Lead facilitator to hand over to the break-out facilitators for introductions and in-depth discussion.*	
5.2 Summary: benefits and potential harms of SDR **Aim:** To summarise the conversation and understand participants key takeaways of the benefits and potential harms of smart data research.	*Participants to be split into breakout groups.* *Facilitator to introduce self and ask participants to recap introductions:* • *Name* • *One exciting thing that's happened since we last met* We've spent a lot of time in the previous sessions hearing from some of the people involved in smart data research about its potential benefits and downsides and thinking about our hopes and fears for smart data research going forward. I just want to spend some time now summarising the key potential benefits and harms as you see them. Remember, it's your opinion we're interested in, so if you don't agree with something that's said we want to know, and if there are things you think are important that haven't been mentioned, feel free to share. *Facilitator to use flipchart to record benefits and harms discussed.*	10:20-10:35

	What do you think are the most important benefits offered by smart data research? • Which of these is the most important to prioritise? **What do you think are the key potential harms of smart data research?** • Which of these is the most important to consider? • Are these potential harms the same for all types of data, or do they differ? *Facilitator to ask participants to look at "on our street" summary in their workbooks (page 4) to remind participants of personas.* **Thinking back to all the people who live on Wisteria Walk, do you think these are the same potential benefits and harms they'd identify?** • If not, which benefits and harms do you think would be most important to them? • Are there any additional benefits or harms they'd prioritise that we haven't considered? • Would their views differ depending on different types of data?	
5.4 Introducing the next sessions	*Lead facilitator in locations in plenary:* We're going to focus the next sessions of our discussion on three key areas that SDR UK is thinking about and that we have been discussing: • Research for the public good • The role of private companies in smart data • Data sharing and the role of data services Today we'd like to work towards developing some final recommendations around smart data research. Based on the discussions we had in the last sessions, we have summarised your key concerns for each area, as well as your suggestions for the rules you would put in place if you were in charge to address them. We'll discuss these rules in more detail	10:35-10:40

	today and use them to develop a set of principles that SDR UK should use to encourage private companies, researchers and data services to carry out smart data research in the way you want.	
5.5 Research for the public good **Aim:** Refine principles from previous sessions and explore any questions or trade-offs that have emerged.	We're going to start by discussing the things that came out of the second workshop, which was all about the public good. *Workbook page 5.* Summarising the key concerns and discussion points from that workshop: You were all relatively open to the idea of smart data research for the public good but had discussions about: • What the research is actually being used for and what impact it will have in the real world • Who gets to decide what the "public good" is • If "public good" always means benefitting the greatest number of people, or if it is also public good if it only benefits a few but addresses disadvantage or inequality, or, for example, a rare disease or condition • How quickly you would expect to see a positive impact and whether public good could also be delayed, or achieved for future generations • Whether you would expect to personally see a benefit from smart data research for the public good Reflecting on these themes now, and keeping in mind everything we have learned about smart data research since, I'm interested to hear: **How, if at all, your perceptions of what "public good" means have changed?** **How, if at all, your views on the potential for smart data research achieving public good have changed?** **If you feel you have (different) answers to some of these questions now?**	10:40-11:25

	As you know, SDR UK have been present for our discussions and are very interested in the outcome of this research, and the principles we agree together. There are a few things related to our discussions about research for the public good that SDR UK wanted us to ask you a little bit more about. **How would you know research has been done for the public good?** • Are some types of research that are more for the public good than others? • What information do you usually need to feel confident it has been for the public good? (e.g. type of issue, impact, scale of impact, people impacted) • How could SDR UK make the public aware about the impact of smart data research on the public good? What, if anything, would you expect them to say, when and where? At the end of that session, we asked you to come up with some rules to ensure that smart data research delivers for the public good. As a reminder, here are the rules we came up with in our second workshop: *Facilitator to ask participants to look at rules from workshop 2 in their workbooks, spending a few minutes gathering reflections on them. If participants veer too much into talking about private companies or data handling, remind them there is an opportunity to review those rules later.* You were happy for your smart data to be used for research for the public good as long as: • The data isn't biased or excludes people and outcomes are fair • There is a clear need for the research and a (potential) real-world impact • There is oversight and accountability to make sure research benefits the public good We're keen to understand what you think each of these means in practice. In your workbooks (pg 5) there is space next to each	

	principle to expand and explain what this means in real life – you could also just give an example if that's easier! *Participants to spend 5 minutes on worksheet in their book before feeding back to the group. Moderator to ask:* **What do you think each of these means in practice?** How practical/feasible does this feel? What would need to be in place for this to work? **Thinking about the whole set, and keeping in mind everything we have learned about smart data research since, what do you think about these rules?** **How, if at all, has your opinion changed? Why?** Now that we've recapped what we saw as the rules for making sure smart data research works for the public good, and delved a little bit deeper to answer some of SDR UK's questions on this topic, we are going to work together to refine these rules into a set of principles for how SDR UK should think about this topic. Working with these short draft statements, it's our job now to refine them to come to a final set of principles that our group are happy with. **Given everything you have learned, which rules do you think are still important that we can use as the basis for our principles?** **Are there any which you think are less important we should delete? Why?** **Are there any other points you would add now, having learned more about the topic of smart data research?** **None of these rules say anything about the types of research that you would consider to be "in the public good". Would you like to add a principle that defines what this means in practice?** **Which one would you most like to see implemented? Is there anything you would change about it?**	

	Now we've spent some time discussing these rules, let's draft three recommendations (or pieces of advice) we would give to SDR UK for how they should think about this topic. These could be our 3 top principles from those already discussed, or they could be new principles based on what we've learned so far. *After discussion, table facilitator to work with participants to generate a set of three principles/priorities for SDR UK related to public good – record in running slides or on flipchart.* *Reconvene in plenary to feedback to group:* *In locations, lead facilitator to ask participants from each table to share their three priorities and decide on a set of 3 for the whole location by voting and/or combining priorities.* *Then join other locations on Zoom to share each location's "winners".*	
REFRESHMENT BREAK	11:25-11:35	11:25-11:40
5.6 The role of private companies in smart data research **Aim:** Refine principles from previous sessions and explore any questions or trade-offs that have emerged.	We're now going to focus on what came out of the third workshop, which focussed on commercial relationships and the role of private companies in smart data research. *Workbook page 6.* Summarising the key concerns and discussion points from that workshop: • You were sceptical about the motivations of private companies for getting involved in smart data research and many of you thought they were motivated by profit only • You don't necessarily trust private companies to do anything for the public good – and if they do, you assume it would be a PR exercise • You're not sure whether public good and private profit can go hand in hand • There were questions about consent – most of you assumed you may have agreed to this use of your data for research via T&Cs but you weren't sure	11:40-12:25

	How, if at all, your perceptions of private companies' role in smart data research have changed? **If you feel you have (different) answers to some of these questions now?** As you know, SDR UK have been present for our discussions and are very interested in the outcome of this research, and the principles we agree together. There are a few things related to our discussions about private companies that SDR UK wanted us to ask you a little bit more about. **What, if anything, would you expect SDR UK to do to support and encourage private companies to share their data with data services?** **How much involvement should SDR UK have in the relationship between a private company and data service?** • How much should SDR UK know about these partnerships? • In some cases, private companies might prefer that their relationship with data services is kept private, which would mean data services wouldn't share who they were receiving data from (including with SDR UK). How do you feel about this? • How much oversight should SDR UK have over this? At the end of that session, we asked you to come up with some rules to ensure that commercial relationships for smart data research are fair. As a reminder, here are some of the rules we came up with in our third workshop: *Facilitator to ask participants to look at rules from workshop 3 (pg 6) in their workbooks, spending a few minutes gathering reflections on them.* You wanted there to be clear oversight and accountability to ensure that research is actually for the public good, rather than for private profit	

	You thought there should be more transparency for consumers on the use of their data There are certain organisations and types of data you are more comfortable with than others and rules should be stricter for sensitive data. We're keen to understand what you think each of these means in practice. In your workbooks there is space next to each principle to expand and explain what this means in real life – you could also just give an example if that's easier! *Participants to spend 5 minutes on worksheet in their book before feeding back to the group. Moderator to ask:* **What do you think each of these means in practice?** • How practical/feasible does this feel? • What would need to be in place for this to work? **Thinking about the whole set, and keeping in mind everything we have learned about smart data research since, what do you think about these rules?** **How, if at all, has your opinion changed? Why?** Now that we've recapped what we saw as the rules for making sure commercial relationships are fair, and delved a little bit deeper to answer some of SDR UK's questions on this topic, we are going to work together to refine these rules into a set of principles for how SDR UK should think about this topic. Working with these short draft statements, it's our job now to refine them to come to a final set of principles that our group are happy with. **Given everything you have learned, which rules do you think are still important that we can use as the basis for our principles?** **Are there any which you think are less important we should delete? Why?**	

	Are there any other points you would add now, having learned more about the topic of smart data research? **Which one would you most like to see implemented? Is there anything you would change about it?** Now we've spent some time discussing what we think fair commercial relationships look like in practice, let's draft three recommendations (or pieces of advice) we would give to SDR UK for how they should think about this topic. These could be our 3 top principles from those already discussed, or they could be new principles based on what we've learned so far. *After discussion, table facilitator to work with participants to generate a set of three principles/priorities for SDR UK related to public good – record in running slides or on flipchart.* *Reconvene in plenary to feedback to group:* *In locations, lead facilitator to ask participants from each table to share their three priorities and decide on a set of 3 for the whole location by voting and/or combining priorities.* *Then join other locations on Zoom to share each location's "winners".*	
5.7 Data sharing and the role of data services **Aim:** Refine principles from previous sessions and explore any questions or trade-offs that have emerged.	We're now going to revisit the fourth and most recent workshop, which was all about how smart data is stored and accessed for research. *Workbook page 7.* Summarising the key concerns and discussion points from that workshop: • You liked the Five Safes (Safe People, Safe Projects, Safe Data, Safe Settings, Safe Outputs) but had questions about who enforces these • And data services may be very safe but who is accountable if something does go wrong?	12:25-12:55

	• You were worried about the security of the data when it is being transferred from private companies to data services • Re-identification using multiple datasets did not feel like a big risk to you, but you still want to make sure that all identifiable information is removed before the data is shared Reflecting on these themes now, and keeping in mind everything we have learned about smart data research, I'm interested to hear: **How, if at all, your perceptions of safe handling of data have changed?** **If you feel you have (different) answers to some of these questions now?** At the end of that session, we asked you to come up with some rules to ensure that smart data is handled safely. As a reminder, here are some of the rules we came up with in our fourth workshop: *Facilitator to ask participants to look at rules from workshop 4 (pg 7) in their workbooks, spending a few minutes gathering reflections on them.* • Make sure the Five Safes are enforced consistently across data services and updated as risks evolve, especially accreditation and training to ensure only trustworthy people work with our data • Have clear accountability and fines/penalties for when something does go wrong • Introduce auditing of companies and researchers, with background checks to ensure no conflicts of interest We're keen to understand what you think each of these means in practice. In your workbooks there is space next to each principle to expand and explain what this means in real life – you could also just give an example if that's easier! *Participants to spend 5 minutes on worksheet in their book before feeding back to the group. Moderator to ask:*	

	What do you think each of these means in practice? • How practical/feasible does this feel? • What would need to be in place for this to work? **Thinking about the whole set, and keeping in mind everything we have learned about smart data research since, what do you think about these rules?** **How, if at all, has your opinion changed? Why?** Now that we've recapped what we saw as the rules for making sure commercial data is handled safely, we are going to work together to refine these rules into a set of principles for how SDR UK should think about this topic. Working with these short draft statements, it's our job now to refine them to come to a final set of principles that our group are happy with. **Given everything you have learned, which rules do you think are still important that we can use as the basis for our principles?** **Are there any which you think are less important we should delete? Why?** **Are there any other points you would add now, having learned more about the topic of smart data research?** **Which one would you most like to see implemented? Is there anything you would change about it?** Let's draft three recommendations (or pieces of advice) we would give to SDR UK for how they should think about this topic. These could be our 3 top principles from those already discussed, or they could be new principles based on what we've learned so far. *After discussion, table facilitator to work with participants to generate a set of three principles/priorities for SDR UK related to public good – record in running slides or on flipchart.* *Reconvene in plenary to feedback to group:*	

	In locations, lead facilitator to ask participants from each table to share their three priorities and decide on a set of 3 for the whole location by voting and/or combining priorities. *Then join other locations on Zoom to share each location's "winners".*	
LUNCH	12:55-13:30	12:55-13:30
5.8 Public involvement	*Table facilitators to spend 10 minutes gathering initial thoughts about public involvement.* We have talked a lot about the different actors in this smart data ecosystem, we've talked about the 'public good' and we've talked about 'consumers'. I now want to look in more detail at this last group and think about where **you** sit in this ecosystem. Let's spend a few minutes getting your initial views: **What do you think your role is in relation to smart data research?** **What, if any, say do you have over your data?** **What are your rights and responsibilities?** **How do you want SDR UK to involve the public?** **What could that look like?** *Back to plenary for presentation:* Like ethics and security, this is a topic that other organisations have been thinking about too, so we've drawn together a couple of speakers who have all thought a lot about this to chat about the different ways SDR UK could hear from the public. [lead facilitator to play the video] Speaker will cover: • What is the benefit of public engagement? What are the potential downsides? • What are the risks of not engaging the public?	13:30-14:05

	• What are the different levels of engagement e.g. inform – consult – engage – decide? • What are some good examples from other data services/programmes? [Q&A with Fionnuala and Lucy] Do you have any questions for the speaker? Now at your tables I'd like you to work together to decide what type of public engagement you'd like to see SDR UK put in place. *In breakout tables:* On page 8 in your workbooks, there is a summary of the different "levels" of engagement mentioned in the presentation to get you started, but we're really interested in your ideas so feel free to suggest something you don't see there. **What do you think of the options here?** • Are there any you think might work well for SDR UK? • Why? • Are there any you think won't work in SDR UK context? • How involved might you want to be in thinking about how smart data is used for research? Do any of these examples feel relevant to you? **Do you have any other ideas for how the public might be involved in SDR UK?** • Why might this work better than the examples you've seen? *Table facilitators to work with participants to generate **a set of three principles** for SDR UK related to public engagement – record in running slides or on flipchart.* *Reconvene in plenary to feedback to group:* • *In locations, lead facilitator to ask participants from each table to share their three priorities and decide on a set of 3 for the whole location by voting and/or combining priorities.*	

	• *Then join other locations on Zoom to share each location's "winners".*	
5.9 Hopes and expectations for SDR UK	*Admin assistant/notetaker to add principles to running slides during this discussion, ahead of section 5.10.* Before we summarise today's session, let's revisit your hopes and fears for smart data research again. **Thinking about everything you have heard, what are your biggest hopes for smart data research?** **And what are your biggest fears for smart data research?** I'm going to share with you a summary of the hopes and fears you and the other participants shared in workshop 1. *Workbooks page 9.* **Can you see anything that has changed?** • Why has this changed? • What has stayed the same? Why is this still important? **What is the best-case scenario?** **What is the worst-case scenario?** **Which feels more likely?** *Facilitators to support table in generating a list of ideas.*	14:05-14:20
5.10 Your recommendations	*Introduce the set of 12 principles/priorities they have generated, displaying each onscreen at the front of the room and discussing at tables.* • Looking at these as a set, is there anything missing? • Is there anything you would change? • Is there anything contradictory? • Which are most important? *Participants to revisit the personas that live on their "street" and check whether their principles will ensure the best outcome for each persona.* *Participants to think about the principles in the context of different types of data*	14:20-14:45

	Participants to think about the principles in relation to their hopes and fears – if SDR UK implemented the recommendations would they feel hopeful *A participant from each table nominated to report back in plenary.*	
5.11 Wrap-up and close **Aim:** To explain next steps.	*Lead facilitator to lead plenary session.* Representative from each group to share their final recommendations. *Finally, lead facilitator to:* • *Thank participants* • *Remind participants of payment details*	14:45 - 15:00

Stimulus

The materials used in workshop 5 include:

- Running slides and participant workbooks. These are available upon request.

Presentation from Fionnuala Ratcliff, Dialogue and Engagement Specialise at ScienceWise, Lucy Farrow, Partner at Thinks Insight and Strategy and Catherine Joynson, Head of Participant Engagement at Uk Bio Bank: What is smart data and using smart data for research.

The purpose of this presentation was to prompt participants to think about what role the public will play in smart data research including:

- What are benefits and downside of public engagement?
- What are the risks of not engaging the public?
- What are the different levels of engagement?
- What are some good examples from other data services/ programmes?

Responses to the Q&A with Pete Stokes, Director of Platform Development for Bennett Institute for Applied Data Science.

Q: Could you give an example of public benefit that a private company said they would achieve using the ONS data?

A: There are many different benefits, but an example is a project by Frontier Economics to evaluate the success of the TechNation Programme (which is taxpayer funded), which seeks to increase investment into the UK.

Q: Do you use AI to de-identify data, or is it used in any other part of the process with data?

A: No. Data are currently systematically de-identified by a member of staff, by removing risky fields (typically these include fields such as Name, Address, Date of Birth, NI Number)

Q: How long does the data services hold onto data / how do you dispose of research data once researchers are done?

A: All data services will have policies on this, but datasets are typically kept for as long as there is a demand for them (as data will typically be used for many projects). A standard policy is that, if no researcher requests access to a specific dataset for 12 months, then it should be removed from the service.

Q: Does data get updated in the datasets when researchers are using it – and if so, what happens to the previous data set?

A: This varies between data services and datasets. Some data are collected periodically (e.g. survey data), and data services then offer access to each new collection alongside the others (for example separate data for the 2021, 2022 and 2023 iterations of the same survey), while other maintain and periodically refresh a single "live" version of the data (as OpenSAFELY does, with GP data).

Q: Should data that has a commercial aspect not be treated with more restriction than data that may be of a more general nature?

A: All data use is restricted to whatever the Data Controller is supportive of, and this typically varies according to the sensitivity and detail of the dataset. Where commercial organisations are permitted to use data through a data service, their applications are usually subjected to greater scrutiny (e.g. than those from academia) to ensure the work will be in the public interest.

Q: Where does research get published?

A: This varies according to the data service. OpenSAFELY publish all research here https://www.bennett.ox.ac.uk/opensafely/papers/. ONS require researchers to share details of where outputs are published with them, and then publish case studies of many of these here: https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice/researchexcellenceandpartnerships/researchoutcomes#have-you-reported-your-outputs-.

Pete Stokes Q&A

Q: Could you give an example of public benefit that a private company said they would achieve using the ONS data?

A: There are many different benefits, but an example is a project by Frontier Economics to evaluate the success of the TechNation Programme (which is taxpayer funded), which seeks to increase investment into the UK.

Q: Do you use AI to de-identify data, or is it used in any other part of the process with data?

A: No. Data are currently systematically de-identified by a member of staff, by removing risky fields (typically these include fields such as Name, Address, Date of Birth, NI Number)

Private & Confidential

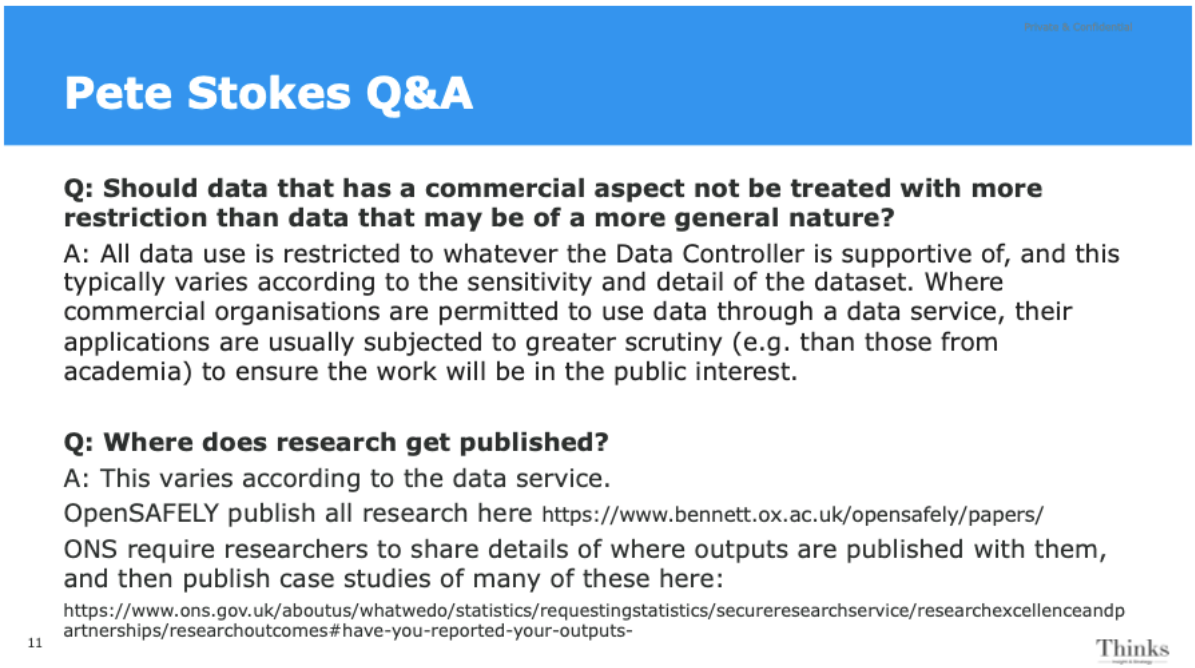
Pete Stokes Q&A

Q: How long does the data services hold onto data / how do you dispose of research data once researchers are done?

A: All data services will have policies on this, but datasets are typically kept for as long as there is a demand for them (as data will typically be used for many projects). A standard policy is that, if no researcher requests access to a specific dataset for 12 months, then it should be removed from the service.

Q: Does data get updated in the datasets when researchers are using it – and if so, what happens to the previous data set?

A: This varies between data services and datasets. Some data are collected periodically (e.g. survey data), and data services then offer access to each new collection alongside the others (for example separate data for the 2021, 2022 and 2023 iterations of the same survey), while other maintain and periodically refresh a single "live" version of the data (as OpenSAFELY does, with GP data).



Information we shared with participants to help them understand the different levels of public engagement

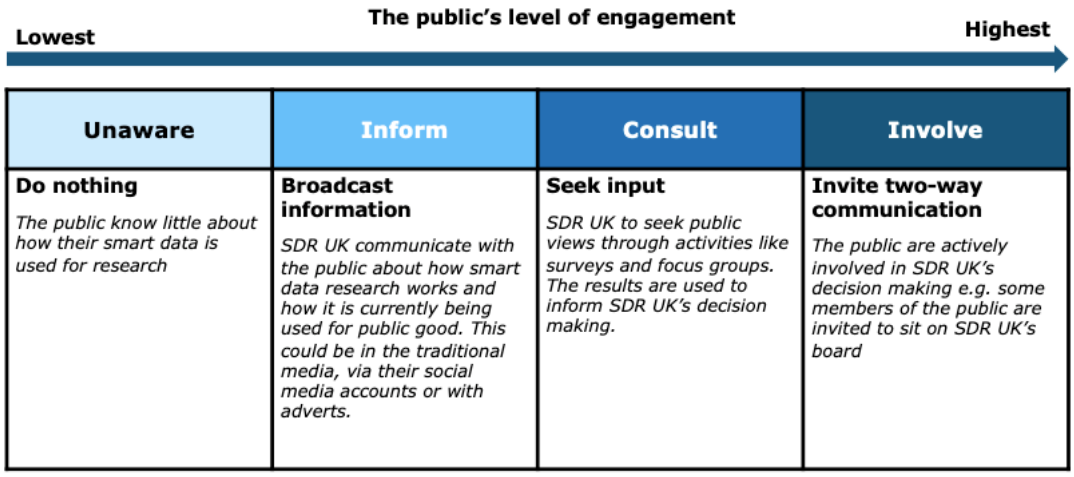
The public's level of engagement (from lowest= unaware, to highest=involve)

Unaware: **Do nothing**
The public know little about how their smart data is used for research

Inform: **Broadcast information**
SDR UK communicate with the public about how smart data research works and how it is currently being used for public good. This could be in the traditional media, via their social media accounts or with adverts.

Consult: **Seek input**
SDR UK to seek public views through activities like surveys and focus groups. The results are used to inform SDR UK's decision making.

Involve: **Invite two-way communication**
The public are actively involved in SDR UK's decision making e.g. some members of the public are invited to sit on SDR UK's board



Participant workbook

Page 1: Recap of SDR UK's role (can be seen in participant workbook for workshop 1)

Page 2: Recap of SDR UK's role

A private company*: *SDR UK calls them data controllers*

They collect smart data as part of their business activities. They decide to share smart data so it can be used by researchers working for the public good

SDR UK's role at this stage of the process:

- Talk to companies to explain the benefits of sharing their smart data with researchers for public good and commercial insight.
- Understand their needs, motivations and challenges

A data service: They work with private companies to take the smart data and turn it into something researchers can use (a data set) and set the rules of who can access it, when and how

SDR UK's role at this stage of the process:

- Fund the data services and ensure best practice standards are met
- Communicate their work to the public, researchers and government.

- Make sure they work together and share learning

Scientists and researchers: They access the smart data and use it to understand the big social challenges we are facing

SDR UK's role at this stage of the process:

- Ensure data services are meeting their needs in terms of data and training
- Encourage use of smart data for research that benefits the public
- Ensure publication of reports and findings
- Amplify the impact and influence of recommendations

Page 3: Benefits and potential harms of smart data research (see case studies below)

Page 4: Research for the public good

Key concerns and discussion points relating to smart data research and the public good:

- What the research is actually being used for and what impact it will have in the real world
- Who gets to decide what the "public good" is
- If "public good" always means benefitting the greatest number of people, or if it is also public good if it only benefits a few but addresses disadvantage or inequality, or, for example, a rare disease or condition
- How quickly you would expect to see a positive impact and whether public good could also be delayed, or achieved for future generations
- Whether you would expect to personally see a benefit from smart data research for the public good

Your rules: What do they mean in practice?

Rule 1: The data isn't biased or excludes people and outcomes are fair

Rule 2: There is a clear need for the research and a (potential) real-world impact

Rule 3: There is oversight and accountability to make sure research benefits the public good

Page 5: The role of private companies in smart data research

Key concerns and discussion points regarding commercial relationships for smart data research:

- You were sceptical about the motivations of private companies for getting involved in smart data research and many of you thought they were motivated by profit only
- You don't necessarily trust private companies to do anything for the public good – and if they do, you assume it would be a PR exercise

- You're not sure whether public good and private profit can go hand in hand
- There were questions about consent – most of you assumed you may have agreed to this use of your data for research via T&Cs but you weren't sure

Your rules: What do they mean in practice?

Rule 1: You wanted there to be clear oversight and accountability to ensure that research is actually for the public good, rather than for private profit

Rule 2: You thought there should be more transparency for consumers on the use of their data

Rule 3: There are certain organisations and types of data you are more comfortable with than others and rules should be stricter for sensitive data

Page 6: Data sharing and the role of data services

Key concerns and discussion points regarding the safe and fair handling of data:

- You liked the Five Safes (Safe People, Safe Projects, Safe Data, Safe Settings, Safe Outputs) but had questions about who enforces these
- And data services may be very safe but who is accountable if something does go wrong?
- You were worried about the security of the data when it is being transferred from private companies to data services
- Re-identification using multiple datasets did not feel like a big risk to you, but you still want to make sure that all identifiable information is removed before the data is shared

Your rules: What do they mean in practice?

Rule 1: Make sure the Five Safes are enforced consistently across data services and updated as risks evolve, especially accreditation and training to ensure only trustworthy people work with our data

Rule 2: Have clear accountability and fines/penalties for when something does go wrong

Rule 3: Introduce auditing of companies and researchers, with background checks to ensure no conflicts of interest

Page 7: Public involvement (see 'Information we shared with participants to help them understand the different levels of public engagement' above)

Page 8: Hopes and fears about smart data research from workshop 1

Your main hopes from workshop 1:

- That smart data research has a benefit for society e.g. improved public services, reduce inequalities and healthier lives

- Education for the public to better understand how data is used, which would improve public trust
- Greater control over how and when sensitive data is used
- That smart data research will support evidence-based decision-making and drive innovation
- That findings from smart data research will give people advice that helps them make better plans for their lives
- That smart data will be used appropriately and be secure
- That findings and outcomes of the research will be communicated to the public

Your main concerns from workshop 1:

- Issues related to data misuse e.g., to target people for scams, data leaks and hacking
- That actors involved in smart data research might not stick to the rules around de-identification and data storage
- Inaccurate or incomplete data leading to inaccurate research findings, or that researchers/companies might manipulate data for their own needs
- That data might be used for research that would negatively impact people, or might be used for private profit or 'corporate greed'
- That research findings won't have any impact

On our street

Number 3: Jazmin and Isaac

- Isaac is 13 and got his first smart phone in Year 7. Jasmin has set up parental controls on the device.
- Jazmin is concerned about Isaac's online safety and privacy, including what data is being collected and who can access it. She thinks the minimum age for most online services is too young for children to properly consent.

Number 18: Paul, Layla, and Finn the cat

- Paul used to work for the NHS but recently got laid off. Since his phone broke, he's been using the internet at the local library.
- Paul doesn't think much about his data and how it's used. He doesn't think there is much data out there about him.
- He and Layla often struggle to pay the energy bills. Paul saw a news article about smart meters being used in research to help local authorities target fuel poor households. He hopes this might help him and Layla.

Number 10: Gareth and Jo

- Gareth is a wheelchair user who finds connecting with people on social media and online gaming essential for his wellbeing.
- They often order groceries, clothing and other household items online. When Gareth needs to get around, he uses his bus pass or UberWAV.
- Gareth isn't too concerned about privacy or what happens with his data as he feels any risk is outweighed by the benefits.

Number 23: Billy and Barbara

- Billy and Barbara are retired. They both have smartphones to make calls, and a smart TV but don't stream any programs. They both have Facebook accounts, which they use with default privacy settings.
- Billy was the victim of identity fraud 4 years ago and lost some savings. They have now stopped using online banking.
- They are worried about who can access their data and how they protect it.

Number 25: Sarah, Jonathan, Ava and Zack

- Sarah is a researcher for a company that behind a popular dating app.
- She has a GPS, loyalty cards and only makes purchases using her debit or credit card. She does all her banking and life admin online. She has smart devices throughout her home.
- Sarah considers herself well-informed about how data is collected and used. She has a good idea of what her digital footprint looks like.