Smart Data Research UK

Smart Data Research UK Public Dialogue 2025

sciencewise

Thinks Insight & Strategy

Acknowledgements

Authors and project team

Thinks Insight & Strategy

Lucy Farrow

Julia Ridpath

Loulwa Al Rasheed-Wright

Teresa Kuhn

Grace Evans

Rebecca McFarland

Niamh Bull

Aleena Numaan

Smart Data Research UK

Joe Cuddeford

Mark Gardner

Sarah Shaw

Olivia Varley-Winter

Sam McGregor

Sciencewise

Fionnuala Ratcliffe

Graphic Science

Dom McDonald

Annabelle Byrne

Ben Johnson

Oversight Group

(Chair) Professor Alison Park, Economic and Social Research Council (ESRC)

Nick Bailey, Urban Big Data Centre, University of Glasgow

Stephanie Borthwick, Which?

Tim Davies, Connected by Data

Javier Ruiz Diaz, Digital Policy Consultant

Colin Griffiths, Citizens Advice

Roger Halliday, Research Data Scotland

Teodora Kaneva, techUK

Jared Keller, Independent Consultant

Yves-Alexandre de Montjoye, Imperial College London

Andy Morris, Walgreens Boots Alliance

Sofi Nickson, Office for Statistics Regulation

Kathy Peach, NESTA

Alison Preston, Ofcom

Octavia Field Reid, Ada Lovelace Institute

Fay Skevington, Department for Education

Pete Stokes, OpenSAFELY, University of Oxford

Richard Syers, Information Commissioner's Office (ICO)

Ros Williams, University of Sheffield

Contents

Acknowledgements 2

Foreword 4

Executive summary 5

Key findings 7

Core values 9

1. Introduction to the dialogue 10

2. Initial reactions to smart data and smart data research 16

3. Delivering public good 24

4. Data quality and inclusivity 38

5. Working with private sector data 46

6. Security, oversight and accountability 54

7. Public involvement 62

Foreword

At Smart Data Research UK (SDR UK), our mission is to unlock the power of smart data to improve lives, and we believe that strong, meaningful engagement with the public is key to ensuring we bring this mission to life.

This public dialogue is the first project of its kind for SDR UK. When we began, we made a deliberate choice to engage early – before our core data services were commissioned. While making some conversations more challenging, this has given us invaluable insights to shape the direction of Smart Data Research UK.

Throughout our discussions, we witnessed how public understanding of smart data research evolved. As participants learned more about its potential to improve public services and drive evidence-based policymaking, their enthusiasm grew. However, they also raised crucial questions about trust, transparency, and governance.

Three key messages emerged clearly. First, the public expects robust processes and safeguards around data quality and research integrity.

Second, transparency is a priority – including about private sector partnerships and how research findings influence policy decisions.

Third, we must better articulate what smart data research is, why it matters, and how it works in practice.

As we move forward, we're implementing several concrete actions.

- We're clearly articulating the robust data governance frameworks we use – like the Five Safes.
- We're establishing transparent oversight mechanisms and creating clear channels for public involvement.
- We're developing ways to explain our work (including technical concepts) that resonate more effectively with the public.

Perhaps most importantly, this dialogue has shown the immense value of involving the public in broad, contextual discussions about smart data research. Their insights have highlighted what truly matters – not just what we might assume is important. This will remain central to our approach as we build a smart data research ecosystem that serves the public good.

We are grateful to all participants for their thoughtful engagement in this dialogue. Your perspectives will help ensure SDR UK delivers genuine public benefit while maintaining the trust and confidence of the communities we serve.

Joe Cuddeford
Director, Smart Data Research UK

Executive summary

Context

Smart data is generated whenever we engage with the digital world – whether we're shopping online, using social media, or getting directions. When this data is used for scientific research to understand and address important social challenges, we call this 'smart data research'.

Digital technologies now provide researchers with many sources of data, which in turn provide new insights about society. Such research can address diverse issues such as disease prevention, combating disinformation, reducing geographic disparities, or tackling climate change.

Smart Data Research UK is the UK's national programme for smart data research. It provides secure data infrastructure and invests in partnerships between academics and companies that hold smart data, to unlock the power of smart data to improve lives. A key part of SDR UK's work is establishing 'data services' – infrastructure to safely manage datasets and provide access to diverse types of smart data for UK researchers.

Journey from data to impact

About this dialogue

To inform the development of its programme, SDR UK partnered with Sciencewise, an internationally recognised public engagement programme, to understand public views on smart data research. We commissioned Thinks Insight & Strategy to lead this dialogue, which explored how the public thinks smart data research should be conducted and governed.

This dialogue involved 16.5 hours of engagement with 72 participants through a mix of online and in-person workshops. A series of five workshops were held simultaneously across five locations in the UK (Belfast, Gateshead, Inverness, London and Newport) over four weeks between September and October 2024. The objectives of the dialogue were to:

1. **Understand people's hopes and concerns towards smart data research for public good and the values and principles that underlie them.**
2. **Understand people's perceptions of what constitutes 'public good' and their priorities within this.**
3. **Identify people's expectations for ensuring that collaborations between publicly funded data scientists and private companies align with 'public good' and ethical principles.**
4. **Understand people's expectations for the handling of their data by researchers and data services in line with their values and principles.**
5. **Use the insight from the dialogue to inform the development of Smart Data Research UK and its data services and their public engagement work in the longer term.**

Key findings

1 Members of the public were unfamiliar with smart data research but grew increasingly enthusiastic as they learned more about it.

Participants began with limited knowledge of smart data research and initial scepticism about their data being used constructively. Early concerns focused on personal data being shared when buying a product or service, and then being misused for marketing or fraud.

As they learned more throughout the dialogue, they became increasingly enthusiastic about the potential for their data to be used for research that benefits the public.

This finding suggests that SDR UK should provide clear communication about smart data research and how it can drive positive change.

2 Participants wanted smart data research to deliver real-world impact.

Participants wanted smart data research to tackle major national challenges including social inequality, health, economic issues, and infrastructure. Their understanding of 'public good' evolved to consider both scale (number of people affected) and size (potential for benefit) of outcomes. While initially focused on urgent challenges, exposure to case studies helped participants appreciate the value of longer-term research impacts.

Low trust in government and public institutions coloured these discussions and led to scepticism about whether research findings would actually influence policy.

This finding suggests that smart data research should demonstrably influence policy and deliver tangible public benefits, with SDR UK playing an active role in encouraging policymakers across the UK to use smart data and research findings.

3 High quality and inclusive data were seen as vital to ensuring smart data research benefits all communities.

When participants learnt about potential biases through incomplete or inaccurate data collection, they became concerned about how flawed datasets could lead to limited public benefit and inequitable outcomes.

Participants also learnt how researchers typically mitigate these risks by using multiple datasets and diverse data sources, which helps ensure more comprehensive and balanced research outcomes.

This finding suggests that SDR UK should implement robust processes to support high-quality, inclusive datasets and reliable research findings. Participants saw a key role for SDR UK in promoting best practices and ensuring research benefits communities fairly.

4 Participants were initially sceptical of private sector motivations, and wanted reassurances that research is done for the public good.

Participants were initially sceptical about private sector involvement but became more comfortable when public benefit was clearly prioritised and safeguards were explained. They wanted transparent private sector collaborations with independent oversight and clear accountability measures.

This finding also suggests that commercial benefits must align with public good, with clear processes to ensure this alignment from project inception through to publication.

5 Initial concerns about security were significantly reduced when participants learned about existing safeguards and oversight.

While security was initially a major concern, learning about the Five Safes framework significantly increased participants' comfort with smart data research. Their focus shifted from general security worries to wanting specific details about oversight and accountability measures.

By the end of the dialogue, several participants acknowledged that creating too many restrictions on how data can be used might lead to 'missed opportunities' for relevant data to be used.

This finding suggests that SDR UK should be clear about how sensitive data remains safe and secure when it is used for research, and clearly communicate how different actors work together to provide the safeguards and systems which prevent misuse of data.

6 Members of the public want to have meaningful involvement in shaping smart data research for the public good.

The dialogue revealed how public input can be most valuable in specific areas – particularly in determining research priorities and defining public benefit, while technical matters like data security might be better left to experts. The dialogue revealed that smart data research, while unique in some respects, shares many public concerns with other types of research data (such as administrative data or health data).

These insights can help SDR UK to refine the priority for future engagement: involve the public in questions and decisions where they can have the most impact, either through unique perspectives or strong public interest. Future public engagements could also be coordinated with other data initiatives.

Core values

Throughout the dialogue, four key values emerged as drivers of trust:

The dialogue process helped to establish the importance of these values, alongside the Five Safes Framework which guides SDR UK data services use of sensitive, de-identified or personal data (discussed further on page 48).

These values provide a framework for SDR UK and other actors in the smart data research landscape to evaluate their work and ensure alignment with public priorities.

1. Introduction to the dialogue

Introducing smart data

What is smart data and smart data research?

Smart data is produced through our daily interactions with the digital world, such as when we purchase products or services, use social media, or get directions. It is collected when we use mobile phones, wearable health devices, store loyalty cards and smart technology.

Smart data research happens when smart data is safely shared for scientific research, to help us understand important social challenges. Smart data research can cover a wide range of issues, from disease prevention to planning public services, understanding behaviours or tackling climate change.

Terminology: It may be useful for readers to know that the term smart data is used elsewhere with a narrower meaning. It is sometimes used to describe the secure sharing of customer data, upon the customer's request, with Authorised Third-party Providers (ATPs), as with 'open banking'.

What is Smart Data Research UK?

Smart Data Research UK (SDR UK) is the UK's national programme for smart data research. Its mission is to bring together companies, academics, innovators and governments to use data to improve lives through better understanding of the world and solving social and economic challenges.

SDR UK is publicly funded by UK Government through UK Research and Innovation (UKRI), and delivered by the Economic and Social Research Council (ESRC). The programme is part of the UKRI Infrastructure Fund, which supports the facilities, equipment and resources essential for researchers and innovators to do ground-breaking work.

The SDR UK programme has four main objectives, to:

- **Provide secure data access** by building long-term partnerships with data owners, and delivering secure and effective digital research infrastructure.
- **Be a trustworthy programme** for the public by demonstrating responsible research practices and championing public engagement in smart data research.
- **Build capability for cutting-edge research** by leading an interdisciplinary research community and solving methodological and technical challenges.
- **Generate social and economic benefits** by supporting research that addresses economic and social challenges and tracking and communicating the benefits of research.

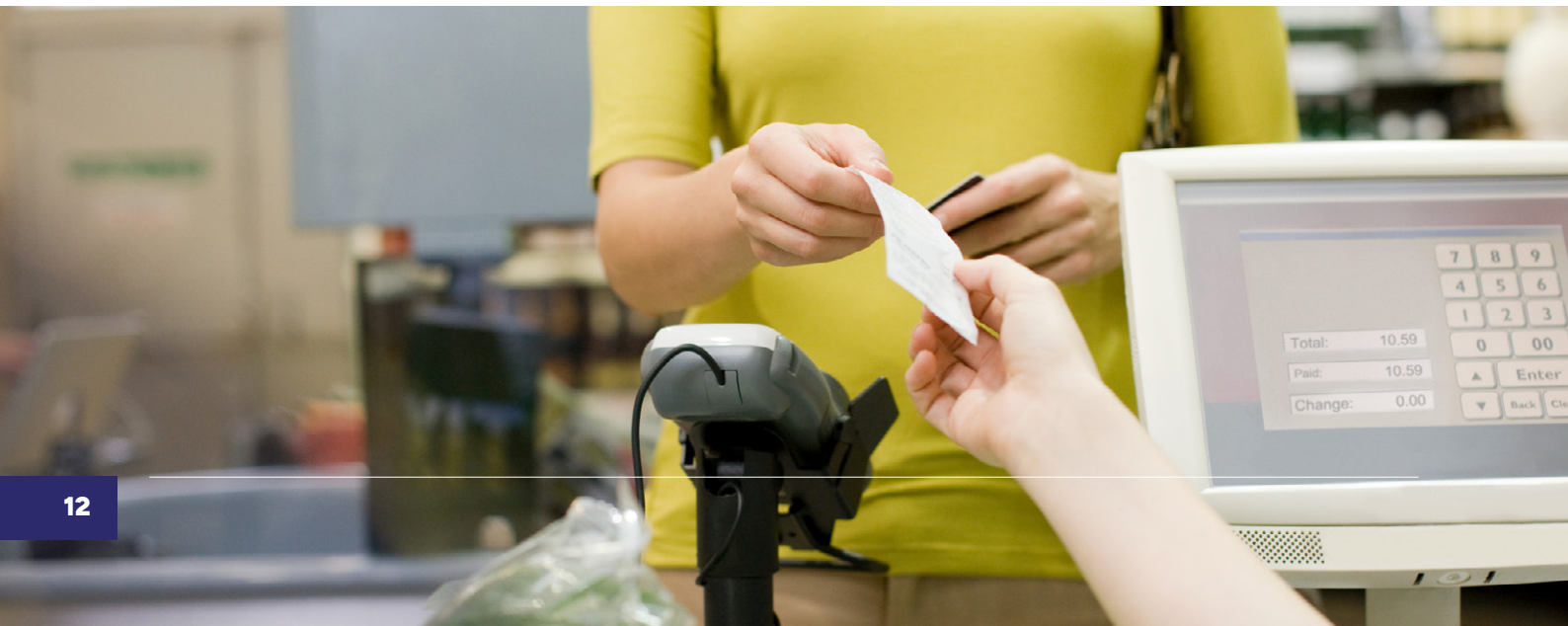
Introducing the dialogue

In March 2024, SDR UK, in partnership with UK Research and Innovation's Sciencewise programme, commissioned Thinks Insight & Strategy to run a public dialogue on smart data research.

Our purpose was to engage people from across the UK to understand what matters most to the public and ensure SDR UK's programme of work takes into account public views, hopes and concerns about smart data research. The dialogue had five objectives:

1. **Understand people's hopes and concerns towards smart data research for public good and the values and principles that underlie them.**
2. **Understand people's perceptions of what constitutes public good and their priorities within this.**
3. **Identify people's expectations for ensuring that collaborations between publicly funded data scientists and private companies align with public good and ethical principles.**
4. **Understand people's expectations for the handling of their data by researchers and data services in line with their values and principles.**
5. **Use the insight from the dialogue to inform the development of Smart Data Research UK and its data centres and their public engagement work in the longer term.**

The dialogue covered a broad range of topics associated with smart data research but it was not intended to be exhaustive. There remain many areas for further exploration for SDR UK to consider when planning further engagement with the public. These are highlighted throughout the report.



What is public dialogue?

Public dialogue is when the public and experts have conversations to help decide how to handle challenging issues.

Smart data research is complex. Data sharing can be controversial. It's important to go beyond immediate reactions and give people the time to understand things, hear different people's opinions, and come to informed views.

Who was involved?

This dialogue involved a range of people with different roles:

- **Public participants:** a total of 72 people from five locations across the UK (Belfast, Gateshead, Inverness, London and Newport). Recruited to reflect the UK population, including those with different levels of digital engagement and different attitudes to data sharing.
- **Specialists:** people with expertise in data, research and engagement who were invited to speak with dialogue participants to provide evidence, or a point of view, to inform participants' deliberation on smart data research

Professional recruiters (accredited by the Market Research Society) were used to recruit participants, who were broadly reflective of the general population in terms of demographic criteria. Quotas were used to ensure the diversity of the sample in terms of demographics, digital confidence and digital footprint (e.g. level of use of devices, apps and social media). These quotas reflected the intended sample and were largely met in the sample ultimately achieved.

A detailed breakdown of the demographics of those who took part can be found in Appendix A, which is on the SDR UK website.

What happened?

Participants took part in five workshops over the course of a month. These were split between full-day face-to-face and evening online workshops. Participants deliberated for a total of 16.5 hours.

Facilitators from Thinks Insight & Strategy led the workshops and group discussions with participants. They supported participants to deliberate on key topics and questions by sharing information and prompting discussion to understand their views. Facilitators referred to stimulus material and drew on specialists and/or attendees from SDR UK to remind participants of key concepts, answer questions and resolve any misunderstandings.

The table below provides an outline of each workshop. Further detail is in Appendix B, which can be found on the SDR UK website.

Table 1: Overview of the workshops

Workshop 1
Introduce participants and build understanding of smart data and smart data 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.
Workshop 2
Understand spontaneous views of what defines "research for the public good" and priorities for research. Explore the rules participants want to ensure research works for the public good.
Workshop 3
Explore participant views on private companies' motivations for being involved in smart data research, and the potential harms of sharing smart data. Discuss rules participants would put in place to ensure commercial relationships are fair.
Workshop 4
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 that participants would put in place to ensure data is handled safely and securely.
Workshop 5
Understand hopes and priorities for smart data research and SDR UK. Develop recommendations for the SDR UK programme and understand people's expectations for public engagement in smart data research going forward.

The public dialogue was guided by an oversight group. This comprised 19 stakeholders with expertise in data, research and public engagement (members of the oversight group can be found in the Acknowledgements). Their role was to provide challenge and support to the dialogue team in relation to the project framing and information provided. They met four times over the course of the dialogue (as described in Appendix A, which can be found on the SDR UK website).

Graphic Science Ltd, an independent evaluator, was commissioned to evaluate the robustness of the process of the dialogue and its findings, and its early impact on stakeholders.

Analysis

Each workshop was audio recorded and notetakers transcribed live notes during the workshops. Transcribed notes were added to an analysis grid. A coding framework was iteratively developed based on moderator brainstorm notes. Workshop notes were coded in the analysis grid using thematic coding to identify common themes, shifts in views and sentiment and areas of difference between participants. The Thinks Insight & Strategy team analysed the coding alongside the participant-generated recommendations to help develop and capture the key areas that participants think SDR UK should consider when developing its programme. Further details on the method of analysis are in Appendix A, which can be found on the SDR UK website.

About this report

The conversations from this dialogue have been summarised into the seven key themes below. Each chapter provides details on:

- **What participants said –** Including their views at the start of this dialogue, how these views changed over time, and where there was more or less agreement between participants. Where these chapters detail particularly notable changes in participants' views, this has been signposted with the annotation: changes in participant perspectives.
- **What this means for SDR UK –** Within each key theme, participants' values, priorities and concerns have been used to develop a reflection, to guide the work of SDR UK and other organisations within the smart data research landscape. This section also includes considerations on the implications of participants' views for SDR UK's work.

Important note on the scope of this dialogue

The research aimed to establish public attitudes towards smart data research as a starting point on a journey of public engagement for SDR UK. As such, the research takes a broad view of the benefits, challenges and trade-offs within smart data research rather than drilling into specific challenges in fine detail. Further engagement with the public should seek to explore any unresolved questions which have surfaced throughout this dialogue.

The findings of this dialogue represent the views of a group of people who are broadly reflective of the population. The sample is not exhaustive, so other views may exist, nor is it statistically representative and should not be taken to generalise the public as a whole.

This dialogue was commissioned during the early stages of SDR UK's development, before its data services received funding. As a result, some important themes and working methods were not covered in the discussions, such as the individual donation of personal data for research.

2. Initial reactions to smart data and smart data research

Participants' understanding of their own data sharing focused on deliberately shared personal data

Top-of-mind examples of data and data sharing

At the start of the dialogue, participants discussed their 'digital footprint'. They were prompted to think about the types of data that might be collected from them over the course of a normal week. Participants generally talked about websites and apps which ask for information about them when used to shop, access services and complete forms or tasks.

Examples given included sharing:

- Personal information to sign up to social media accounts.
- Financial information to shop online or use online banking services.
- Medical information to access NHS appointments or records.
- Location information in order to use maps or travel apps.

In the examples they raised, participants generally focused on the types of data they gave deliberately as part of buying a product or signing up for a service. Much of the data participants spontaneously discussed was personal, personally identifiable, and/or sensitive, which informed their initial reactions to smart data research.

Participants did not spontaneously mention types of data that would be collected more broadly as they went about their daily lives, such as bank payment data or loyalty card data.

These examples would be presented to participants later in the dialogue through case study examples.

> **"Passport details, location (data) from apps like Instagram or step counting. Amazon has my card details and address and phone number."**
>
> Participant, Inverness, Workshop 1

> **"I've had to sign into public Wi-Fi in the airport, sharing my age, address, email address, phone number, and what country I am travelling from. That was lots of my information. I had to register some new apps while I was away and put in my information."**
>
> Participant, Newport, Workshop 1

Differentiating between different types of data

Participants did not generally differentiate between different types of data about them, that might be used for research purposes. Throughout the dialogue, they referenced different types of data interchangeably, even when prompted by facilitators to consider smart data, specifically. This reflects findings from other dialogues on data, where the public typically views data as a large mass of information – which can feel complex, overlapping and intangible[1]. It also suggests that establishing a boundary between smart data and other types of data does not necessarily reflect the way the public perceives 'data'. In Workshop 1, participants were provided with the following definition of smart data:

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

Although facilitators shared an explainer of identifiable and non-identifiable data in workshop 2 (see Appendix B on the SDR UK website), participants did not always differentiate between identifiable personal data and non-identifiable data. They were concerned about the use of identifiable personal data for research purposes but were reassured when they learnt about the de-identification process for smart data research. However, as participants referred to both types of data interchangeably throughout the dialogue, they often raised concerns about identifiable data, even within the context of smart data research (where data is de-identified). Further detail on these concerns in the context of data security is provided in Chapter 5.

Participants referred to multiple types of data when discussing the potential impact of smart data research, even when prompted by facilitators or presented with case studies. For example, when discussing case studies from the health sector, participants referenced researchers using census data and medical records – as well as smart data from wearable devices and health apps. This reflects the reality of combining multiple data types for research, as was demonstrated in case studies presented to participants (e.g. a study exploring social exclusion by combining Housing Benefits data with data from the property website, Zoopla). However, this did mean that participants often referenced data not generated by digital interactions (such as data held by public services) when discussing the implications of smart data research.

> **"Gathering information on employment in areas where there is transport and not-easy transport, how could that get linked to allow people to access work to then access goods and the rest of it? You'd need to look at how people are transporting themselves to work, is it cars, trains, or buses? You'd pull that together to see if the economy of that area is a good one, and compare it to the types of work being done."**
>
> Participant, Inverness, Workshop 2

1. For example, a dialogue exploring public trust in data, conducted by Thinks Insight & Strategy on behalf of the Centre for Data Ethics and Innovation (currently known as the Responsible Technology Adoption Unit within the Department for Science, Innovation and Technology): https://assets.publishing.service.gov.uk/media/61eace40e90e07037d96983c/Trust_In_Data_-_Publishable_Report__1.pdf

Participants had initial concerns about data sharing

Participants raised two main areas of concern about the sharing of personal data between organisations:

1. Marketing and targeted advertising

Almost all participants spoke of their own experiences of unwillingly receiving targeted advertising based on their online behaviour (e.g. seeing advertisements for shoes after searching for shoes). Participants had mixed views on how and why this occurs, with a lack of clarity on the role of consent:

- At best, participants felt they had perhaps not realised they had failed to 'opt out' (e.g. by not reading terms and conditions in detail, not switching off tracking across their smartphone apps).
- At worst, participants suspected that their personal information had been 'sold' to third parties – with a lack of clarity on whether this is done legally.

> **"Marketing disclaimers that you just click to get through. No one reads the small print – you are giving an agreement without knowing it."**
>
> Participant, Inverness, Workshop 1

2. Scams and fraud

Participants frequently mentioned feeling worried about being a victim of online scams or fraud, based on the information they were sharing. This was particularly in relation to financial information they shared to shop or bank online.

Many participants mentioned hearing about instances of 'hacking' and theft of personal information, which made them feel that this was a significant risk associated with data sharing.

> **"The amount of online bank fraud is through the roof. I'm always wary of online fraud."**
>
> Participant, Gateshead, Workshop 1

> **"The NHS got hacked the other day, my son needs to wait longer for his blood test. That's really sensitive information about people's health."**
>
> Participant, London, Workshop 1

Older participants were typically more worried about these two areas than younger participants. For example, younger participants were often more accepting of the use of their personal data for marketing and targeted advertising. They saw this as a commonplace aspect of modern life, and were fairly comfortable sharing their data if it would lead to improved products, services or experiences. They also felt that they were 'savvy' enough to avoid falling for scams and broadly knew how to identify these.

> **"I've thought about [data sharing] but it's not a concern. If you're not doing anything illegal it shouldn't be a concern. I'm just worried about spam, I don't want emails and phone calls."**
>
> Participant, London, Workshop 1

> **"If someone hacks into your bank account there are protections but if someone gets access to your location how safe are you then? I'm worried about my daughter showing her location all the time."**
>
> Participant, Inverness, Workshop 1

Smart data research was a new topic for almost all participants at the start of the dialogue

The majority of participants were not aware of smart data, or smart data research, before taking part in this dialogue. They found it difficult to envisage what smart data was compared with other types of data, what smart data research might look like in practice, and how it might differ from other types of research.

Throughout the dialogue, participants continued to refer to multiple types of data when thinking about the potential uses and impacts of smart data research. They understood that only de-identified personal data would be used for smart data research when this was introduced in Workshop 1 and found this reassuring. However, this was not always recalled when discussing the risks and benefits associated with smart data research throughout the dialogue.

SDR UK enables smart data research via its data services. A wide range of data types, research topics and actors could be involved, and these were referenced in the materials presented to participants. Because SDR UK's data services weren't established at this point, it wasn't possible to present exactly what data would be used going forward. This uncertainty affected discussions more than people not understanding the topic. Important to remember when communicating about similar projects in the future.

Hopes and fears for smart data research

Participants were introduced to key concepts explaining smart data research in Workshop 1. They expressed some hope and optimism about its potential to benefit society. Participants also recognised that gathering information, such as smart data, leads to better understanding of individuals, society and issues facing the country. They were generally positive about increased understanding being an inherently 'good' thing in this context.

While some participants expressed hope and optimism, generally participants were not particularly excited or interested in smart data research. They didn't know much about it or how it could impact society (covered below in more detail). Questions remained about the extent to which smart data research could make an impact and lead to real-life changes for them.

The concerns they raised about data sharing – that it could lead to unwanted targeted advertising or even security breaches – meant they were also worried about the potential for this research to 'go wrong' or be used by 'bad' actors.

> **"It connects people's lifestyle choices and what they do – ailments and medications are dependent on different things. I suppose that information collected can be used for the good of the general population. So, for instance, if there is a lot of a certain type of illness in one area that comes through, that can be looked into."**
>
> Participant, Newport, Workshop 1

> **"I find it quite scary that they can know so much about us, without us giving them the information. I feel differently about data I provide versus what is collected unintentionally."**
>
> Participant, Inverness, Workshop 1

> **"When you get into insurance companies, you put your details online. Then the next week you get calls – they sell your information."**
>
> Participant, Gateshead, Workshop 1

SDR UK's role was not always clear

In Workshop 1, and throughout the subsequent workshops, facilitators shared information with participants about who SDR UK are, their role and how they work with other actors within smart data research. For example, the diagram in Figure 1 was used to outline SDR UK's role alongside private companies, data services, scientists and researchers.

It took time for participants to feel confident about each actor and their role in the process – and there were participants who were still unclear about this at the final workshop.

Some participants continued to describe a role for SDR UK that differed from the information shared (even when prompted by facilitators). For example, participants often assumed that SDR UK would play a regulatory type of role in overseeing companies providing data for research, and some came to see this as a gap that SDR UK should fill (see Chapter 5).

> **"How many private companies are involved? Are there any regulators of them?... I just assumed that there would be a Research Ombudsman that would make sure the data is being shared properly."**
>
> Participant, Gateshead, Workshop 1

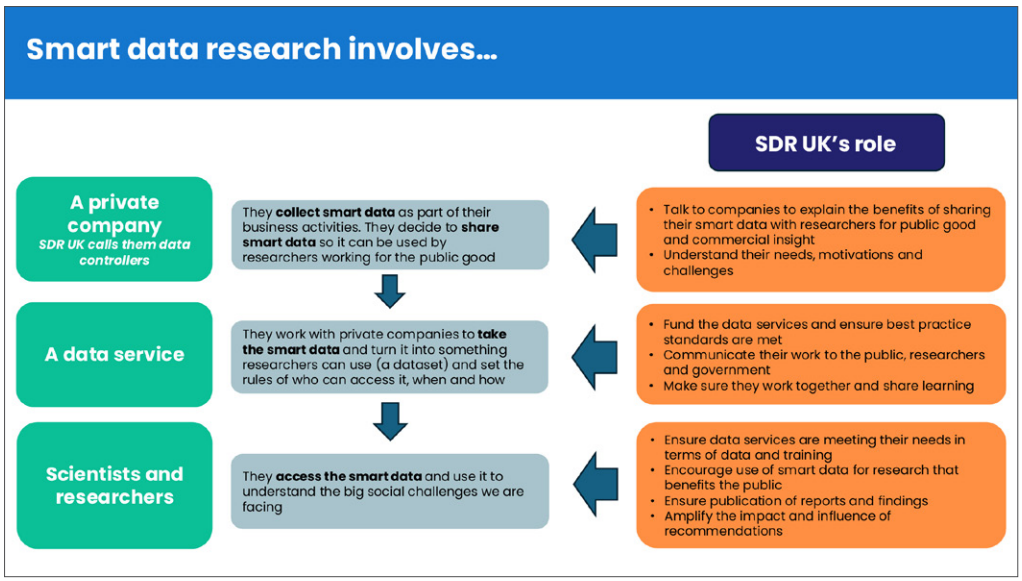


Figure 1. Diagram shown to participants explaining different actors in the smart data research ecosystem

Response from SDR UK

Smart data research is a new and complex area. This is why it was so important for us to do a dialogue at the outset of our programme, to help shape and influence SDR UK from the get-go, rather than asking people about things that were already well established. While this early engagement brings valuable opportunities, it also means some key components are still in development and so harder to grasp.

Throughout the dialogue, knowledge – or lack of it – is highlighted as a recurring theme. But as participants learned more, they became increasingly supportive of the potential for their data to be used for research that benefits the public. This points to a need for us to provide clear communication about what smart data research is, why it's conducted and how it works, to build trust and engage the public going forward. We must also think carefully about how we explain our role, making clear what we are (a family of organisations providing data services and support to researchers) and what we are not (a regulator or a private company).

Despite explanations across multiple workshops, misconceptions remained about some crucial distinctions: between identifiable and de-identified data and different types of data (smart data vs administrative data, etc.). This means we may need to reframe our work with the public in future by developing better ways to explain these technical and institutional distinctions to resonate more effectively. This may require us to move beyond traditional explanatory approaches and develop new frameworks, metaphors, or engagement methods that can better convey these differences.

3. Delivering public good

Definition

SDR UK seeks to deliver research that benefits the public. But they did not enter the dialogue with a fixed definition of 'public good' beyond the broad mission to bring together companies, academics, innovators and government to understand our world better and help solve social and economic challenges. They were keen to understand what the public thought about when they heard the words 'public good'.

People think about 'public good' in terms of national challenges

Throughout the dialogue, participants viewed and discussed public good primarily through the lens of major national challenges.

They focused on the public services underpinning – and struggling to address – these challenges. Participants frequently referred to current and ongoing challenges that have impacted them personally, as well as the country more broadly – such as the cost-of-living crisis.

There were three main areas that participants felt were priorities to address, to benefit the public good:

- **Health:** The area mentioned by almost all participants in relation to improving people's health or 'saving lives' and addressing NHS pressures, such as long waiting lists.

> **"Healthcare and medicine at the top because research and that sort of area leads to treatments and vaccines and good healthcare practices."**
>
> Participant, Inverness, Workshop 2

> **"Using the data to help with health services is the most important thing that smart data could be used for."**
>
> Participant, London, Workshop 2

- **Economy:** Mentioned by most participants as essential to improving people's standard of living and opportunities. Within economic growth, participants also spoke about the importance of education and skills development. They felt that improvements in the economy would have a 'knock-on' effect on improving services across the board (e.g. by increasing tax revenue which can be used for funding).

> **"Socioeconomic growth is the driver for everything."**
>
> Participant, Belfast, Workshop 2

“Looking at which areas have high unemployment and need either training or help with employment. Using a skillset in order to improve the economy.”

Participant, Gateshead, Workshop 2

- **Infrastructure:** Mentioned by most participants across all locations, with an emphasis on transport, housing and public facilities (e.g. libraries, recreation). Frequently associated with broader local area investment and economic growth.

“In Northern Ireland, the infrastructure isn’t very good or public transport. Here you can’t get a bus past 11 o’clock. Improving this would be a benefit for the economy as well.”

Participant, Belfast, Workshop 2

Research pillars

In Workshop 2, participants were provided with information about SDR UK’s four smart data research opportunities – or ‘pillars’ (see Figure 2 below):

- Productivity and prosperity
- Health and wellbeing
- Digital society
- Sustainability

Figure 2. SDR UK’s four research opportunities, shown to participants in Workshop 2

The priorities raised spontaneously by participants in their definition of public good were consistent with the first two pillars. When they were then prompted with information about SDR UK's four pillars, they therefore felt it was clear how smart data research in these areas could benefit the public good.

When people discussed public good, they didn't bring up digital technology or sustainability much on their own. While a few participants raised environmental issues as important to consider, this was rarely considered a higher priority than health or the economy.

Inequality central to public good definition

Participants consistently spoke about addressing inequality, ensuring that all members of the public were given equal opportunities, benefits and outcomes. They felt this was an important part of public good.

Participants drew on their own experiences, and those of others in their workshop groups, to reflect on differences for people living across the country. For example, differences in access to health services, levels of deprivation and availability of transport links.

> **"This is research that helps the public at large. Use the data to ease inequality like health disparities in two different areas. Use the data to make improvements to society."**
>
> Participant, Newport, Workshop 2

> **"I've been through cancer myself. I've got to travel a hundred miles to get an X-ray, so I'd like to see data collected for people with cancer needs, hospital needs and equipment to help people in [remote] areas."**
>
> Participant, Inverness, Workshop 2

Concerns about inequalities were particularly prominent in discussions amongst participants living outside London. These participants expressed frustration with regional differences in public services and infrastructure. A few expressed a view that central Government does not consider or understand the needs of those living in the devolved nations and northern England.

> **"There's a project to increase transport within the North East...and having that research to say maybe where there's inequalities in [accessibility] and obviously that would then have an impact on being able to access shopping, fuel, things like that."**
>
> Participant, Gateshead, Workshop 2

Benefits for the public good should translate into real-life impacts

Participants typically described benefit for the public good as 'improvements' across key areas of challenge – such as health, the economy and local infrastructure.

They wanted to see any initiative focused on benefitting the public good resulting in 'better' experiences for people (i.e. the 'public') in real life. They expected smart data research to focus on the causes and impacts of these types of challenges, to inform and support improvements for people.

These discussions reflect findings from wider public engagement on research for the public good[2].

Examples of suggested areas for research across participants' three key priority areas included:

- **Health:** Increased life expectancy or quality of life for people suffering from illness, quicker or easier access to a GP or hospital appointment.
- **Economy:** More job opportunities, more 'money in your pocket', higher attainment levels for school children, fewer people using food banks.
- **Infrastructure:** More public transport options, cheaper transport, reduced congestion in town centres.

> **"Maybe having a bit more money in our pockets at the end of the day...And in schools, for example, if there's kids with certain needs, there's enough staff to deal with it so the school is not struggling to cope with certain issues."**
>
> Participant, Newport, Workshop 2

While participants consistently raised similar priority areas for the public good (as outlined above), a few noted that 'improvement' could mean different things for different people. For example, different people in society are likely to have different needs, priorities and expectations, based on their own personal circumstances and attitudes.

> **"I think it would be different things for different people. Not everyone wants or needs the same thing. What is beneficial to me may not be for everyone else. There's going to be a lot of different scales for different people."**
>
> Participant, London, Workshop 2

2. OSR & ADR, A UK-wide public dialogue exploring what the public perceive as 'public good' use of data for research and statistics (2022)

Scale vs impact: Balancing growth and public value

When defining the public good, participants recognised that there was no easy answer to questions about who should be prioritised. At the start of the dialogue, most felt that research should benefit the majority of the population. This reflected their view that the public good should address large-scale challenges for real change.

> **"It would help the public at large rather than one small section of the public."**
>
> Participant, Newport, Workshop 2

Changes in participant perspectives: Views changed over time. By the final workshop, many had adopted a broader definition of who should be served by the research. This was influenced by discussions about the severity of a particular challenge on people. Take health research – participants were comfortable with benefits for a small group of people (e.g. those suffering from rare diseases) if it meant reduced suffering or saving lives. Targeting even a small group of people in this context would have broader benefits (e.g. families and carers of ill people).

This was also informed by discussions about inequality, where participants noted that addressing these issues requires targeting research to understand the needs and experiences of particular groups of people (e.g. certain demographic groups or communities).

> **"It is better to have a bigger difference on a small number of people. This will help them to improve their lives."**
>
> Participant, London, Workshop 2

Unresolved questions

While participants increasingly embraced a more inclusive view of who benefits from public good research, several points of disagreement remained. Most accepted that benefiting a smaller population could qualify as public good, but debates persisted about the required magnitude of impact. Participants struggled to reach consensus on three key questions: how to measure the severity of impact across different contexts, what minimum group size could justify research investment, and how to balance serving specific populations versus the broader public. These tensions highlighted the ongoing challenge of operationalising research for the 'public good'.

Changes in participant perspectives: Participant views changed about when they thought impact should be realised. At the start of the dialogue, most wanted smart data research to lead to 'immediate' or short-term impacts on people's lives. This was often because they felt that current challenges are so severe that change needs to happen now (e.g. in relation to cost of living).

Over the course of the dialogue, participants increasingly accepted that smart data research could lead to longer-term improvements. They heard about case studies and how the academic research process worked. This prompted people to recognise that it takes time for real-life change to 'filter down'.

Case studies about medical research were particularly convincing in highlighting the importance of conducting research now on medicines or illnesses to benefit future generations.

Participants also discussed areas where future impact would be more acceptable – and feasible – such as infrastructure developments, which take time to complete.

> **"I have a slightly better understanding of what public good is, the research that you are doing has to have some real-world impact at some point, obviously not straight away, you need time to implement it. Not every single time will a goal be reached."**
>
> Participant, London, Workshop 5

> **"With medical research for the public good, it would affect future generations but not necessarily right now."**
>
> Participant, Newport, Workshop 5

Unresolved questions

While participants acknowledged the value of research with longer-term benefits, they grappled with how to weigh these against immediate needs. Those who supported future-focused research often cited intergenerational benefits, particularly participants with children and grandchildren. However, two key questions remained unresolved: under what circumstances long-term research should take precedence over short-term improvements, and which research areas warranted such a long-term approach.

Real-life impact: a counterbalance to concerns about data sharing

The case for making smart data available for research rests in large part on the public good that comes of it. It's the social licence for data sharing to be undertaken.

During Workshop 1, participants learned that scientific research deemed necessary and in the public good can reuse people's data without explicit consent (see Figure 3). This information aligned with participants' broader concerns about their limited control over personal data collection and sharing.

Changes in participant perspectives: Participants were initially surprised to learn that data could be used beyond its original purpose – a practice most were unaware existed. However, their perspective shifted as they considered the potential real-world benefits of smart data research. This potential helped balance their concerns about data control, security risks, and commercial interests (discussed further in Chapter 4). When introduced to public good as a criterion for research approval, participants found this framework strongly reassuring and came to view it as an essential safeguard in the data-sharing ecosystem.

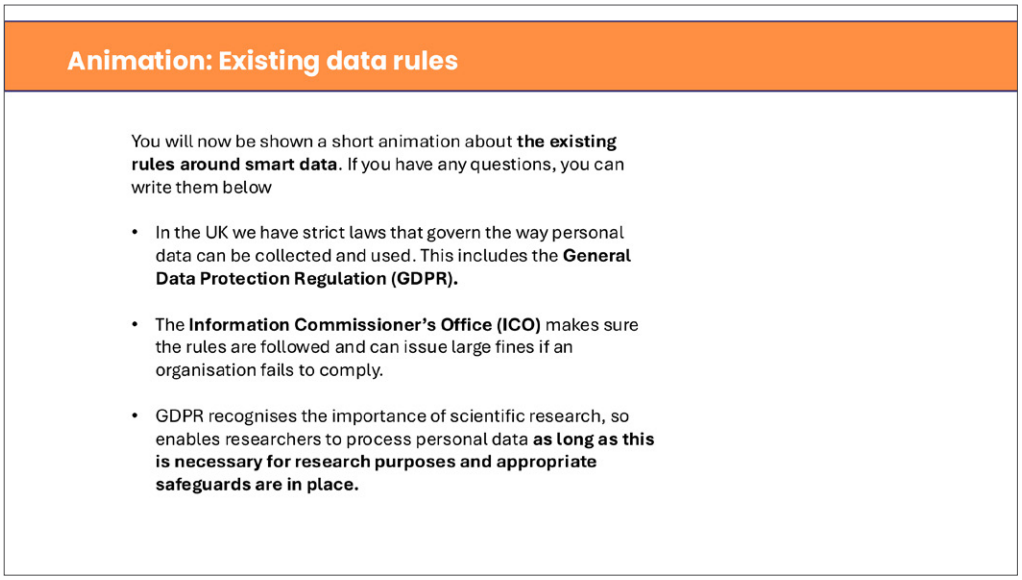


Figure 3. Explanation of existing data rules provided in participant handbook and animated film

Low awareness of research and lack of trust in government

There was low awareness of publicly funded research at the start of the dialogue. Participants did not have much prior knowledge about the aims of this type of research. What and who it involves. The systems in place to facilitate and manage it. It was also not clear how publicly funded research would be used to inform decision-making – and therefore, how it could affect the lives of our participants. This meant people were not very excited about smart data research, indifferent to its potential to make a difference.

> **"It goes back to the fact we already know there's a problem. Why do we need more data to tell us there's a problem? It's not going to fix anything."**
>
> Participant, Inverness, Workshop 1

Throughout the dialogue, and particularly when discussing public good, participants expressed low levels of trust in public institutions and government. They didn't believe these institutions had the ability or inclination to ensure smart data research led to positive real-life impact. Participants typically referenced the term 'government' when discussing policymakers. The term was often used as a catch-all to describe decision-makers across local and national government, and public institutions, such as the NHS.

In expressing this distrust, participants referenced broader events from recent years, such as the UK political context, COVID-19 pandemic and cost of living crisis. These were seen as failures to protect public interests in decision-making.

This dialogue included people from Belfast, Gateshead, Inverness, London and Newport. Distrust was particularly strong amongst participants living outside London. Participants in these areas were more likely to distrust government, and feel 'left behind' or ignored by public institutions (e.g. in policy decision-making and investment). They were often more cynical about the motivations of those involved with smart data research, and their ability to create positive change. They doubted smart data research could have a significant influence on decisions.

> **"The difficulty for us comes from seeing them, the government, abusing [data] in the past. Using it for political and personal gains. For me, I would be supportive if it was doing the things it's suggesting it wants to do. Because, when you look at it historically, things haven't changed, people who live in areas of social deprivation are still unwell or don't have opportunities. What's going to change with this data that hasn't before?"**
>
> Participant, Inverness, Workshop 1

This distrust was reflected in three main concerns voiced by participants about government decision-making. These concerns underpinned doubts about the likely impact of smart data research:

- **Government and policymakers are not using evidence to make policy decisions at all:** There was a sense amongst participants that government doesn't use any evidence to make big policy decisions. Participants believed that politicians took decisions based on their own opinions and then would "find" data or research findings that supported their point of view.

> **"I wish the council and government would actually use it. The people who should be using it and learning and making changes."**
>
> Participant, Newport, Workshop 1

- **Government can't use data effectively when making decisions:** Participants expected that the data that could be available to policymakers is not being used properly. There was a perception that a lack of joined-up thinking in government prevents data from being used in the right way to make data-driven policy decisions. Many participants questioned whether data and research findings were properly understood by policymakers, or were not being used in the right way to actually improve the public's lives. For example, they perceived a lack of effective data use in public health decisions, particularly during the COVID-19 pandemic. If research was being regularly used to inform policymaking, then this was not clear or well understood by the public.

> **"What's the point of collecting all that data on COVID if [the Government] didn't use it"**
>
> Participant, Inverness, Workshop 1

- **Misuse of data in government:** A few participants across the locations were concerned about government misusing data. They thought that smart data research could potentially be used to control or discriminate against people. Used by political parties to benefit their own motivations, rather than being used for the greatest public benefit.

> **"How can I trust the man not to get information and use it against me and apply restrictions to me? How do I know it will be interpreted correctly, it's open to abuse. Do we all trust our government these days?"**
>
> Participant, Belfast, Workshop 1

Changes in participant perspectives: As participants learned more about the opportunities for smart data research, they became increasingly excited and enthusiastic about the potential for research findings to positively impact policy decisions. Sharing tangible examples (beginning in Workshop 1) of where research led to a positive public impact or influence on government decisions increased confidence in the ability of smart data research to have influence
and impact.

> **"Social issues should be prioritised, you can justify to policymakers why they should be making changes to social issues if you have smart data research."**
>
> Participant, Inverness, Workshop 2

Participants thought that smart data could be used by policymakers to better understand patterns, which could be used to improve public services. They thought this would be of particular benefit in more deprived areas and for people who experience inequalities. They saw potential for government at a national and local level to use smart data to improve the lives of its population by identifying population groups, or geographic areas in particular need, for example, related to energy needs, transport or health issues.

> **"Identifying vulnerable people and having the resources within the local areas, the local governments that can softly enter those people's lives and help."**
>
> Participant, London, Workshop 1

Changes in participant perspectives: After learning more, support for smart data research strengthened. As support, confidence and enthusiasm grew, so did participant sentiment that smart data research should be used by government and other public institutions. Participants saw the public good as an important criterion for their smart data to be used. They wanted clarity on this to balance out their perceived lack of control over their data (mentioned in Chapter 2).

Participants recognised that SDR UK is a publicly funded body, with funding directed towards supporting smart data research projects. Nevertheless, participants often felt that their data is being 'taken' or used without their full consent. Nevertheless, they believed that if it was being collected anyway, it should directly inform policymaking. They emphasised that failing to share research findings with policymakers would waste public resources.

> **"It would be a waste of money if it was not put into policy and not on politicians' agenda, if all this money and effort is put into this research and nothing comes out of it."**
>
> Participant, Inverness, Workshop 5

Scepticism remained that smart data research will change anything

Participants continued to express a lack of trust in policymakers using the research findings for public good, or without 'spinning' findings to suit a pre-determined narrative. Some felt that rather than using the findings to shape policy and decisions, government would use or manipulate the framing of findings to support policies that are already established.

> **"NHS is an example because it is too hard to make the changes that need to be made. The government is more focused on getting votes."**
>
> Participant, Inverness, Workshop 2

Participants were concerned that smart data research, even with the best intentions, would have little positive impact on the public if it's not followed up with action from policymakers. They were keen to see research with clearly defined policy actions, rather than academic papers that they feared would only sit on the shelf. They wanted policymakers to take findings seriously and use them in their decisions and planning. At the end of Workshop 1, a few participants spoke about their hopes that smart data research could allow government to take a more proactive, rather than reactive approach to policymaking.

“They should be proactive and not reactive. For example, if there is a pattern of people with more untreated cancer in Newport, then the government should be encouraging people there more to get screenings.”

Participant, Newport, Workshop 1

To alleviate these worries about the potential impact of smart data research, participants felt that there needed to be greater effort to put research in front of policymakers. They also felt that more could be done to encourage policymakers to use smart data research to inform their decisions, for example by expanding their knowledge on how it could be used and its potential impact. Further discussions with the SDR UK team in Workshop 5 uncovered a role for SDR UK to encourage policymakers to proactively seek out research when faced with a complex challenge to solve.

Response from SDR UK

Participants wanted smart data research to tackle major national challenges including social inequality, health, economic issues, and infrastructure. Enthusiasm about smart data research’s potential benefits grew after participants learned more about how it could identify patterns and improve public services.

However, there was a complex relationship between participants’ views of smart data research and their trust in institutions such as government. Their growing appreciation for the benefits of smart data research was tempered by concerns about whether, and how, public institutions would use this research in practice. Participants were sceptical about whether:

1. Research findings would genuinely influence policy decisions rather than being used selectively to support pre-existing positions
2. Benefits would be equitably distributed across regions, given existing disparities in public services and infrastructure
3. Government at both national and local levels would follow through with meaningful action based on the research

This points to a need for us to play an active role in encouraging policymakers across the UK to use smart data and research findings. We will implement clear reporting on what we are doing to get research in front of policymakers in government and local government, and how data and research findings are used in policymaking.

It will also be important to ensure distribution of research access and outputs across regions. This is in fact part of our founding mission of expanding access to smart data across the UK.

4. Data quality and inclusivity

Participants lacked knowledge about how research works

Before taking part in the dialogue, participants' personal experience with research were limited to things like observing the chief scientific advisor throughout the COVID-19 pandemic, having a research department at their place of work or having friends and family who work in a university.

Research – and smart data research more specifically – generally felt abstract. They found it difficult to picture what the smart data research process would look like in practice for researchers. They had questions about what researchers would do with smart data, and how they would be vetted or authorised to use it.

Participants generally had more trust in researchers to be 'genuinely' interested in benefitting the public good, compared with private companies. However, they had questions about the processes in place to ensure this – and that the work done by researchers would actually support the public good.

> **"It says approved researchers, who is going to approve them? And what is the criteria? Is Tesco an approved researcher?"**
>
> Participant, Newport, Workshop 3

> **"The researchers should be vetted as they are going to have the information. It should be made sure that the researchers are genuine."**
>
> Participant, Newport, Workshop 3

> **"Researchers are more trustworthy than politicians. They only care about the data, they're not going to skew it, their agenda is only to look at the data and what they find is what they find – they're more trustworthy to me."**
>
> Participant, London, Workshop 1

Learning about data bias inspired participants to think about data quality

During the workshops, participants initially focused on data security concerns, paying less attention to issues of data quality and bias. It wasn't until prompted near the end of Workshop 1 that they began considering problems of data exclusion. This exploration of data quality and bias risks in smart data research revealed that these critical issues weren't naturally top of mind for most participants.

Initial thoughts and concerns about data quality came from the broader concern about accuracy of the research. For example, making sure everyone was 'counted' in the data – especially those who were digitally disengaged.

> **"It would be interesting to know [if local councils are using this data for planning] if it improves safety in a high-capacity area. If it's something not a lot of people will use it's a waste. Could be helpful to track population movement but it needs to be accurate."**
>
> Participant, London, Workshop 1

> **"[In response to Persona: Number 23 – Billy and Barbara] My fear would be that they get left behind. They release some of these benefits we get online, such as shopping discounts or things. If they don't keep up with the times, they might get left behind."**
>
> Participant, Inverness, Workshop 1

During Workshop 2, The Alan Turing Institute's Maxine Mackintosh, who leads the Diverse Data programme at Genomics England, explained how smart data research could inadvertently exclude certain populations without proper management (see Figure 4).

Figure 4. Explanation of how bias might enter data-driven research, provided by Maxine Mackintosh in Workshop 2

Maxine's presentation developed participants' views on the notion of data quality, encouraging many to think deeply about data bias and associated issues. For example:

- Excluding people through the types of questions that research looks at, if it focuses only on where there is the most amount of data
- Missing trends across people who are not well represented in data e.g. people from minority ethnicities being grouped together in data, or people who are non-binary being excluded in categorisation of data.

From this point in the dialogue, participants would often refer to Maxine's presentation and would bring up issues of data quality and bias frequently when discussing different elements of smart data research.

> **"I found it really interesting as well, especially how common data bias is. I had no idea that certain people were excluded, I just thought that it was designed to include everybody. I thought that it was good that the smart data is for a broader spectrum of people, from different backgrounds and religions."**
>
> Participant, Inverness, Workshop 2

> **"What I heard was a mirror to what we were talking about in the previous section. The data that is there is inaccurate, and it's an issue I hadn't thought of. The people who may be in the minority or have different opinions are not listened to. The majority are normally focused on."**
>
> Participant, Newport, Workshop 2

Changes in participant perspectives: After Maxine's presentation, participants were reassured that their initial concerns about the inclusion of digitally disengaged people had been recognised by researchers, and they were seeking to address it.

> **"Reassured. It is good that the biases are being taken into account and the researchers are aware of that."**
>
> Participant, Inverness, Workshop 2

With new information about how bias can define datasets, participants began to express concerns that minority groups would always be excluded. Some described it as 'typical' that these individuals wouldn't be counted in smart data research and there was a sense of pessimism about what could be done to plug gaps. Concerns about data bias were especially prevalent amongst participants in London.

> **"Unsurprised. Due to the work I do, it is like society doesn't see these particular groups."**
>
> Participant, London, Workshop 2

Following Workshop 2, participants consistently applied their new understanding by examining data integrity in the case studies and considering potential impacts on minority groups. They emphasised that high-quality data should accurately represent all users and explicitly identify any demographic gaps in the dataset.

Participants saw inclusion as vital to the integrity of research

Participants worried that the exclusion of key groups would result in incomplete and incorrect research. Concerns about exclusion often focused on people who did not have access to digital devices or were not online (such as children and the elderly). Finding a way to include these people, or at least mitigate the consequences of any exclusion, became more of a priority as the dialogue continued.

Geographic differences also became apparent here. Participants in Belfast and Inverness raised concerns about physical locations with poor internet connections, where apps often don't work or there is a lack of digital infrastructure. Participants were conscious that these areas might be 'left behind' in research.

“They would have to notice that there are pensioners and young people who don’t use these digital societies. So even though there is lots of data, there is also some groups it doesn’t include. I think the majority of the UK have access to the internet, but there will be pockets in some areas where the internet isn’t very good.”

Participant, Newport, Workshop 2

“There’s 10 million in London, 5 million in Scotland, if you look at everyone, it’s always ‘London, London, London’ so we lose out, because nothing happens in the north.”

Participant, Inverness, Workshop 5

There was interest in using diverse datasets to minimise bias

As the dialogue continued, participants increasingly mentioned the importance of using multiple data sources to combat ‘bad’ or incomplete data coming from one company. They thought that, in general, smart data research projects should avoid working with single data sources to minimise the risk of bias. Participants were not concerned about risks of re-identification in combining datasets (see Chapter 6). Many supported the idea that large datasets should be made available or purchased to plug data gaps. In the context of this conversation, they also saw a role for SDR UK in increasing participation in data sharing, both from individuals and from private companies, to broaden datasets.

“Try and get data from more than one source.”

Participant, London, Workshop 5

Participants called for oversight to ensure high-quality datasets are available

Participants were concerned that bad 'quality' or biased data could lead to unintended harms. They worried that:

- Researchers would not consider or care that their datasets were incomplete.
- Researchers would not consider that key groups were missing from their datasets when reporting their findings.

They argued that this would result in missing perspectives and behaviours of key groups in research findings, and therefore from decision-makers thinking.

Participants therefore returned to the idea of oversight and accountability, calling for checks that datasets are not missing key groups. They felt an "independent body" would be best able to fill this role.

> **"Are they collecting the correct information? They get so much money from the government to do this project, then what happens? It's assessing value for money."**
>
> Participant, Gateshead, Workshop 5

> **"They might not be the correct data without SDR oversight"**
>
> Participant, London, Workshop 5

> **"Bias could be harmful, if certain people don't participate in the data collection, then it would not be accurate."**
>
> Participant, Newport, Workshop 5

> **"They should ensure the data they are analysing is up to date."**
>
> Participant, Gateshead, Workshop 4

Response from SDR UK

It's clear that our participants want processes in place to support high-quality datasets and research findings. They saw risks around data bias and exclusion. Participants want SDR UK and researchers to be transparent about data limitations while actively working towards better, more complete datasets. Independent verification was another priority for some participants, to identify where datasets are incomplete or show bias against certain groups.

While these issues are not unique to smart data research, we have a role to create the right conditions for high-quality research through our data services. We'll do this by implementing strong principles across our data services to ensure that smart data research produces reliable research findings.

Robust ethical and accountability mechanisms already exist within the UK research ecosystem. Researchers are challenged and held accountable for their outputs through established processes such as peer review, where methodology and data representativeness are critically evaluated. Research institutions, funding bodies, and publishers also maintain ethical standards and data quality requirements. Rather than creating a new independent oversight body, we will focus on improving transparency around dataset limitations and promoting best practices, encouraging researchers to proactively address potential biases and gaps in their datasets as part of good research practice.

5. Working with private sector data

Participants were initially sceptical about the motives of private companies in smart data research

Participants came to the dialogue with certain expectations about how private companies behave. These were based on a general perception that private companies' main interest is to generate profit, given shareholders expect a financial return.

This meant participants began the dialogue thinking that private companies would seek to profit from the use of the smart data they collect; that commercial interests would be at odds with SDR UK's ethos/desire to use smart data for national interest, unless companies deliberately set themselves up to do things differently.

> **"Private companies, I think they're selling your details to other companies, not all of them."**
>
> Participant, Gateshead, Workshop 1

Participants' mistrust in private companies' use of their data stemmed from:

- **A perceived lack of transparency in the consent process:** Participants spoke about the barriers to not sharing data, such as accepting cookies and consent "hidden" in terms and conditions which are not transparent or accessible. They reflected on their experience of accessing websites where they felt they were forced into accepting cookies to be able to access the site. Many participants shared their belief that companies purposefully make it difficult for the public to understand the consent process, and almost impossible to engage in some services without providing consent for data sharing.
- **Experience of targeted advertising or scam calls and emails:** Previous experience or awareness of scams resulted in participants suspecting that this was the result of their data being "sold" by private companies.

"I think whenever I look online I get bombarded. I was looking at a holiday and had umpteen emails about holidays, how do they know that?"

Participant, Belfast, Workshop 1

"I've been hacked as well – last week found 120 messages from my phone to my contacts asking for money, on Facebook and WhatsApp, had to cancel everything and change all my passwords."

Participant, Inverness, Workshop 1

Due to their concerns, participants started the dialogue worried that private company involvement in smart data research might go unchecked. They were most concerned about:

- **Biased findings:** Some expressed concerns that researchers could be influenced by commercial partnerships (e.g. funding), so the findings from projects commercially benefit a company rather than benefit the broader public.
- **Security risks in sharing their data:** Many assumed their data would be insecurely shared with researchers, and without their consent, as they suspect is already happening between private companies.

"I'm not 110% convinced that universities are the best place for this data analysis to happen, they have their own biases..."

Participant, Belfast, Workshop 2

Changes in participant perspectives: For most participants, their strength of feeling and sense of acceptability around this area did shift over the course of the dialogue. They became more pragmatic about the needs of private companies to have some benefit commercially for them to spend time and resource being involved in smart data research.

"A business is a business, they need to make money to survive"

Participant, London, Workshop 5

"It's not a problem if they get a profit out of it, as long as something good comes out of it as well.

Participant, Newport, Workshop 5

Participants learned about the Five Safes framework (further explored in Chapter 5 and outlined in Figure 5) and this led to greater confidence that:

- Projects that are publicly funded would have to have a public benefit (safe projects)
- The data that is used would not be personally identifiable (safe data)
- The data that is used will be held safely and securely in data services, with restrictions around who can access it and for what purpose (safe settings).

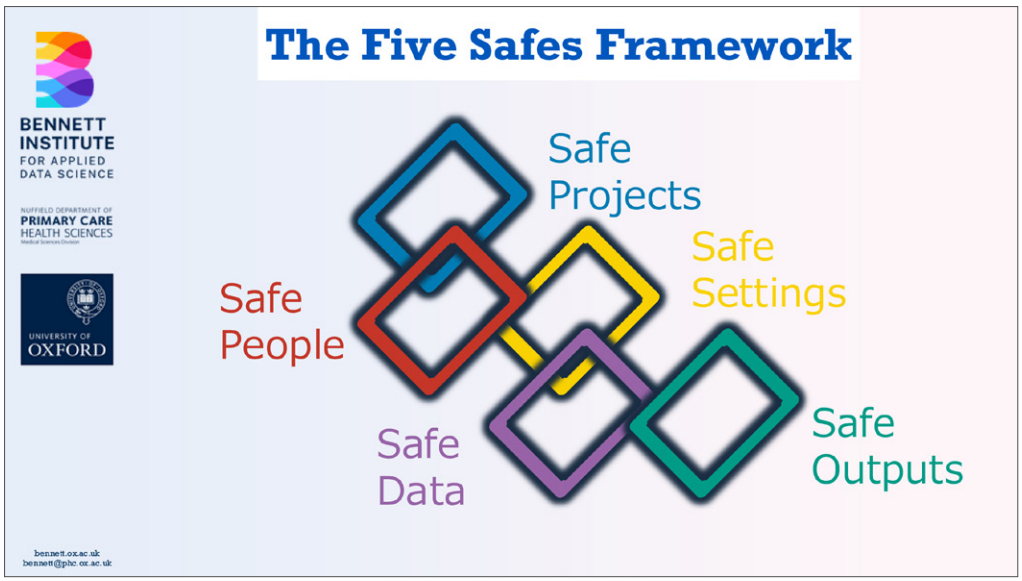


Figure 5. One of the overviews of the Five Safes framework, shared in Workshop 4

In Workshop 3, participants heard from a Chief Data Officer (CDO) from a large UK retailer. He shared an example of how customer data could be used for public benefit through smart data research. The CDO described how his organisation wanted to get involved in smart data research because it had the potential to deliver good outcomes for the nation, which is part of their organisational mission.

Although some participants remained sceptical about the company's motivations, others saw the argument that a company could be purpose-driven – interested in contributing to society, as well as commercial.

Additionally, participants recognised that involvement in research for the public good would generate good publicity for private companies (despite this retailer not publicising their involvement). For those participants, this felt like an acceptable commercial benefit for the company (likened to 'corporate social responsibility' initiatives). They felt that this might drive public support for the organisation and lead to longer-term commercial gains.

> **"The private companies would also get recognition for it. So, if it has made a big difference and had this positive impact, it would be good advertising for the company."**
>
> Participant, Inverness, Workshop 3

Further discussions with the SDR UK team about how the relationship between private companies and data services works in practice helped to reassure participants. They were particularly interested in how a balance between public benefit and private gain plays out in smart data research.

Participants were reassured to hear that the public good would be a requirement if a researcher from a company wanted to access a secure dataset. This helped to build trust that smart data research would achieve a balance of private and public interests. Specific processes offered reassurance, such as:

- Approval would not be given for research funded by SDR UK if there was only a commercial benefit to the company, rather than a public benefit
- Understanding that researchers would need to satisfy the SDR UK approvals processes that the research was in the public interest, rather than the company's interests.

> **"I guess it would be making sure that the goals and outcomes were all assessed and thought out ahead of time. There would need to be sort of consensus formed about what would constitute significant enough public benefit, to justify that kind of research."**
>
> Participant, Inverness, Workshop 2

By Workshop 5, the views of many participants had shifted. A greater understanding and support of smart data research, and reassurance around the role of private companies resulted in some participants coming to believe that there may be opportunities to be realised around private company involvement in smart data research. Some participants felt that the repurposing of smart data collected by private companies, to be used for a public benefit, was positive. They viewed this as the public getting a benefit back from the private company, that they otherwise wouldn't receive.

> **"It's good at the end of the day, company gets the profit and public's gaining from it."**
>
> Participant, Gateshead, Workshop 5

A few participants even considered that established private companies could have an important role in promoting the opportunities and impact of smart data research on the public good.

> **"We will promote your company, based your use of data to support public good to improve their PR."**
>
> Participant, London, Workshop 5

A minority of participants with the least trust in private companies did not change their mind. They felt that commercial gain and public benefit could not co-exist in smart data research, and were hesitant about any commercial gain at all.

> **People at the top often have vested interest in this company and that company and shares here and there... Even if something is for the greater good – there are still cracks there.**
>
> Participant, London, Workshop 5

Unresolved questions

Many participants came to accept that private profit and national interest can go hand-in-hand in principle. There was not enough time in this dialogue to discuss and deliberate the types of commercial benefits that might be acceptable to people. For example, when discussions touched on the idea that private companies could be compensated for the time spent sharing data, some felt this was reasonable while others did not. They also recognised that it might be difficult to identify a specific public benefit at the time of granting access to smart data for research, and that this may only be understood once the research has been conducted.

Participants want transparency from private companies

Participants thought it was essential to know which companies are involved in smart data research and how.

Many felt that the presentation from the UK retailer, in Workshop 3 was an exemplar for private companies in being transparent about their involvement. They felt positive about what they perceived to be the open, transparent and passionate delivery of the speaker's presentation. This challenged their previous perception of private companies concealing their true motivations from the public, in long and incomprehensible terms and conditions.

> **"[The speaker] seemed to be proud of how he is using [company] data. You could see he valued trust, and I trusted him with the data. It seemed very trustworthy."**
>
> Participant, Gateshead, Workshop 3

For a few participants, the retailer's decision not to publicise its involvement in smart data research to the public at large was not transparent enough. If a company was not gaining a benefit from the positive publicity of being involved in smart data research, where else they might be benefiting? These participants cited concern about a 'hidden agenda'.

> **"[The speaker] seemed nice, passionate and informative. But he is one person. I feel cynical about it all."**
>
> Participant, Newport, Workshop 3

Oversight to ensure private involvement in smart data research aligns with public good

Participants agreed that for the broader public to have faith in this alignment between commercial and national interest there needs to be some form of oversight.

They called for an independent body to provide oversight and governance of the relationship that data services and researchers have with private companies. Participants said this would provide the reassurance they needed to ensure that where private companies are benefiting, this does not outweigh the public good.

Participants suggested that an independent body could:

- **Audit research projects:** Although reassured about the need for research to demonstrate public good for it to be funded, many participants wanted this to go a step further, with

checks on projects to ensure public benefit is maintained and prioritised as an objective of the research. Where this isn't the case, participants deemed it acceptable for funding to be removed.

- **Take action against misuse:** Many participants thought that punishments such as fines or restricting private company involvement should be implemented if a private company fails to adhere to rules (such as the Five Safes).
- **Recognise good actors:** Participants also wanted incentives or accreditation for private companies that maintain good practice. Some participants suggest certification or the use of a watermark to recognise private companies' contribution to public benefit through smart data research.

> **"It could be not just fines, but penalties. Like being restricted from certain kinds of research, blacklists or having to undertake more training."**
>
> Participant, Newport, Workshop 5

Response from SDR UK

Participants weren't always clear how companies are already using smart data for their own purposes. They were initially sceptical about the motivations of private companies for sharing data with SDR UK. But they became more comfortable when we explained the change SDR UK is seeking to bring about (safe access to de-identified data for research) and the safeguards in place.

Participants wanted transparency about how companies are involved in smart data research. We will put transparency at the forefront of smart data research. This will include regular public updates on research outcomes and clear communication about partnerships.

One concern raised by participants was the perceived lack of transparency in the consent process. Once data has been de-identified, it becomes difficult or even impossible for researchers to trace it back to specific individuals. This makes it impractical for people to withdraw their consent in certain research scenarios. This is acknowledged by UK data protection law, which provides a number of lawful bases for processing data that do not rely on explicit consent. For SDR UK, this reinforces the importance of meaningful public engagement, to ensure that research is being done responsibly to retain broad public support.

We appreciate participants' desire for independent oversight to ensure private sector involvement aligns with public good. It's important to acknowledge that comprehensive governance structures already exist. Rather than creating a new independent body, which would duplicate existing oversight, we believe enhancing transparency around current governance processes would more effectively address public concerns.

6. Security, oversight and accountability

Security of smart data was a top concern for participants from the start of the dialogue

Participants were worried about how their data would be shared and handled, particularly by private companies. They wanted to know whether their personal information might end up in the 'wrong hands' at any point in the smart data research process. This concern persisted despite participants being informed about the de-identification of smart data, reflecting strongly held broader security concerns. They worried this might happen through mismanagement, system or human error, or malicious intent (e.g. hacking).

In this context, participants were worried that:

- Smart data researchers might be able to access their sensitive (and in some cases embarrassing) personal data and identify them with it.
- Smart data might be sold and used in a way to profit from the public.
- Bad actors might gain access to their smart data and use this to steal their identity or money.

These concerns remained prominent throughout the first three workshops, with participants continuously returning to them in their discussions.

> **"I knew about data you use if you never have Wi-Fi or something, you use data on your phone. That's how I classed data. I never thought about smart data before."**
>
> Participant, Inverness, Workshop 1

> **"I'm worried about my bank details, if I put that into something, I'm always checking if it's legit before I do it. It's scary to be hacked."**
>
> Participant, Inverness, Workshop 1

Participants maintained that the safety of data was a key risk that required attention

Participants maintained that data security was a key risk associated with smart data research over the course of the dialogue. This was despite being informed about the de-identification of smart data and being reassured by specialists and representatives from SDR UK that security breaches within smart data research are rare.

In Workshop 4, participants heard from a specialist in applied data science at the University of Oxford. He presented information about how data services work, and the robust safeguarding processes in place to manage the sharing and handling of smart data for research (see Figure 6 below). Participants found his presentation surprising but while most felt reassured by hearing of the strong security standards, a (vocal) minority of participants remained doubtful.

> **"You're out there, we're doing what we've got to do, we're all on the internet. But is our data safe? And how safe is it? I know that [Specialist] was saying, 'Yes, it is safe'... But how safe?"**
>
> Participant, Inverness, Workshop 4

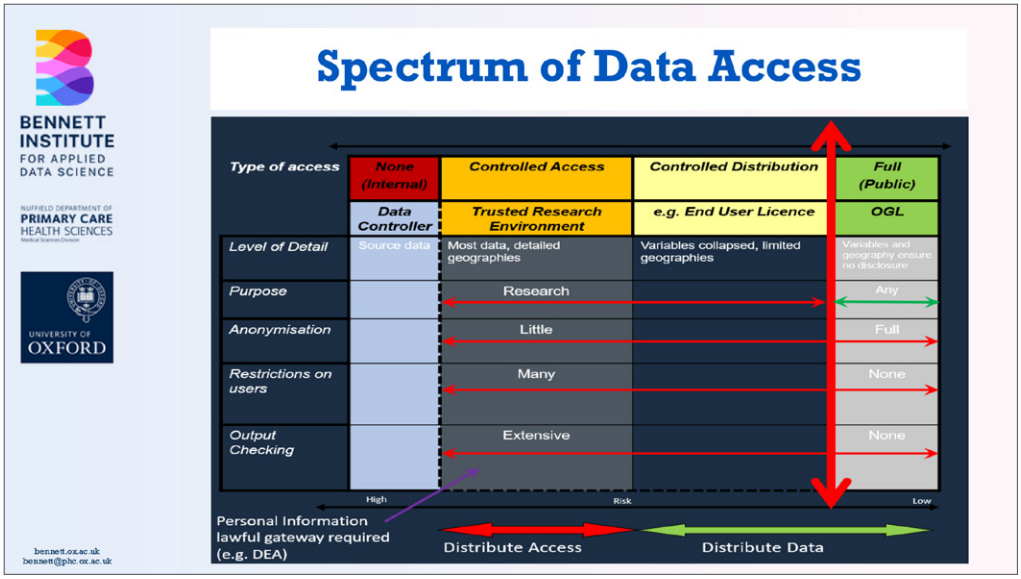


Figure 6. Part of an overview of data access shared by specialist in Workshop 4

Learning about safeguarding processes strongly reduced levels of concern about data security in practice

Changes in participant perspectives: Concerns about data security did diminish over the course of the dialogue, after hearing about the safeguarding processes in place. This was the area where participant sentiment shifted most during the dialogue: they moved from a place of high concern to feeling much more reassured. Once they had this reassurance, participants became more accepting of smart data research, including towards the elements where they had most scepticism such as private sector collaboration. This suggests that offering participants clarity on security processes plays a key role in facilitating meaningful discussion about smart data research. This is explored in further detail in Chapter 7.

> **"I think I would have to know a bit more about how they are keeping the data secure and what sort of processes they use before I can make my overall decision on it."**
>
> Participant, Newport, Workshop 4

By the end of the dialogue, several participants acknowledged that creating too many restrictions on how data can be shared or used might deter private companies from contributing to the research – leading to 'missed opportunities' for relevant data to be used.

More broadly, participants thought that certain types of data might require stricter controls than others. For example, more sensitive information such as personal, financial and health data.

> **"There needs to be exceptions to the rule on not creating too many restrictions."**
>
> Participant, London, Workshop 4

> **"Maybe a rule about sort of sensitive data, financial and health needs to have. I don't know if it's VPN or whatever, but much, much more strict protocols in order to access it."**
>
> Participant, Newport, Workshop 4

> **"I think health records, financial information, personal data, names, addresses, phone numbers, all that kind of stuff needs more protection than public data. I think public data doesn't require as much security as all that sensitive data."**
>
> Participant, Inverness, Workshop 4

Changes in participant perspectives: Another significant change was brought about when the Five Safes framework (Figure 5 in the previous chapter) was introduced. Participants found this helpful in two ways:

- **Informative:** Seeing the safeguarding measures mapped out across each step of the research process helped to explain how smart data research works, who is involved and how researchers will use data. This helped participants to understand the smart data research process in more

detail, which built their confidence. For example, learning that it is a vetted and controlled process, rather than individual researchers operating wholly independently.

- **Reassuring:** The Five Safes indicated that the overall process was structured with security in mind. This built their confidence that security risks were being mitigated throughout the smart data research process.

> **"Something I hadn't considered is that the researcher is trained before they have access to the data. This surprised me – in a good way."**
>
> Participant, Inverness, Workshop 4

> **"The framework is good. A lot of thought has gone into them. It shows they are thinking about the safety of our data."**
>
> Participant, Inverness, Workshop 4

Participants were particularly reassured by two aspects of the data security framework: Safe People and Safe Projects. These elements, both independently and in combination, addressed key concerns about security and public benefit prioritisation.

Safe People

This framework focuses on minimising data security risks, which participants identified as primarily stemming from human error. Their concerns centred on researchers potentially mishandling data, either through improper security practices or deviation from intended public benefit purposes. Participants showed particular interest in researcher qualifications, suggesting the development of an accreditation system to certify smart data researchers.

Safe Projects

This aspect ensures that public benefit remains a core priority from the inception of any smart data research project. This principle became especially significant when discussing private sector collaborations, where participants emphasised the importance of maintaining an appropriate balance between public good and commercial interests.

> **"You need to trust the people over everything. If the people are safe and everything has been input as best as it can then everything should run thoroughly."**
>
> Participant, London, Workshop 4

> **"Safe projects. I think there are projects with commercial aspects, and projects that are for the public good. It is important that the commercial aspects are restricted. I think safe projects are the most important, but it is tied with people."**
>
> Participant, Belfast, Workshop 4

Re-identification was not a key area of concern

Participants were less concerned about the risk of re-identification. This topic was introduced in Workshop 4 through a hypothetical example where combining multiple smart datasets might enable re-identification.

Participants did not feel that re-identification was likely to occur; they thought that it would be time-consuming for someone to seek to re-identify an individual, and that it was therefore unlikely to happen. However, several participants did note that re-identification could present a risk for vulnerable people, if not for them personally. For example, people who have experienced domestic abuse, and victims of crime.

On the whole, participants did not feel that further specific safeguards were required, beyond those within the Five Safes, to address re-identification. However, it's important to note that this dialogue could not address all the potential data types that could be used for smart data, or how they might be combined, nor did it involve people with lived experience of these types of harms: specific cases may elicit higher levels of concern, either for the general public or for differentially affected groups of people.

> **"This might be relevant for someone who has experiences like domestic abuse, so if someone found out where they live through their Uber data, they would be reluctant to share their data."**
>
> Participant, Newport, Workshop 4

Participants wanted to know how oversight and accountability would ensure safeguarding processes are followed in practice

Participants felt more could be done to clarify how oversight and accountability would support safeguarding processes. They had specific questions about the following areas:

Oversight: Who would be ensuring that all processes are followed as they should?

- Participants thought it was important that any oversight would be independent, to allow them to oversee processes from an unbiased perspective (e.g. compared with actors involved in the research project who would have a vested interest in how data could be used).
- This was particularly important when participants discussed the fact that the public would not consent – or feel that they had consented – to their smart data being used for research projects. In this context, participants thought that oversight would be important to 'represent' and safeguard the public's interests.
- Participants also saw a role for an independent 'overseer' in managing researcher accreditations and giving a recognised 'stamp of approval' for data services.

Accountability: How would those overseeing these processes, as well as those involved in carrying them out, be held accountable? How would processes be enforced?

- Participants thought that accountability would play an important role in ensuring that processes are followed in practice by researchers, data services and private companies.
- They thought that accountability was also important for promoting transparency around how and why data is collected, shared and used.
- They were strongest in their views that private companies should be held accountable if something were to go wrong, reflecting the distrust that many participants had in the private sector throughout the dialogue. This was particularly true for the most cynical participants.

> **"Who makes sure everything is done properly? You know? This should be overseen by someone who is accountable, otherwise there will be blaming."**
>
> Participant, Newport, Workshop 4

"I know they're saying there is a safeguard there, but how strong is that safeguard? As I've said, in the four meetings we've had, I've been hacked twice, and I've had to change all my details. I'm not really sure on all this at the moment. I need more reassurance that the safeguard that is there is going to actually help me."

Participant, Inverness, Workshop 4

"I feel a bit more confident in the framework for looking after data and keeping it secure. Confidence from what he said about the number of layers of protection. The only thing I didn't know or perhaps could be asked is he's always on about these researchers having accreditation. Who does the accreditation?"

Participant, Newport, Workshop 4

Response from SDR UK

It is useful to hear how reassuring participants found the Five Safes framework – the safe access to sensitive data (e.g. personal data, de-identified data, or sensitive anonymised data), for research.

As in the previous chapter, we believe enhancing transparency around current processes would help address public concerns. For example:

- Explaining how we support and implement the Five Safes, describing the legal basis of what we do, the safeguards and the oversight in place, and our overall approach to public engagement.

Our governance structure delineates responsibilities between SDR UK and regulatory bodies, establishing formal channels for coordination and audit with external oversight agencies.

Internal compliance monitoring systems are in place, supported by regular reviews and updates. Importantly, our structure includes mechanisms for incorporating public feedback into our processes, ensuring the system remains responsive to public needs and concerns.

This systematic approach creates a robust foundation for secure and ethical smart data research that serves the public good while maintaining the highest standards of data protection.

7. Public involvement

Public understanding and trust

Participants found it difficult to envisage how smart data research would work in practice, and how it would be used to inform decision-making for the public good.

> **"I never thought about smart data before. I knew [companies] was using data for different things, like browsing, but I didn't really click the smart data idea."**
>
> Participant, Inverness, Workshop 1

At the start of the dialogue, participants were most concerned with informing the public that smart data research is happening and where their data was going. They highlighted three areas where they wanted more information about smart data research:

- **How data is being used:** needing transparency about how the public's data was being collected and used by private companies and public institutions.
- **How consent is taken and what it means:** needing accessible and transparent information about what consent means, and what people are agreeing to when they interact with a digital product or service.
- **Risks of data sharing:** needing to warn people about the potential harms if they unknowingly shared their data (e.g. data security and privacy).

> **"If someone is really ill-informed, like me, that's scary. If I'm clicking something and I'm actually doing something bad that's scary – what am I consenting to? Could be allowing them access to all parts of my life."**
>
> Participant, London, Workshop 1

> **"Needs to be transparent, who has it, how is it being used, what third party it is going to, it's scary if you don't know who has it."**
>
> Participant, Inverness, Workshop 1

At this stage of the dialogue, most participants didn't identify a particular actor within the system as being responsible for informing the public on these topics. However, several participants did say that private companies have a responsibility to raise awareness and increase transparency when collecting and sharing smart data.

“Educate people on what their information is being used for and how it can benefit them.”

Participant, Newport, Workshop 1

“I hope that [private companies] have more consent or understanding when going on websites, more of an explanation of where your data is going and why it’s used.”

Participant, Newport, Workshop 1

Changes in participant perspectives: As participants learnt more about smart data research many became more comfortable and accepting of their smart data being used for research purposes. Trust improved as they became more informed about what smart data research looks like in practice, how it could benefit the public good, and what safeguards would be in place to ensure it is conducted securely and ethically. When reflecting on their own journey of understanding, participants used words such as ‘privileged’, ‘excited’ and ‘reassured’ to describe their experience and change in attitude to smart data research.

“For me it’s all about education, in Workshop 1 I didn’t know about anything really. But now that I know, I feel much more positive that some of my previous concerns are no longer a concern anymore.”

Participant, Inverness, Workshop 5

“We just need to be informed, if SDR UK could tell us that this is what they do and how they work and go between the government.”

Participant, Gateshead, Workshop 5

At the end of the dialogue, several participants shared their hopes that the wider public would become more informed about smart data research. They felt that increased awareness about how data is used more broadly would make the public more accepting of their smart data being used to benefit the public good—for example, going beyond the areas of concern they raised in Workshop 1 (e.g. data used for marketing or scams) to raise awareness of broader uses of data for decision-making.

“We have been giving out data, not realising what the data is being used for, or could be used for. We need more education out there so people have a fuller understanding of how they are giving up their data, like on loyalty cards and things. We have been involved but not knowing what we are involved with.”

Participant, Inverness, Workshop 5

“A month ago, none of us had even heard of this. We are the privileged ones to know about this.”

Participant, Newport, Workshop 5

By Workshop 5, as participants' support for smart data research had built, they had some more specific guidance on what the public should know about the smart data research process. They felt this should include:

- Information about what smart data research is, illustrated with real-world examples of how it is used
- Details about how data is kept secure and the processes around this
- Examples of the real-world impact of smart data research on public good.

Unresolved questions

Although participants wanted the impact of smart data research to be communicated to the public, discussions did not dig deep into where responsibility for communicating this lay. While some thought researchers should do this, others thought that the responsibility would sit with private companies.

Participants wanted information about smart data research to be transparent, easy to understand and access

Participants had some conditions for information that is to be shared with the public about smart data research. They wanted information to be:

- **Transparent about the impact of smart data research, even where there is not one:** Participants recognised that there might be times when smart data research does not lead to real-world benefits for the public. In these cases, participants thought it was still important to communicate about it.

> **"Whether the results are good or bad, publish the results – even if this wasn't the outcome you [SDR UK] anticipated. Transparency and accountability of the outcome."**
>
> Participant, London, Workshop 5

> **"To be more transparent so we can trust them more."**
>
> Participant, Newport, Workshop 5

- **Easy for the public to access, in a way that best suits their needs and abilities:** Participants were concerned that currently research is only accessible via academic papers, which are difficult for the general public to understand, and often hidden behind paywalls. They also wanted information to be shared through a range of methods to maximise public engagement and ease of access.
- **Easy for the public to understand:** ensuring that information is easy to interpret and understand, including:
  - Using simple language, without jargon, to explain data and smart data research
  - Using real-life examples and case studies to highlight the impact that smart data research is having
  - Use videos, animations or infographics to share complex information (particularly about the smart data research process).

> **"It would be good if the government had a website with categories to show which data is being used for what, and to see what stages they go through, for the public to see what's being discovered step by step."**
>
> Participant, Inverness, Workshop 1

A few participants also felt that information should be kept up to date and relevant—for example, ensuring that communications about smart data research include findings from research conducted in the last year. Participants thought it would be useful for the public to be regularly updated about progress in this space, with ongoing communication to keep the public abreast of any advancements.

> **"I suppose having some transparency, making sure the sources and the methods and the findings are all openly accessible to the public, just to keep the trust going and accountability."**
>
> Participant, Inverness, Workshop 2

Unresolved questions

Participants remained divided on how much information they truly wanted to be aware of. Some participants felt they needed to understand only the 'basics' of smart data research to be reassured about its benefits. Those participants would be satisfied with understanding what smart data research is and the processes around it to ensure public good and data security (e.g. the Five Safes framework). Other participants wanted more detailed information, such as access to research reports and understanding exactly who is involved in what research (and the motivations behind it).

Participants want the public to be engaged in SDR UK's programme

Participants did not only want the public to be informed about smart data research but also engaged in decision-making about how it works in the UK. In Workshop 5, participants were introduced to a framework for public involvement in decision-making (see Figure 7 below), and asked to consider how it might be applied to SDR UK future public engagement.

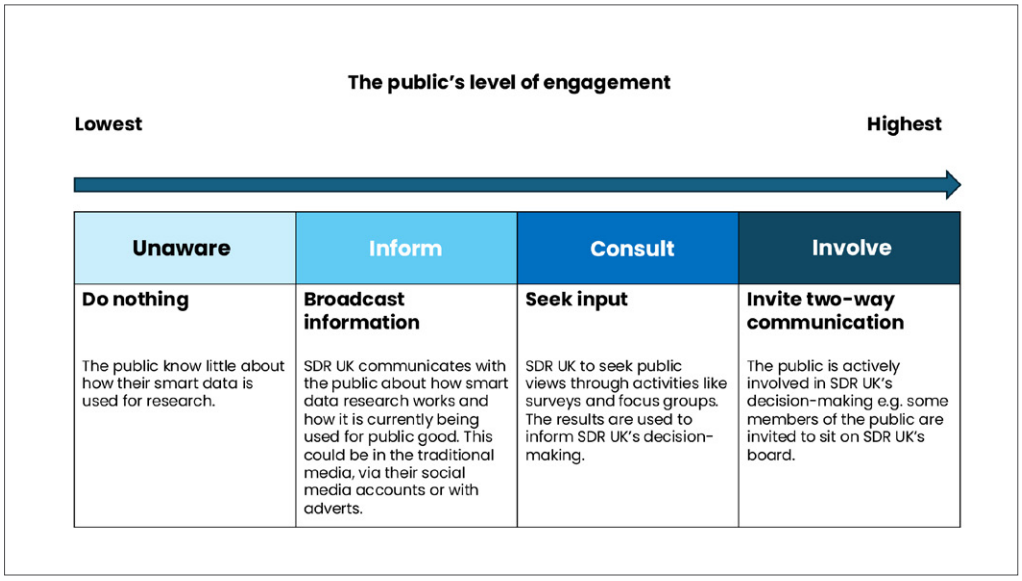


Figure 7. Framework for public involvement in decision-making.

The theme of personal control – or lack of it – was raised throughout the dialogue. This diagram prompted participants to think about how engagement could give the public a greater sense of control over their data. The potential of public engagement as a way of putting control back in their hands was an appealing prospect. Many pointed to a fundamental belief that as the public's data was being used for smart data research, they should have a say in how.

> **"We should be able to know what data is being collected on us."**
>
> Participant, Belfast, Workshop 2

> **"I understand your point that data should be used to benefit society. My biggest issue is that I have no control over my data—I can't opt out. If I wanted to use Uber, I am automatically opting in to them using my data and I don't even know what they are doing with this data."**
>
> Participant, London, Workshop 4

> **"It would be good if the public were able to give their suggestions on what they want to be researched for the public good. Medical research should be done, but there should also be public driven research. The public should definitely have a say."**
>
> Participant, Newport, Workshop 5

Changes in participant perspectives: Participants maintained that the public should be involved in SDR UK's decision-making.

Reflecting on their experience participating in the dialogue, participants commented on the amount of new information they had to engage with to answer the questions at hand. Some said they still felt far from experts on smart data research and worried that it wouldn't be realistic to give members of the public enough information and expertise to be able to discuss every element of smart data research. They also suspected there were parts of the programme that would be highly technical and confusing for the average person to comment on.

Participants felt that there were decisions in which it would not be necessary to involve the public, most likely because they were too technical or where they could not have a useful impact. By contrast, they thought that the public should be engaged to understand their priorities when defining what constitutes the 'public good' – an area which they felt was much more feasible for members of the public to input.

> **"I don't think the public need to know the ins and outs of smart data to inform it, as they have experts."**
>
> Participant, Newport, Workshop 5

> **"I think its self-explanatory, the Five Safes are put there for a reason, no one should be concerned really."**
>
> Participant, Gateshead, Workshop 5

Information about security safeguards had significantly increased participants' confidence in SDR UK's work. The Five Safes framework, in particular, helped address key security concerns, though participants emphasised that these safeguards must be supported by proper oversight and accountability measures.

When discussing ongoing public involvement, participants suggested that widespread communication about these security measures could reduce the need for extensive public consultation. They indicated that if safeguards were properly implemented and monitored through robust oversight, additional public input on security matters would be less crucial for SDR UK's decision-making process.

Some participants were drawn to the idea of consulting the public as a matter of principle – that publicly funded institutions should involve members of the public within their organisational processes and structures for decision-making.
However, most participants were less interested in consulting the public as a matter of seeking input. They felt that this would be relatively superficial, and had questions about how it would work in practice and the value it would add.

By contrast, most participants thought that public input would be more meaningful and impactful if sought through a two-way involvement. They also hoped that having members of the public on SDR UK's board would act as an important check, holding them accountable for delivering public good.

“The public should be informed, and they should give their input. I like the ideas of surveys and focus groups but I’m not sure if we are educated enough to make the decisions.”

Participant, Newport, Workshop 5

Participants were most enthusiastic about being involved in decision-making where they felt they could have the most impact:

- **To offer a unique perspective:** Some participants mentioned involving the public in decisions as a way to combat data bias could be particularly powerful. Hearing from those living in vulnerable circumstances or who were traditionally under-represented in datasets was seen as a method of helping ensure smart data research works for everyone. Some participants, particularly those in Wales, Scotland and Northern Ireland, spoke of the importance of involving the public on local issues where they would be able to offer a perspective that researchers or SDR UK might be lacking.
- **Public good:** Participants had a lot to say about defining the public good, as they could form a view based on their own experiences, values and beliefs. Participants emphasised how the public does not need to be educated on the technicalities of smart data research to support discussions about public good but can bring their own lived experiences to help make judgements. They also had a lot to say about data security as this was an area of high perceived personal risk and concern, suggesting an interest in contributing to further strategic (but not technical) conversations about the topic.

“I know SDR UK is in its infancy, so having things like what we are doing today, going out to the public and getting different perspectives is important to do often with different groups and different areas.”

Participant, Inverness, Workshop 5

“I think that last question about having the public involved in what topics get researched is a great point, with the academics you can have things that get missed and getting the public involved in topics would be great.”

Participant, Belfast, Workshop 5

Response from SDR UK

Participants expect SDR UK to play a role in raising awareness of what smart data research is, and how data is used for research and decision-making that impacts people's lives. We will consider how to talk about smart data research, and associated social or ethical issues, in an engaging and informative way.

We will also consider how to inform the public most effectively. This includes using a variety of formats and channels, while ensuring that concepts are explained clearly and without jargon.

Communications will also play an important role in providing transparency about how publicly funded research makes a difference.

Participants wanted SDR UK to share the impact of smart data research, even where there may have been none. Alongside SDR UK's own role in sharing this information with the public, SDR UK could also produce or share guidance for researchers to help them better communicate with the public about impact.

Participants wanted their data to be used to make decisions about areas that will deliver impact for the public good. They felt that involving them in decisions about smart data research would be a way of ensuring that real-life impact for the public is considered and provide a better sense of control over how their data was being used. We will set out how we will involve the public in decisions where they can have the most impact, either because they can offer a unique perspective or because it is a topic that the public is interested in and passionate about.

This dialogue has demonstrated the value of involving the public in a broad way, covering key issues in context and alongside wider concerns. This will help to surface what matters most to the public, rather than risk constraining discussions to what researchers or SDR UK might expect to be important to them.

sciencewise

Smart Data Research UK

Thinks Insight & Strategy

Stay in touch

Email us
info@sdruk.ukri.org

Find us
sdruk.ukri.org

Bluesky
sdruk.bsky.social

LinkedIn
Smart Data Research UK

250325