



# A UK-wide public dialogue exploring what the public perceive as ‘public good’ use of data for research and statistics

October 2022

# Acknowledgements

---

## Authors and project team

<b>Dr Fran Harkness</b>	Co-Director, Kohlrabi Consulting
<b>Cornelis Rijnveld</b>	Researcher, Kohlrabi Consulting
<b>Yuncong Liu</b>	Researcher, Kohlrabi Consulting
<b>Shayda Kashef</b>	Public Engagement Manager, ADR UK
<b>Dr Mary Cowan</b>	Research Specialist, Office for Statistics Regulation

## Project advisory group members

<b>Dr Penny Babb</b>	Head of Policy and Standards, Office for Statistics Regulation
<b>Dr Jo Duffy</b>	Senior Research Policy Manager, Economic and Social Research Council
<b>Dr Emma Gordon</b>	Director of the ADR UK Strategic Hub and Director of ADR England
<b>Dr Suzanne Hall</b>	Director of Engagement, The Policy Institute, King's College London
<b>Aidan Peppin</b>	Senior Researcher, Ada Lovelace Institute
<b>Dr Paul Richards</b>	Open Research Lead, UK Research and Innovation
<b>Rebecca Veitch</b>	Head of Research Integrity, UK Research and Innovation
<b>Elizabeth Waind</b>	Senior Communications and Engagement Manager, DARE UK (Data and Analytics Research Environments UK)
<b>Dr Simon Whitworth</b>	Head of Data Ethics, UK Statistics Authority

## ADR public engagement steering group members

<b>Elizabeth Nelson Gorman</b>	Public Engagement, Communications and Impact Manager, Administrative Data Research Centre Northern Ireland
<b>Harriet Baird</b>	Impact and Knowledge Exchange Manager, Scottish Centre for Administrative Data Research
<b>Professor Kerina Jones</b>	Associate Director of Information Governance and Public Engagement, Swansea University

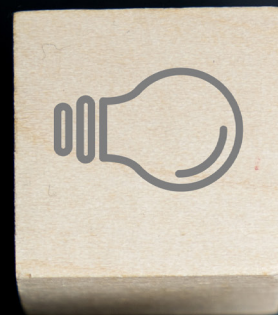
The team thank the members of the public for giving their time and sharing their invaluable views with the research team, and all the many people across the country who helped make the workshops a reality. We would also like to thank the Project Advisory Group and the ADR UK Public Engagement Steering Group for their helpful guidance and advice on the project throughout the process.

We would like to thank Dr Anna Price (OSR), the facilitators, and colleagues in Scotland, Wales, Northern Ireland, and England for their contribution to the workshops. We would like to acknowledge Helen Kennedy, Professor of Digital Society, University of Sheffield, for providing feedback on an early version of this report.



# Contents

Acknowledgements.....	2
Foreword.....	5
Executive summary.....	7
Introduction.....	11
Method.....	13
Analysis.....	16
Findings (expanded).....	19
Summary.....	37
References.....	39
Appendix A: Glossary of terms.....	41
Appendix B: Participant demographic information.....	42
Appendix C: Workshop agenda.....	43



# Foreword

---

## Dr Emma Gordon, ADR UK

At ADR UK, maximising the power of administrative data for the public good underpins everything we do. In practice this means developing and strengthening partnerships with public, government and academic stakeholders as we realise our mission of transforming the way researchers safely access the wealth of administrative data for research that aims to improve people's lives. By working in this way, ADR UK is bridging the gap between government and academia, enabling policy to be informed by the best evidence available.



The ADR UK programme began as a three-year pilot to test the feasibility of our mission. In 2020, in the final year of our pilot phase, we published a [review](#) of public attitudes towards the sharing and linking of administrative data for research. It found that the public is broadly supportive of this use of administrative data if three core conditions are met: that the research serves the public interest; data is protected from identification and re-identification; and that there is trust and transparency on behalf of those handling the data. This work provided a steer to develop the [ADR UK Public Engagement Strategy](#) and informed the vision for our current investment period (2021-2026).

Engaging directly with members of the public informs every stage of our work. Listening and responding to public views is essential to demonstrating trustworthiness and maximising the public benefit of research using administrative data. This public dialogue is the first project of its kind for ADR UK and therefore serves as a milestone for the programme. The insights generated from this work will be carefully considered as we seek to inform practices across the ADR UK programme. They also provide a starting point for deeper exploration of issues related to the use of public sector data for the public good. We hope that this work can serve as an important resource for others working with data and statistics.



## Ed Humpherson, Office for Statistics Regulation

In OSR we have a vision that statistics will serve the public good, and we believe that strong meaningful engagement with the public is key to ensuring that we can bring this vision to life.

This report presents fascinating and novel insights into an important subject and makes a strong contribution to our understanding of how the public perceive the public good of data for research and statistics.

The findings illustrate an important point: there is no one 'single' public. Participants themselves grappled with the concept of defining 'the public', showing it is not perhaps as straightforward as it might seem. The findings also illustrate that 'the public' are not just one homogenous group, as we can see that participants did not always agree with each other. Whilst this may show that we were successful in recruiting a diverse sample of participants and creating an environment where they felt comfortable respectfully disagreeing with each other, it also shows the importance of inclusive discussion. Over the course of the workshops, through listening to each other and considering other viewpoints, participants developed increasingly richer and more nuanced insights about the public good which has helped us to develop a valuable resource.

Another key finding to highlight was that participants saw clear distinctions between the phrases public good, public interest, and public benefit; phrases which are often used interchangeably by those who write about data for research and statistics. This is a fascinating insight which emphasises the importance of choosing language thoughtfully and carefully for public-facing communications.

This work is an exciting and important step for us. It helps us understand what more could be done to serve the public good from the perspective of the public. We will consider this evidence carefully as we evaluate the role that statistics play in the lives of the public, and consider how to ensure that statistics are serving the public good.



# Executive summary

---

Recent years have seen ever increasing possibilities to generate insights from administrative data collected from the public by public services, bringing an even greater need to include the public in conversations about its future use (DARE UK 2022; Waind, 2020).

The potential for research and statistics to be used to inform decision-making and improve the understanding of our society has been envisioned as serving the public good, by both academics and professionals working in the voluntary and public sectors (HDR UK 2021a; Involve, 2018). Public good (also sometimes termed 'public interest' or 'public benefit') is considered to be central to decision-making as public sector data can only be used to produce statistics if its purpose is considered to serve public good in accordance with the Digital Economy Act (2017).

Yet, so far, although the public regularly cite the public good use of data for research and statistics as a core condition of their support for data sharing and use, there is no consensus on how public good is understood by the public (Cowan & Humpherson, 2020; DARE UK 2022; HDR UK, 2021b; Waind 2020).

To address this, [ADR UK \(Administrative Data Research UK\)](#) and the [Office for Statistics Regulation \(OSR\)](#), supported by independent researchers Kohlrabi Consulting, undertook a deliberative dialogue exploring public perceptions of 'public good' use of data for research and statistics.

In June 2022, 68 members of the public took part in a series of UK-wide deliberative workshops, with analysis of 24 hours of conversation. Whilst a consensus was not always reached by participants, the evidence has been synthesised into five broad themes. Many of these findings feed into one another; together they provide insight on points to consider when using data and statistics in a way that is consistent with serving the public good.



## Findings

The following findings are derived from the feedback of participants in this research and aim to improve the understanding of what the 'public good' is from the perspective of members of the public. ADR UK and OSR will consider these findings, alongside other evidence, to inform how their work can maximise the benefits associated with data for research and statistics.

### 1

#### **Public Involvement: Members of the public want to be involved in making decisions about whether public good is being served**

Participants expressed a preference for meaningful public engagement to help inform decision-making concerning the use of data for research and statistics and explored several forms of how the public could be involved in decision-making. Participants articulated that inclusive public panels, with diverse members of the public, should play a central role in decisions made about data and statistics. They suggested that this role be supplemented with public conversations around the wider use of data for research and statistics, with requirement for continuous efforts to engage with the public.

### 2

#### **Real-World Needs: Research and statistics should aim to address real-world needs, including those that may impact future generations and those that only impact a small number of people**

Participants suggested that the value of data being used for research or statistics should be assessed by need, rather than by the number of people who would benefit, suggesting that serving the public good does not refer to serving the needs of a specific number of people. Participants felt that addressing social inequity and social inequality was a particularly pertinent reason to enable access to data for public good.

Participants asked that the public have full and transparent access to the decision-making process of Data Access Committees (see [Appendix A Glossary of terms](#)) to understand how public good was intended to be served. Although this information may already be publicly available, participants felt it could be more easily accessible. Participants articulated that they would like to see transparency from Data Access Committees regarding the impacts of proposed projects, including on how projects aim to address issues related to equality and inequity.

## 3

**Clear Communication: To serve the public good, there should be proactive, clear, and accessible public-facing communication about the use of data and statistics (to better communicate how evidence informs decision-making)**

Participants felt strongly that the public would benefit from greater awareness of the practices, motivations and outcomes of public good use of data for research and statistics. While information does exist on websites, and is shared across social media channels, participants felt these messages often did not penetrate their personal networks. Proactive communication that is clear and accessible – both regarding the use of language and availability of information, with the aim of reaching broader audiences, was viewed as a solution.

An example suggested was a national campaign to raise awareness about the public good use of administrative data for research and statistics. Public awareness of data use and associated practices was perceived as a way to support further democratic accountability for those who are responsible for ensuring data is used in a secure way.

## 4

**Minimise Harm: Public good means data collected for research and statistics should minimise harm**

Many participants felt a personal responsibility that data about them should not contribute to anything harmful; for example, data should not be used to perpetuate stereotypes about certain groups of people. To mitigate potential harms, participants suggested consulting members of the public, particularly those with lived experience, about potential uses of data for research or the interpretation of statistical patterns. Engagement with those who have relevant lived experience may particularly inform appropriate interpretations of statistics, including language.

The Five Safes framework was explained as an example of data security measures, for which participants showed support. As well as wanting more widespread public awareness of the security around data access, participants desired increased accountability from those working with data and statistics, though with whom precisely was not explored. Suggestions given were having a named 'data protection lead', a whistle-blowing procedure if misuse of data was identified, a public telephone line, and public awareness of the repercussions of data misuse.



## 5

**Best Practice Safeguarding: Universal application of best practice safeguarding principles to ensure secure access to data should help people feel confident to disclose data**

For 'good' to be truly realised, participants felt that a framework such as the [Five Safes](#) should be universally applied for the public to feel confident that public sector data is being used in a way that they can trust. Further, participants felt that even more data collected by public services should be securely stored and linked, and good quality evidence should be shared, in order to inform policy and decision-making. Some participants expressed wanting to maximise the use of available data via more data linkages to better understand multifaceted needs.

Participants had sympathy for the variety of reasons some members of the public may have for not wanting to disclose their data to public services. However, even participants who would prefer not to disclose their data wanted decision-making to be based on evidence representing everyone in society. Participants suggested research was needed to understand why people do not want to disclose data about themselves, alongside greater awareness of the role of administrative data in research for the public good. Related to this were hypothetical discussions around synthetic data potentially filling in gaps where datasets were patchy or lacked enough information. There was consensus that this was not a public good use of data, though participants were supportive of the use of synthetic data for other things, such as training or developing code, as long as synthetic data wasn't being used in place of real data to make decisions.

# Introduction

---

Data about people is generated every day. A record is created when people use a public service which may include their demographic information such as date of birth, age, sex and gender, and further information related to the reason for using that service. This administrative data can lawfully be repurposed for statistics. This type of data can also be used for research, if the purpose meets a number of conditions including being de-identified before it is accessed by a researcher, and serving public good, a condition enshrined into law with the Digital Economy Act 2017 (DEA).

Public engagement and involvement exploring how the public feel about storing, sharing and examining administrative data over the past decades has indicated a general public consensus that data can be used in these ways when used in the name of public good (Aitken et al. 2019; Atkinson et al. 2017; Elias 2021; Kispeter 2019; Scott 2018; Waind 2020).

The question remains as to what the public perceive as 'public good' use of this data when used for research and statistics. Under the DEA 2017 legal framework, 'public good' (sometimes referred to as 'public interest' or 'public benefit') is broad. Legal public good uses of data are: to provide evidence for public policies, services or decisions which benefit our economy, society, or quality of life; to extend understanding of social, or economic trends and events; or to improve quality or understanding of existing or proposed research (UK Statistics Authority, 2018).

The last few years have seen progress in attempting to measure or evaluate the public good use of personal data. A 2018 consultation with professionals from the public and voluntary sectors by Involve, Understanding Patient Data, and the Carnegie UK Trust, generated a framework to evaluate whether public benefit has been delivered by a data sharing initiative in the context of health data. A checklist included criteria such as whether data use demonstrates positive outcomes, delivers what people need or want, or generates tangible, visible improvements for people. Those attributes quantified public good.

Their investigation recommended that the public ought to be part of these discussions to create a shared understanding of what public good means with regard to data use. The public should be involved in shaping the future of (their) data use itself. This principle was a key message from the 2022 DARE UK (Data and Analytics Research Environments UK) dialogue with members of the public: that great efforts should be made to meaningfully include members of the public in decisions about how data about them is used.

Building on that work, ADR UK (Administrative Data Research UK) and the Office for Statistics Regulation (OSR) formed a partnership this year (2022) to explore public perceptions of 'public good' use of data and statistics. This conversation aimed to understand public perceptions of both data use for research and of statistics, which is the information produced from analysis and interpretation of data.

Both organisations have conducted investigations into public feeling about data or statistics use (Cowan & Humpherson, 2020; Waind 2020). Their respective reviews identified a gap in the evidence of exactly what the public perceive to be data or statistics serving the public good.

ADR UK and OSR have therefore collaborated to develop a UK-wide public dialogue with online and in-person workshops in order to build on existing knowledge and create a resource exploring the primary question of what do the UK public perceive as 'public good' use of data and statistics? The project also explored the following sub-questions:

- How should 'public good' be defined and/or measured when making decisions about sharing data for research?
- What uses of data and statistics are considered to be in the 'public good'?
- Are some uses of data and statistics 'more' in the public good than others?
- Are there conceptual differences between the phrases public good and public interest, public benefit, public welfare, common good, greater good, societal benefit or other similar phrases (which are sometimes used interchangeably in the literature)?

**ADR UK** is a programme made up of four national partnerships across the UK, transforming the way researchers access the UK's wealth of public sector data to enable better informed policy decisions that improve people's lives. Key to its mission, is the joining up of the abundance of administrative data already being created by government and public bodies across the UK and making it available to approved researchers in a safe and secure way. ADR UK is therefore bridging the gap between government and academia enabling government policy to be informed by the best evidence available. ADR UK is funded by the Economic and Social Research Council (ESRC), part of UK Research and Innovation (UKRI).

**OSR**, informally known as the 'UK's statistics watchdog', is the regulatory arm of the UK Statistics Authority, which is an independent statutory body that operates at arm's length from government as a non-ministerial department. OSR's vision is simple: statistics should serve the public good. OSR's priorities are to protect the role of statistics in public debate, uphold the trustworthiness, quality, and value of statistics and data used as evidence, and to develop a better understanding of the public good served by statistics.

# Method

---

**This public dialogue included four in-person workshops which took place across the UK in London, Cardiff, Glasgow and Belfast, and one online workshop for those who were unable to join in person. In total, 68 people participated in the initial workshops, and ten were invited to a follow-up workshop to clarify workshop discussions and bring together UK-wide viewpoints.**

## **Deliberative approach**

Deliberative discussion was selected as the most appropriate methodology due to its techniques to generate informed opinions from a starting point of little to no understanding. The issues under discussion in this study were very complex. The terminology, starting with what data and statistics actually mean can be complicated, let alone the laws and common practices within data and statistics use.

The premise of deliberation is to logically and incrementally build up a shared understanding through interactive activities and knowledge-sharing, while giving participants the tools to interrogate their learnings and original viewpoints. These tools are comprehensive yet accessible materials which incorporate both the pros and cons and a more nuanced range of perspectives on the topic of interest, and extensive time and space for reflection and questions. Small groups of differently situated people, supported by a neutral facilitator, allow participants to learn from each other and safely develop their ideas.

The deliberative design fulfils several best-practice criteria produced by recent reviews of public involvement in data research (Aitken et al., 2019; Jones et al., 2020). Members of the public are enabled to feel empowered conversation, rather than the answers being set up for them.

## **Advisory Group**

A Project Advisory Group was created to ensure key stakeholders were involved to inform how the dialogue was conducted. People invited to be members of this group had relevant expertise and the ability to utilise the dialogue findings to inform their respective organisation's processes and policies: members of the Project Advisory Group are listed on page 2.

The ADR UK Public Engagement Steering Group also provided oversight at key points in the project.

## Recruitment

Inclusivity and diversity were high priorities when considering participants for this project, therefore, recruitment was done via a community-based approach. This means recruitment was undertaken mostly by 'community researchers', or lay members of the public living in a city or rural areas in each of the four nations. Up to three community researchers were employed in each nation to produce their own nation-specific recruitment strategy. This involved considering how to break down local barriers, modes of invitations, as well as ways to participate in the project. To support inclusion, the recruitment team had an equal gender split and included a range of age groups, ethnicities, religious beliefs, and interests.

Potential participants were either engaged with directly via local networks and community services, or indirectly via physical leafletting, posters, and online social media groups. To help ensure a range of participants were recruited for this project, community researchers targeted a wide variety of services, community spaces, and local businesses, in locations spanning the UK.

When appropriate, trusted community groups across the UK were also offered compensation in the form of a £20 voucher to help disseminate the project advertisement to networks beyond the community researchers.

## Participants

A total of 72 participants registered for the initial workshops, with 68 attending. Participants were asked to confirm that they were over 18 and were currently living in one of the four UK nations. Fifteen spaces were allocated to people living in each of the four nations of the UK. The other formal limitation on participation was that participants did not work with or study data or statistics, as the aim was to understand the perspective of the general public. To improve accessibility of the events, participants were offered the option to take part online, with consideration for those with at-home caring responsibilities, and special arrangements were offered for those who joined in person. Participant demographics are included in [Appendix B](#). In recognition of their valuable time and input, participants were offered a £150 digital voucher for the initial round of workshops and £60 for the follow-up workshop.

## Procedure of the main workshops

The workshops took place in June, with five workshops repeating the same content and format. Each in-person workshop involved a maximum of 10 participants who were split into groups of four to five participants each; each group had a facilitator and separate note-taker. Participants in the online workshop were split into five breakout rooms of four to five participants, also with a facilitator and separate note-taker per breakout room. The online breakout rooms were assembled to include at least one person from each of the four nations to ensure representation

from across the UK. Both in-person and online workshops ran between 10:00 and 15:30, with a coffee break and a lunch break. All audio was recorded and facilitators followed a script to ensure topics were covered consistently.

Each workshop began with facilitators covering workshop aims and agenda with space for questions and further explanations. Prior to the workshops, participants were communicated with by phone call, text and email, as per their preference, to manage expectations and address any potential concerns.

The workshops included two short 'explainer' presentations from representatives from ADR UK and OSR to contextualise the use of data for both research and statistics. With this exception, participants spent the entirety of the workshop undertaking interactive activities and group reflections. Topics and questions explored are detailed in [Appendix C](#).

Participants were encouraged to explore the nuances of data and statistics use by engaging with hypothetical case studies of different uses of data for research and statistics, and a range of real-life perspectives on data for research and statistics. Participants were given time and encouragement to talk to each other and help each other reflect about each issue, as well as being prompted by their facilitator.

## **The follow-up workshop**

A follow-up online workshop took place in July, roughly a month after the initial workshops, to clarify topics and themes discussed in the initial workshops and bring together UK-wide perspectives. Ten of the original participants were invited to attend the follow-up workshop. As far as possible, a participant was invited from each of the original workshop groups. The follow-up workshop had three aims:

- to validate analysis of the workshops to ensure what participant feedback has been accurately communicated,
- to answer questions about the initial themes, as advised by the Project Advisory Group,
- for participants to explore practical applications of their views with the intention to develop some guidance for how the public perceives 'public good' use of data for research and statistics.



# Analysis

---

Transcripts of the workshop recordings were coded by a qualitative researcher from Kohlrabi Consulting. Based on these codes, two researchers from Kohlrabi Consulting identified key themes and subthemes, along with supporting quotes and examples for context. Themes were validated by workshop facilitators and representatives from ADR UK and OSR; workshop facilitators also contributed to analysis.

The preliminary analysis was presented to the Project Advisory Group ahead of the follow-up workshop to identify areas requiring further clarification. Participants in the follow-up workshops reviewed initial themes and explored outstanding questions. The findings of the follow-up workshop were coded and incorporated into the analysis by researchers, allowing for both interrogation and validation of results.

## **Strengths and limitations of the project design**

The recruitment design for this project aimed to break down barriers to participation, including accessibility of instructions, affordability, feeling welcomed, and flexibility for those who had caring responsibilities. Many participants fed back that they were pleased to have the opportunity to contribute to this topic.

The deliberative design of the workshop enabled participants to incrementally build up their own understanding of the project themes, coming from little or no understanding. It also provided a space for discussion-based learning. The workshop evaluation forms suggested participants felt they were adequately informed to confidently participate in the discussions. Participants particularly valued hearing directly from representatives from ADR UK and OSR, whose explainer presentations aimed to familiarise participants with the existing processes and protocols associated with public good use of data for research and statistics.

Despite the strengths of this project, there were also some limitations. The sample of participants chosen for this project does not represent the demographic breakdown of the UK public (this was not a condition of recruitment) and a small sample cannot represent the entire British public. Further, the fact that each workshop lasted one day meant that the discussion did not explore all possible relevant topics. For example, there was little opportunity to explore research ethics and integrity, and to probe further every time participants referred generally to 'organisations' without clarifying what they meant. However, the findings offer novel insights into a topic which required exploration, and provide a resource for further developing the

understanding of public good.

Despite efforts to replicate the workshop experience for each nation and online, a significant change was made following the first workshop in London. Representatives from ADR UK and OSR remained in the room for the duration of that workshop. Upon reflection, it was decided that they would leave the room after delivering their explainer presentations for subsequent workshops. This was to allay any potential concerns of participants possibly feeling uncomfortable articulating distrust or criticism towards either ADR UK or OSR.

Equipping members of the public with the understanding required for deliberative discussion may carry the risk of silencing participants' instinctual responses to the material presented. To ensure a balance between informed deliberation and more instinctual perspectives, there were two activities aimed at exploring participants' interpretations of 'public good' and their experiences of data collection and use at the start of the workshops before the ADR UK and OSR explainer presentations. This structure enabled facilitators to track if and how participants' views on public good changed as they learned more about it in the context of the use of data for research and statistics.

Similarly, although group discussions help to build understanding, they may carry the risk of some participant's views not being heard. Use of repeated questions in different formats throughout the workshop, including written responses, in addition to calling on each participant to speak supported researchers to draw out individual perspectives.



## Interpretations of 'public good'

Whilst 'public good', 'public interest', and 'public benefit' are phrases sometimes used interchangeably in the literature (Cowan & Humpherson, 2020; Waind, 2020), participants shared different interpretations of these phrases suggesting that they largely considered 'public good' to mean something different to 'public interest' and 'public benefit'.



Analysis of the language participants used when speaking about 'public good', compared with 'public interest' and 'public benefit', suggested that 'public good' meant a wholly good thing, while 'public interest' was often understood as something interesting to the public, but not necessarily something good, while 'public benefit' was often considered as a tangible benefit, but not always good for the wider public.

Participants associated words such as 'advancing' and 'improving' and 'knowledge' with 'public good', which seemed to indicate an understanding of public good being theoretical sometimes, or involving incremental change. The form that 'public good' could take was discussed within the context of achieving 'the greatest good' (see **Real-World Needs**), which indicated that theoretical or actual changes needed to be wholly positive. However, the terms 'public benefit' and 'public interest' seemed to focus on current situations, rather than theoretical or future ones. In contrast, participants expressed that 'public benefit' was something tangible but not necessarily good, or achieving the greatest good. But they associated 'public good' with positive impact, regardless of whether it materialised immediately or in the future (see **Real-World Needs**).

Words such as 'research' or 'data' also carried slightly different meanings for some participants, dependent on personal histories or knowledge. An observation of this discussion is the public may consider commonly used phrases related to data and statistics to have different meanings, which suggests those communicating about data and statistics should not assume knowledge or shared understanding when communicating to the public.

# Findings (expanded)

The following section details the insights derived from this research, supported by anonymised verbatim quotes from workshop participants to illustrate our findings.

**1**

## **Public Involvement: Members of the public want to be involved in making decisions about whether public good is being served**

Participants repeatedly returned to the question of *who* decides what constitutes public good.

The participants felt that, practically, it was implausible for the majority of the public to be formally involved in deciding what public good is within the context of the use of data for research and statistics. Several participants adamantly expressed that the public already have a voice in interpreting what public good is: that of the democratically-elected government who use it on their behalf, though some participants echoed a strong distrust of politicians. This supported discussions about a public role in the guardianship of data and statistics; if the outcomes of their use would affect the society they lived in, then, as citizens, they want their say.

“Things that are interpreting public good, or if it reaches into whether its public good or not, that should be decided by the public.”

— Workshop participant

A shared understanding of the use of data for research and statistics for public good was identified as something which would help people feel part of tacit public involvement and agreement. Across their different backgrounds, the participants established that they had very little knowledge about the use of data for research and statistics before these workshops.

“The public good should not be politicised or defined by politicians. I don't trust MPs; I would rather it was someone working in a coffee shop.”

— Workshop participant

"If they do de-identify it, I'm fine with that. but it's not okay to just allow people to use data in whatever way they want and for me to say it's got nothing to do with me. Its use is affecting my life and the lives of my children."

— Workshop participant

Participants saw responsibility for building public understanding as belonging to everyone involved in handling data or statistics, but one group attracted particular focus. Participants suggested that the public services which collect data, such as GPs, hospitals, and schools, have a responsibility to communicate the purpose for data collection and how it might be used. Many participants articulated that transparent communication 'on the ground' would enable the public to feel respected, and to feel tacitly part of the discussion on whether to use data for public good. It was also wondered whether contentment with disclosure of personal information may increase if services' motivations for collecting data considered more sensitive, such as ethnicity and religion, were transparent. Participants felt that no public service should assume the trust of the public.

"[I want] the conditions for everybody to have this conversation of what underpins the notion of public good. Every citizen should have the ability to make the judgement."

— Workshop participant

The follow-up workshop made clear that participants wanted at least some members of the public to have been consulted in decisions concerning the public good use of data for research or statistics, as they will best understand public perspectives. This was specified as undertaken on a case-by-case basis, over and above the legal framework for data sharing and use. Although there was general support for the DEA framework, most participants expressed that they did not want public good decisions left only to the government, due to concern about politicised or potentially self-serving interpretations of public good.

"It's important for people to speak for themselves. The government might not know what is public good. The people, where the data is coming from, should be able to say something, and not just the government."

— Workshop participant

## Practical applications suggested by participants

Participants expressed a preference for members of the public to be included on public panels, which aim to inform whether a use of data is serving the public good. It was important to participants that panels should include representatives from the groups likely to be affected by the data use, and make efforts to include wider perspectives to potentially mitigate unintended effects (see [Minimise Harm](#)).

In addition to involvement in decision-making, participants wanted data organisations to listen to the public perspective in a variety of settings, using inclusive methods, across the country. For example, they wanted organisations which produce statistics to

proactively seek the views of the public in interpreting statistical patterns, or organisations which enable access to public sector data for research to proactively seek public views in developing and prioritising their areas of research interest. Similarly, researchers from public institutions using such data to meaningfully engage the public with their research.

Participants involved in the follow-up workshop responded positively to being engaged more than once in the project; an ongoing relationship between project partners and participants was viewed as meaningful.

### 2

#### **Real-World Needs: Research and statistics should aim to address real-world needs, including those that may impact future generations and those that only impact a small number of people**

Participants tackled the question of which public is being served by the use of data for research and statistics. In the first round of workshops, there appeared to be several prioritisations:

- the needs of the largest majority versus the needs of fewer people with 'higher needs',
- the current needs of people as opposed to the needs of a future public,
- the needs of people in the United Kingdom compared to people across the globe.

The common principle was the idea of a public good use of data for research and statistics being for 'the greatest good' or the most good possible within a given context. This was still interpreted in a number of ways.

Many participants understood 'the public' to mean the majority of people. For them, the greatest good was the greatest number of needs being met at once. This was seen as fair as there was trust that outcomes of data and statistics use would eventually benefit everyone so eventually society will be a better place. One participant called this "a domino effect".

Others felt as strongly that public good use of data and statistics should be aimed at the highest need, regardless of who or how many might benefit. This was particularly driven by concern about social inequity. However, participants were unable to reach consensus on this, with some participants insisting that it was unfair to intend for a use of data or statistics to serve part of society rather than a general public. Emphasis on inequity ran parallel to a desire for equality; an observation that even participants passionate about addressing inequity felt that benefits for people who are better off are still compatible with public good.

Data use for 'the greatest good' corresponded with some participants' belief that 'the public' does not have to refer to the 'current public'. Using data for research and statistics could intend to improve the lives of future generations, or through international data sharing for people across the world.

However, there was no consensus on this point. A few participants disagreed and suggested the use of data for research and statistics should be limited to covering the interests of the people whose data was being studied. This was partly related to the idea of ownership of data because it was suggested that people should see a return on their 'donation' of data about them (see [Clear Communication](#)).

"Everybody should be able to benefit from it. If just a part of people is benefiting from and the rest are still stagnant, it's not okay, that is not a fair use of data. It's one-sided."

— Workshop participant

Discussions concerning the principle of the greatest good in the follow-up workshop clarified that there was not a set number of people who should be served by the use of data or statistics. Participants agreed that the value of a use of data or statistics could be assessed by need, rather the number of people who would benefit.

Regarding the topic of equitable data use, participants expressed that society is already unequal and inequitable, and that public good use of data for research or statistics should aim to address these issues. Participants also felt that data should not be released for research purposes if it can fuel inequities or inequalities. For example, some uses of data for research may fit into a common interpretation of public good if the application leads to tangible improvements for some, however, participants felt if a consequence of such research were to widen inequality gaps then it would not be considered public good use of data.

Participants communicated that decision-makers granting the use of data for research should have an understanding of the impacts of the project proposed, akin to how local councils fill out an impact assessment for their activities (see [Minimise Harm](#)).

Participants assumed knowledge would help guide decision-makers and enable fair data distribution.

The follow-up workshop reiterated that participants were happy for public good use of data for research and statistics to encompass more intangible uses such as research for understanding, or challenging or validating established evidence, within the context of achieving the greatest good. Participants felt that society would benefit with greater understanding or better evidence for decision-making (see [Real-World Needs](#) and [Minimise Harm](#)). Participants spoke about not wanting to get in the way of science by impeding advances in knowledge.

"Society is unfairly stacked. Things like your postcode, your socioeconomic demographic, or the school you went to can affect your life. You have to recognise this and notions of "public good" should reflect that."

— Workshop participant

"They should have an understanding of the impact of the projects being proposed, a good custodian of data wouldn't release data to a project if it is going to fuel inequality."

— Workshop participant

"Wisdom and maturity are when you plant a tree knowing very well that you might not be there to enjoy its shade. The point is there are future generations that might enjoy the benefits of that research and data."

— Workshop participant

## Practical applications suggested by participants

It was very important to participants that publicly-funded research and statistics using administrative data should aim to address a real-world need, which could include an issue that may impact future generations or one issue impacting a small number of people.

Although participants stressed that no one person should have more of a right to benefit from data for research and statistics than another, addressing social inequality and inequity ranked highly among participants' interpretations of public good. In the follow-up workshop, participants suggested that in

practice, addressing social inequalities and inequities could be prioritised in the process of interpreting whether a use of data for research or statistics is in the public good.

It is important to note that participants did explicitly express distrust for the decision-

making process concerning the sharing and use of data, but they desired a better understanding of how and why decisions are made. It was important to them that this process be publicly communicated and that equity be at the core of decision-making.

### 3

#### **Clear Communication: To serve the public good, there should be proactive, clear, and accessible public-facing communication about the use of data and statistics (to better communicate how evidence informs decision-making)**

Participants were not initially concerned about how the public good use of data for research and statistics would materialise; participants expressed wanting to experience tangible changes. Local service provision, national policymaking, and research with clear applications such as clinical health research, were examples participants listed as uses of data and statistics which generate tangible benefits. Most prominently, communication concerning improvement to quality of life was noted as a good use of data for research or statistics.

The participants' desire to experience tangible changes was fuelled by a sense of disproportionality between the amount of personal data collected about people (ethnicity and religious beliefs was particularly noteworthy), compared with a perceived lack of impact on use of that data to improve people's lives. To illustrate their frustration, participants gave examples of different statistics related to services, such as that of shortages in the NHS and social care staff, or ethnic profiling by police, which were perceived to not be utilised as evidence to improve services.

Participants argued for decision-makers to publicly communicate that they are proactively using statistics on fundamental societal issues, such as poverty, health and social care staffing, and education. Over the course of the workshops, there was a continuous perception that lack of political will or funding was limiting the potential uses of data to serve the public good (see [Minimise Harm](#)).

"I feel like a robot, you know? I feel with the council tax. I feel with my wage is split in half. I feel like I'm giving, I'm not receiving. Working class always struggling. And that's us. That's the public."

— Workshop participant

Recognising the nuances in this discussion, participants wanted members of the public to be able to understand how data was being used for public good. Proactively communicating the decision-making process for the use of data for research and statistics, and what potential uses these decisions had been considered for was important, for instance why some projects are granted data access over others (see Public Involvement). This was viewed as promoting accountability and transparency.

"I wasn't aware until this workshop about the whole process of decision-making about data. I would like to see how the actual decision is made, what it is measured against, if there was a way that the public could see that, and how you actually find that information. If there's more openness around the process then hopefully more trust could be built."

— Workshop participant

Participants expressed a desire for more transparent and accessible public-facing communication. Participants indicated an interest in information translated into a range of languages, communicated via a range of channels, and packaged in a range of formats that accommodate people with different sight, hearing and learning needs. For participants, this meant prioritising offline methods for information sharing, such as physical information leaflets at places where data is collected, such as job centres or GP practices, each with contact information welcoming questions or feedback from the public about how their data might be used for research purposes. They suggested that online notices or webpages on service websites should also be more accessible.

"People need to be able to understand it. I think we need to change how we're talking to people. It needs to be in a language that the community understands. And not just in language, but for other needs for people that are blind, for people who have dyslexia. I don't see anything set up that way."

— Workshop participant

Participants expressed that greater understanding of data and statistics would enable them to make decisions about their lives, and were interested in increasing their own usage of statistics. Participants wanted to be able to make decisions about their lives using statistics without fear of confusion, misinterpretation, or politically-biased narration. They felt this would empower them to both better understand the use of statistics in decision-making, but also to maximise their own use of statistics for personal decisions.

Perceived missed opportunities to serve public good, or miscommunication of an activity, were identified as a cause of disengagement between the public and those working with data and statistics; participants understood this as eventually resulting in less public good being served. Participants spoke of hearing about statistics in the newspapers, which they understood as a call to action, and of sharing their data every time they attend a public service, which they understood as burdensome. The perception that no action followed the collection of this data for research or the production of statistics was understood as the motivation for some choosing not to disclose their data with public services. Despite these concerns, several participants recognised that change is a slow process which might be occurring unbeknownst to them.

"Because on the news they'll say, 500 black people out of so many get stopped. but knowing that information hasn't stopped the police from doing that."

— Workshop participant

"When we give our data, we don't ever find what happens with the results. And it's so important, you know, that we, as the public are made aware. Implementations within policy, or strategy or law even need to be shared for those people who have given that data over."

— Workshop participant

"I'm giving information, but nothing is being done according to my needs, the service provision doesn't meet any of my needs. Then tomorrow you come again to ask for information and we don't want to [give it] because we don't know what you'll do with the information. Then we become hard to reach communities and the cycle just perpetuates."

— Workshop participant

"I think a lot more could be done in a sort of more productive way, in a more positive way, with the data, to inform policy and, and to sort of provide a direction in regards to, you know, what sort of areas could be improved with that data."

— Workshop participant

"It's important to say to people okay we've collected this data and this is what we've discovered and this is what needs to be done. 'We are working on it but we need so much time or money to do this, but this is what the end result will look like. It's not going to happen today because there is a process, but we are aware that this needs to be done.' So, feedback."

— Workshop participant

"I think unless you understand statistics, you wouldn't understand that actually [information was derived] through statistics. And they just give that headline without saying what was excluded from it. Then the public is being misled so it's not public good."

— Workshop participant

## Practical applications suggested by participants

Participants wanted statistics and research using administrative data to be communicated in a clear and accessible way to the public. As far as possible, participants felt that the public want to hear what changes to evidence, understanding, decision-making or implementations of policy have resulted from the use of data for research or statistics. Participants acknowledged that greater awareness of data safeguards, and an understanding of why their data is collected, may yield more support for the public to share their data for research. In these discussions several participants asked "why do they want to know that?" querying why certain data is collected.

Participants perceived a sense of accountability associated with public awareness of the use of data for research and its practices, as it enables the public to make informed opinions. There was a strong appetite for those working with data and statistics to broaden their reach via offline communication, such as information leaflets or posters in public services, to explainer videos

and digital infographics displayed prominently on websites.

Participants wanted to better understand the entire process of using data for research or statistics, including why data is being collected, the problem being addressed with the data, timelines of intended activity, and potential (or actual) results. Participants felt that members of the public would like to be able to individually determine whether the use of data for research or statistics had served the public good, to the best of the producers' ability. This meant having the knowledge to generate an informed opinion. For instance, some participants spoke about how they felt confident making informed decisions during the pandemic based on their understanding of Covid-19 statistics.

Participants also suggested organisations working with statistics should be vocal in mainstream news and communication channels in their critical appraisal of statistics to help individuals make their own assessments of statistics.

## 4

**Minimise Harm: Public good means data collected for research and statistics should minimise harm**

"It's also important to consider the tone of communicating. When you are speaking to communities facing multiple disadvantages, living in deprived areas, you must be very careful about the tone of your language, avoiding stigmatising use of the data if the picture it paints is grim, is very important."

— Workshop participant

Many participants communicated concern about potential negative unintended effects that a use of data for research, or an interpretation of statistics, could have on interconnected areas of society, either immediately or over time. For some participants, despite de-identification, they saw public sector administrative data as belonging to them and therefore wanted a say in how it is used. Even those who considered de-identified data as "*just numbers*", felt a personal responsibility that data about them should not contribute to something harmful, akin to other moral life choices such as reducing our carbon footprint.

Participants shared genuine fears that well-intentioned uses of data aimed at particular groups may inadvertently result in negative impacts by stereotyping and discrimination.

This potential harm appeared to be a greater concern than data breaches or data loss. The Five Safes framework, a set of best practice principles developed by the Office for National Statistics to facilitate responsible sharing and use of data, includes checking that individuals cannot be identified in outputs generated from data analysis ([UK Data Service](#)). The Five Safes framework was included in the ADR UK explainer session. Participants reacted positively towards this but wanted a further step to be included regarding responsible use of language. Specifically, more sensitivity around labelling was strongly recommended.

"Irrespective of whether my name is attached to it, I hate to think that my data is contributing to something that harms someone. If it's not being used for good, then I'm part of that bad. I want to feel ethically good about what I take part in."

— Workshop participant

"If you take a case of 100 numbers, you see Somali, Muslim, you see the good ones aren't mentioned. You've got the good and the bad in everything, but you just see the bad."

— Workshop participant

Participants were frustrated with experiences of stereotypes and assumptions about others. This related to their discomfort about missing data, as they felt decisions may be made about some people based on assumptions rather than on evidence (see [Best Practice Safeguarding](#)). Concern about inaccuracies undermined a growing recognition of the public good of using data for research or statistics to validate or challenge accepted evidence. This was thought to be of value through helping research and statistics be more truthful and representative of the public.

Some participants stressed they did not solely want quantitative data to decide a course of action, as they feared it may be misinterpreted. They emphasised the importance of contextualising data with public engagement; they stated a preference for consulting groups likely to be impacted by the data, to validate quantitative data, rather than designing a policy, service, or piece of infrastructure based on research from data and statistics alone. This might involve, for example, asking local people if they will use a service that the data suggests they might benefit from or asking people with lived experience of an issue why statistics are showing a particular association rather than making assumptions.

"I detect some harmful consequences from the language used in studies. If, for instance, those labelled as deprived and poor areas attract that negative media attention connected to the populations living there in, you know, labelling them as poor, you could subject them to stigmatisation and turn them into targets of attacks."

— Workshop participant

"And then I think for it to be interpreted by a diverse group of people from different backgrounds and different political opinions and different types of jobs. And from there you can decide where to act."

— Workshop participant

These points related to earlier concerns about missing data (see [Best Practice Safeguarding](#)), and the desire for data and statistics use for the greatest good (see [Real-World Needs](#)). Without people being included in the interpretation of statistics that are used for evidence-based

decision-making, those people do not have a public voice and decisions might be made that do not positively impact their needs. One participant felt that if data or statistics are used for the greatest good, that good will eventually reach everyone by improving society as a whole. Others felt that this was not enough; that some people were not being impacted at all. This relates to an earlier finding to improve the representativeness of statistics, research, and decision-making.

In terms of more visible harms from data use, of data breaches or data loss, the majority of participants felt the framework of the Five Safes allayed their concerns. The only improvement suggested was greater accountability, or public awareness of existing accountability frameworks, of data protection. In the follow-up workshop, participants wanted to be reassured that each step of the Five Safes was practiced as laid out in the framework. This conversation returned again to independent regulators. It was also suggested that transparent, publicly-available whistle-blowing policies and named leads would reassure the public that organisations using administrative data were taking protection from harm seriously.

Unintended consequences from interpretations of statistics were identified as another risk. Two participants spoke of people they knew changing their behaviour in a way that harmed their well-being after learning about some newly published statistics. For example, one participant knew young people who had begun to carry knives for protection after reading that statistics had shown knife-carrying had increased. Participants suggested that, before publication, statistical outputs should be sense-checked by a diverse group of people as one way of mitigating potential risks.

"There's a massive amount of missing data, for example on the traveller community, you haven't got any responses at all, we don't know what their NHS experiences are like. The risk of that is that decision-making is made without you."

— Workshop participant

"There's still a worry on who checks that they're being practiced. In other sectors there's always a clear whistle-blowing process. To show that they are aware that there is a possibility for misuse, that they have got a process for if something happens. The organisations show that they have it at the back of their minds, a platform of what to do."

— Workshop participant

"The statistician doesn't have that lived experience. We are then assuming that this person has gone out of their way to read up on the challenges of these different groups of people."

— Workshop participant

Although some participants were reassured after learning about the role of OSR in regulating statistics for the public good, many still felt that some public figures were misusing statistics without any repercussions. Some participants stated those who misuse statistics should be fined, or have more public notoriety due to their wrongdoing. There was not a defined suggestion for action but expressions of frustration about being misled by public figures.

## Practical applications suggested by participants

Participants concluded that potential harms of data use should be anticipated before access to data for research or statistics is authorised, perhaps by consulting a variety of groups or any groups relevant to the data or statistics. Participants wondered if that information could be fed into the public good test that is part of the Five Safes process, adopted by many organisations responsible for making data available to use for research and statistics. Participants were also extremely concerned about the potential damage of research or statistics outputs if they were misinterpreted. Participants wondered whether members of the public with relevant lived experience could be consulted as part of the interpretation and publication of statistics. This would help data users avoid language or interpretations that might fuel stigma or discrimination, and anticipate any other potential harms. A separate use for this suggestion was to validate the interpretation of the statistics with the population they concern. This might

involve, for example, asking local people if they will use a service that the data suggests they might benefit from, or asking people with lived experience of an issue why the data is showing a particular association, rather than making assumptions.

In addition to the Five Safes framework, the participants argued for accountability when harm is caused. Explicit suggestions were that organisations should make publicly available whistle-blowing policies, clear consequences of data misuse, and name the people who are ultimately accountable for any breach of the Five Safes. This is also in the context that whilst the UK Statistics Authority is responsible for the application of the Digital Economy Act as a legal gateway to access data for research and statistics, there is no such body responsible for the overall application of the (much more widely used) Five Safes framework across the UK, though individual checks exist across the stages of the Five Safes framework.

## 5

**Best Practice Safeguarding: Universal application of best practice safeguarding principles to ensure secure access to data should help people feel confident to disclose data**

Participants expressed a hope for data for research and statistics to be supplying the best evidence to all decision-makers, including members of the public. Three sub-themes emanated from these discussions: data safeguards, missing data and missed use of data.

**Data safeguards**

Despite varying degrees of scepticism around data sharing and use, with some considering de-identified data as 'just numbers' and others as 'my data', there was consensus that robust safeguards should be in place to protect data and individual's privacy. The Five Safes framework was shared with participants as an example of how some organisations work with data.

"The Five Safes were reassuring to me...For me it was a solid thing, I wouldn't be afraid of where the data is going after that."

— Workshop participant

As detailed in the earlier themes (see [Real-World Needs](#) and [Clear Communication](#)), participants thought that if public services openly communicated reasons why data was used then they would be more likely to disclose their information. Many participants also articulated that they would feel more comfortable about disclosing data knowing a security framework such as the Five Safes was being practiced.

**Missing data**

Maximising evidence was seen as vital to data and statistics being used responsibly. Many participants related missing data at community level, and poor research and statistics, to incorrect understanding and decision-making. Even those who did not like sharing their information wanted to improve, what one participant referred to as, the "*the real data*". Although some participants referred to their own discomfort or fatigue with disclosing data to public services, there was a sense that more data could be better used to serve public good. A number of participants perceived that patterns could not be examined to

"If you have good data you should be able to see all the links, the dynamics. For example, with competing needs, if you have data about cyclists and about road users, you should be able to meet all the demands."

— Workshop participant

uncover unintended consequences, and plans could not be made for the future, if data was not “integrated” between services (see [Real-World Needs](#)).

Some participants expressed a range of discomfort with disclosing some types of data which relate to how a person identifies themselves. Some participants felt disinclined to provide their data as a by-product of negative treatment by public services, or because services had not explained why they needed data, further underlining the findings in [Clear](#)

[Communication](#). Participants with teenage children spoke of their teenagers not comprehending why they should provide their data, so instead providing incorrect data or leaving sections of a form blank. A significant proportion of participants expressed concern that demographic data could be used to discriminate, or could feed into the stereotyping of certain groups (see [Real-World Needs](#)). However other participants insisted that the existence of structural discrimination meant that data collection is very important.

“They're unnecessary. You're asking what religion I am, where I'm from, am I really British?”

— Workshop participant

“My 16-year-old daughter came to me saying why do they want to know if I'm straight, if I'm Catholic, is it because I'm gay are they going to give me the job? She couldn't understand it no matter how I tried to explain it to her. She just thought it was prying. Should I put my weight down? She was getting angry about it. It doesn't affect how I serve coffee.”

— Workshop participant

A common assertion was that, if poor representation within data was the result of people failing to disclose personal information, then researchers should make more effort to investigate why some people choose not to disclose their data. It was expressed by participants that engagement with these people could help their voices be heard in decision-making, and progress understanding of how to address missing data. As detailed in the earlier themes (see [Real-World Needs](#) and [Clear](#)

[Communication](#)), participants thought that if public services openly communicated reasons why data was used then they would

“I'm not disclosing. Instead of looking to make up the numbers go and ask why people aren't disclosing.”

— Workshop participant

In the follow-up workshop, participants explored the use of synthetic datasets

(see [Appendix A](#) Glossary of terms) to compensate for missing data. These were viewed as protecting the public, as they did not contain information on personal identities, while being valuable to science in supporting analysis and researcher development. Participants stated key to synthetic data serving the public good was it not being used in place of real data or used in place of actual evidence, whether that be for service provision, the advancement or correction of knowledge, or policy and decision-making. Participants also emphasised that it should not take away from the need to address the reasons behind missing data.

"I would be more comfortable filling in [a form if I knew why] I'm being asked for that information. It would be helpful for organisations like ADR UK to have programmes, short clips, videos, about data and how they use it. And adverts online. Posters. So that everybody everywhere knows it's actually useful for me to give my data. It's put into people's minds on a daily basis in different ways."

— Workshop participant

As reported in the [Clear Communication](#) finding, in the follow-up workshop participants articulated that they were happy with validating or challenging existing evidence being a public good use of data for research or statistics. It was felt to fit into the participants' motivation for truth and honesty in statistics; attempting through research to develop better quality and more up to date estimation of the facts. If updating statistics allowed for more people's data to be included, some participants were quick to wonder if this would make societal knowledge more representative of society.

## Missed use

The missed use of data, or the deliberate not sharing of data that could be used for research and evidence-based decision-making (Morrow, 2020), was viewed by many participants as harmful. There was concern that the public good could be undermined by having a process that was too lengthy to access data for research, or restrictive as a result of public service organisations or data custodians refusing or delaying access to data. Participants stated data sitting unused in storage was not a good use of resources. Although participants did not want robust data

"There's no point in finding out how many children have been put in child poverty a year and a half from now. We need to focus on the data that's really needed at that time and, and try and push it through to justify spending on the communities."

— Workshop participant

security compromised, they wondered if any part of the process could be sped up for priority issues.

Some participants suggested that well-resourced private organisations might be able to maximise what could be achieved on behalf of the public sector. Among this group were people who believed private organisations to be more transparent and more efficient than the public sector organisations. Caveats were that there would have to be adequate restrictions and regulations in place and that public organisations would need to be transparent about the involvement of the private sector, and would need to be able to demonstrate public good values.

"If I give my data, I want to see results, I want to see positive results, that's what I expect, that's fine. When it is monetary value, or normal interest, it's fair enough, as long as it's positive."

— Workshop participant

In contrast, other participants were vehemently opposed to private organisations getting involved with public sector data, as they anticipated the profit motive would take precedence over truth.

A number of technical issues in data for research and statistics were explored to understand how they fit into discussions concerning public good. Participants perceived that maintaining datasets (in a secure way), rather than deleting them after use, and re-using datasets to validate research in different populations, served the public good. However, they argued that organisations keeping the data should be a trustworthy public or publicly-funded body. According to participants, private organisations should therefore not be allowed to keep datasets without explicit consent.

## Practical applications suggested by participants

For 'good' to be truly realised as the participants understood it, it was felt that a best practice framework, such as the Five Safes, should be a universally applied framework to allow the public to know that publicly collected data is being used in a way

that they can trust.

As detailed earlier, greater awareness of data safeguards, and an understanding of why their data is collected, may yield more support for the public to share their data for research (see [Clear Communication](#)). To further improve the

representativeness of research and decision-making, researchers should make more effort to consult with people who were less likely to disclose their information. This would help unheard voices inform the evidence being created, and progress understanding of how to address missing data.

Data custodians could explore how they could safely share and link more data to become

evidence to be used for public good. This would enable policy and decision-making that centres on people's needs, and was hoped to address the sense of disproportionality between the volume of data collected from people and the benefits experienced in everyday life. created, and progress understanding of how to address missing data.



# Summary

---

**This report, commissioned by the partnership of ADR UK (Administrative Data Research UK), and the Office for Statistics Regulation (OSR), culminated from the need to provide a public, shared understanding of how the public themselves interpret the public good. The primary research question, what do the UK public perceive as ‘public good’ use of data for research and statistics, was explored with 68 participants from a range of backgrounds, who are routinely asked to contribute their data when interacting with public services.**

Participants felt that members of the general public should be involved in the interpretation of the public good, as they did not think a concept and responsibility such as this should be politicised or decided on behalf of the public (see Public Involvement). Participants were clear that the public good use of data for research or statistics should be assessed by need, rather than the number of people who would benefit (see [Real-World Needs](#)). Participants also suggested that, in order for data for research and statistics to serve the public good, organisations who collect, use, or produce data for research and statistics have a duty to communicate their work in an accessible way with members of the public (see [Clear Communication](#)). Many participants felt a personal responsibility that data about them should not contribute to something harmful (see [Minimise Harm](#)). Lastly, participants suggested that more data collected by public services should be made available in order to shape policy and decision-making that centres on people's needs, ensuring that this is done in a responsible way (see [Best Practice Safeguarding](#)).

Participants also discussed many different ideas for how their thoughts on public good could be implemented. Some of these points are listed below:

- Members of the public should sit on public panels to contribute to deciding whether data for research and statistics is serving the public good
- Decisions on whether data for research and statistics are serving the public good should prioritise whether social inequality and social inequity are being addressed
- Critical appraisals of statistics should be more prominent in the public eye to aid individuals in making their own assessments of the public good of statistics
- Proactive public-facing communication methods for a lay audience, including tested language, should be built into standard communication strategies. Participants wanted to

experience a tangible benefit resulting from their data being used for research and statistics, and want to hear more about this, including through physical communication strategies (e.g., displaying leaflets in public areas)

- In order to identify unforeseen problems with data for research and statistics, consultation including those with lived experience should be carried out and their evidence should be used to help mitigate potential harms
- A best practice safeguarding framework to ensure the security of data, such as the Five Safes, should be universally applied and understood to allow the public to know that publicly-collected data is being used in a way that they can trust. This is in addition to the oversight provided by UK Statistics Authority for the application of the Digital Economy Act as a legal gateway to access data for research and statistics. To promote accountability of responsible use of data for research and statistics, those enabling access to data or producing statistics should support mechanisms for whistle-blowing and public punishing of harmful behaviour.

## Next steps

For ADR UK and OSR, these findings offer important insights into what these participants, who are members of the public who do not work in data and statistics, think about the public good, and how they think the benefits associated with data and statistics can be maximised. ADR UK and OSR will consider these findings, alongside other evidence, to inform how their work, and the work of others, can maximise the benefits associated with data for research and statistics for the public good.



# References

---

Aitken, M., Jones, K. H., Kaarakainen, M., Lugg-Widger, F., McGrail, K., McKenzie, A., Moran, R., Murtagh, M. J., Oswald, M., & Paprica, A. (2019). Consensus statement on public involvement and engagement with data-intensive health research. *International Journal of Population Data Science*, 4(1).

<https://ijpds.org/article/view/586/2829>

Atkinson, S., Badger, S., Milne, R., & Brayne, C. (2017). Ethical, legal and social issues in dementia research. Dementias Platform UK.

<https://www.dementiasplatform.uk/about-us/our-achievements/research-project-reports>

Cowan, M. L. & Humpherson, E. (2020). The Public Good of Statistics. What we know so far. Office for Statistics Regulation.

<https://osr.statisticsauthority.gov.uk/publication/the-public-good-of-statistics-what-we-know-so-far/>

Data Analytics Research Environments UK & Kohlrabi Consulting. (2022). Building a trustworthy national data research infrastructure: A UK-wide public dialogue.

[https://dareuk.org.uk/wp-content/uploads/2022/05/DARE\\_UK\\_Building\\_a\\_Trustworthy\\_National\\_Data\\_Research\\_Infrastructure\\_Public\\_Dialogue\\_May-2022.pdf](https://dareuk.org.uk/wp-content/uploads/2022/05/DARE_UK_Building_a_Trustworthy_National_Data_Research_Infrastructure_Public_Dialogue_May-2022.pdf)

Elias, P. (2021). Promoting public engagement with longitudinal research: A report to the Economic and Social Research Council. Warwick Institute for Employment Research.

<https://www.ukri.org/wp-content/uploads/2022/03/ESRC-020322-Promoting-public-engagement-with-longitudinal-research.pdf>

Health Data Research UK. (2021a). Population Research UK. Design and Dialogue Consultation.

<https://www.hdruk.ac.uk/wp-content/uploads/2021/12/Annex-1-PRUK-Design-and-Dialogue-first-consultation-summary.pdf>

Health Data Research UK & Kohlrabi Consulting. (2021b). Public and Participant Involvement in Population Research UK (PRUK): A scoping review of different approaches and recommendations.

[https://www.hdruk.ac.uk/wp-content/uploads/2021/07/Annex-2-PRUK-PPIE-Scoping-Review\\_final.pdf](https://www.hdruk.ac.uk/wp-content/uploads/2021/07/Annex-2-PRUK-PPIE-Scoping-Review_final.pdf)

Involve, Carnegie UK Trust, & Understanding Patient Data. (2018). Data for Public Benefit.

<https://www.carnegieuktrust.org.uk/publications/data-for-public-benefit/>

Jones, K. H., Heys, S., Thompson, R., Cross, L., & Ford, D. (2020). Public involvement and engagement in the work of a data safe haven: a case study of the SAIL Databank. *International Journal of Population Data Science*, 5(3), 1-10.

<https://ijpds.org/article/view/1371/2815>

Kispeter, E. (2019). Public support for accessing and linking data about people from various sources: Literature review. Warwick Institute for Employment Research, University of Warwick.

<https://www.ukri.org/wp-content/uploads/2022/03/ESRC-020322-Public-support-for-accessing-and-linking-data-about-people-from-various-sources.pdf>

Morrow, E. (2020). Administrative Data: Missed use vs miss use. Administrative Data Research UK.

<https://www.adruk.org/news-publications/news-blogs/administrative-data-misuse-vs-missed-use-133/>

Scott, K. (2018). Data for Public Benefit: Balancing the risks and benefits of data sharing Involve, Carnegie UK Trust and Understanding Patient Data.

[https://www.involve.org.uk/sites/default/files/field/attachemnt/Data%20for%20Public%20Benefit%20Report\\_0.pdf](https://www.involve.org.uk/sites/default/files/field/attachemnt/Data%20for%20Public%20Benefit%20Report_0.pdf)

Waind, E. (2020). Trust, Security and Public Interest: Striking the Balance; A review of previous literature on public attitudes towards the sharing and linking of administrative data for research. Administrative Data Research UK.

[https://www.adruk.org/fileadmin/uploads/adruk/Trust\\_Security\\_and\\_Public\\_Interest-Striking\\_the\\_Balance-ADR\\_UK\\_2020.pdf](https://www.adruk.org/fileadmin/uploads/adruk/Trust_Security_and_Public_Interest-Striking_the_Balance-ADR_UK_2020.pdf)

UK Data Service. What is the Five Safes framework?

<https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/> [last accessed 15 September 2022]

UK Statistics Authority. (2018). Research Code of Practice and Accreditation Criteria.

<https://www.gov.uk/government/consultations/digital-economy-act-part-5-data-sharing-codes-and-regulations/research-code-of-practice-and-accreditation-criteria>

# Appendix A: Glossary of terms

---

**Accredited researcher:** Someone who has been trained to carry out quantitative research. Their credentials have been approved by an independent body, the UK Statistics Authority. They are allowed to securely access to de-identified unpublished data for a specific research purpose under the Digital Economy Act 2017 and Statistics and Registration Services Act 2007.

**Administrative data:** Information created when people use public services, such as schools, hospitals, the courts, or the benefits system.

**Data Access Committee:** Data Access Committees evaluate applications from trained and accredited researchers for the use of de-identified data for research.

**De-identified data:** Personal information such as names and addresses have been removed from the data before it is shared with accredited researchers so that the data do not directly identify individuals and are not reasonably likely to lead to an individual's identity being ascertained (whether on its own or taken together with other information).

**Statistics:** Producing information from data. For example, collecting everyone's age in the room is an example of data, but using that data to calculate an average age makes it into a statistic.

**Synthetic data:** Synthetic data is a version of a dataset that uses made up data rather than actual data, ranging from very low to very high levels of fidelity. The made-up data is generated at random and is made to follow some of the patterns of the original dataset. Like any data, synthetic data can only be accessed with permission and with the right kinds of safeguards around it. If synthetic data is shared with researchers, this is on the understanding that this is not real data, and is only being shared to raise awareness about how the real data is structured to support training and engagement. It is never 'passed off' as real data.

# Appendix B: Participant demographic information

## Age

18-24: 6  
 25-34: 20  
 35-44: 19  
 45-54: 8  
 55-64: 13  
 65-74: 1  
 Blank: 1

## Residence

England: 19  
 Wales: 19  
 Scotland: 16  
 Northern Ireland: 14

## Ethnicity

Arabic: 2  
 Black, African Caribbean: 18  
 Chinese: 1  
 Human: 2  
 Middle Eastern: 4  
 South Asian: 16  
 White: 21  
 White European: 4

## Gender

Female: 40  
 Male: 28

## Social position

Low income: 6  
 Working class: 20  
 Working class, educated: 10  
 Middle-class: 19  
 Homemaker/carer/ill: 8  
 Retired: 5

Note: a) Social position and ethnicity were free text responses; responses for the remaining questions were pre-fixed categories.

The attendees were asked to complete a voluntary, anonymous demographic information form. The demographic characteristics of those who responded indicate a 60:40 female to male ratio. Participants were mostly young to midlife adults, with two thirds of participants between 25-44 years: few being under 24 or over 64. Social class responses showed roughly 53% defining as working class or low income. Around 80% were working, with those not working retired, caring for relatives or small children, or unable to work due to ill-health.

# Appendix C: Workshop agenda

---

## Workshop content

### Sessions

1: Group introduction

### Questions explored

"What does 'public' mean? What does 'good' mean? What do they mean together?"

"Do a certain number of people need to benefit for something to be considered in the 'public good'?"

"'Public good' is something referred to by other phrases, such as 'public benefit' or 'public interest'? Do either of these phrases have different connotations to you than 'public good'?"

2: Public services data activity

"How do you think that data about you might be used for public good?"

"Can anyone provide a reason why they would not want their data to be used?"

3: Data explainer and Q&A

4: Reflection on data use

"Does data use count as 'public good' if some people benefit while others' situation remains unchanged?"

"Does it matter if the organisation using data for public good also makes money as a by-product of that use?"

5: Data explainer and Q&A

6: Reflection on statistics use

"If a use of data is considered to be strongly in the public good, would you be comfortable for sensitive data to be used for that purpose?"

**Sessions**

7: Case studies of data and statistics use for public good

8: Perspectives on data and statistics use for public good

9: Final reflections

**Questions explored**

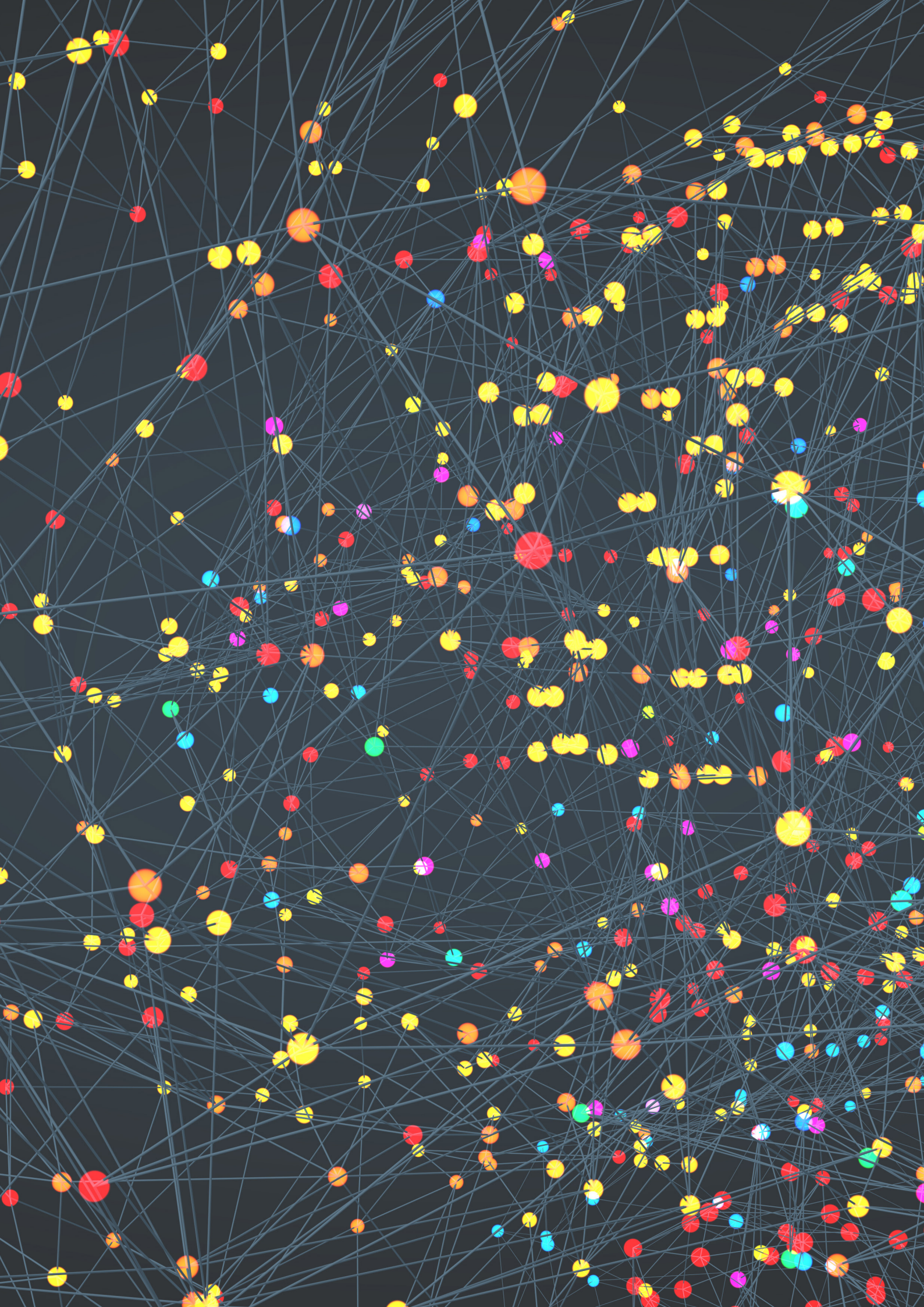
"If a use of data is considered to be strongly in the public good, would you be comfortable for sensitive data to be used for that purpose?"

"Does this view seem surprising to you? Why do you think this person feels this?"

"Is this view part of your personal definition of what public good is?"

"How would you tell that something is an acceptable use of data or statistics?"

"In your dream society, what could data and statistics be achieving for people?"





## Get in touch:

-  [press@adruk.org](mailto:press@adruk.org)
-  [adruk.org](http://adruk.org)
-  [@ADR\\_UK](https://twitter.com/ADR_UK)
-  [ADR UK \(Administrative Data Research UK\)](#)
-  [ADR UK](#)

-  [regulation@statistics.gov.uk](mailto:regulation@statistics.gov.uk)
-  [osr.statisticsauthority.gov.uk](http://osr.statisticsauthority.gov.uk)
-  [@statsregulation](https://twitter.com/statsregulation)
-  [Office for Statistics Regulation](#)