open data institute

Luminate
Building stronger societies

RSA
21st century enlightenment

About Data About Us

By Renate Samson, Kayshani Gibbon and Anna Scott

September 2019

Contents

About this report

The About Data About Us report was produced by the Open Data Institute (ODI) and the RSA (Royal Society for the encouragement of Arts, Manufactures and Commerce) and was commissioned by Luminate. Research was carried out by the ODI, the RSA and Luminate. The report was written by: Renate Samson (ODI), Anna Scott (ODI) and Kayshani Gibbon (RSA), in collaboration with Jeni Tennison (ODI), Peter Wells (ODI), Kitty von Bertele (Luminate), Toby Murray (RSA) and Charlotte Holloway (RSA)

About the ODI

The ODI works to build a strong, fair and sustainable data economy by helping governments and businesses around the world get data to people who need it. It is independent, nonprofit and nonpartisan, founded in 2012 by Sir Tim Berners-Lee and Sir Nigel Shadbolt. From its headquarters in London and via its global network of startups, members and nodes, the ODI offers training, research and strategic advice for organisations looking to explore the possibilities of data.

About the RSA

The RSA believes in a world where everyone is able to participate in creating a better future. Through our ideas, research and a 30,000 strong Fellowship we are a global community of proactive problem solvers, sharing powerful ideas, carrying out cutting-edge research and building networks and opportunities for people to collaborate, influence and demonstrate practical solutions to realise change.

About Luminate

Luminate is a global philanthropic organisation focused on empowering people and institutions to work together to build just and fair societies. We support innovative and courageous organisations and entrepreneurs around the world, and we advocate for the policies and actions that will drive change across four impact areas: Civic Empowerment, Data & Digital Rights, Financial Transparency, and Independent Media. We work with our investees and partners to ensure that everyone has the opportunity to participate in and to shape the issues affecting their societies, and to make those in positions of power more responsive and accountable. Luminate was established in 2018 by Pierre Omidyar, the founder of eBay, and his wife Pam. The organisation is part of The Omidyar Group. www.luminategroup.com

Executive Summary

Our data lives are complicated.

Never before has data played such an integral and granular role in how we live. On a daily basis, we are asked to make decisions about personal data about us – consenting to it being gathered and used for many purposes.

We are only just starting to grasp the impact that these decisions have on us, and others. We must think differently about data, and the rights and responsibilities around it.

We must engage with and listen to people about how they feel, and stop writing off 'the public' as being complacent or ignorant about data protection issues, as they often are by people in power and in the media.

Recently, 'data ownership' has been raised by some as a possible way to give us control over the data about us. But, given that data about us is rarely just about us as individuals, but usually about other people too – many have criticised ownership as an overly simplistic solution. Critics have said that instead we need to strengthen our 'data rights' and the responsibilities to maintain them, with a more systemic approach including legislation, regulation, policymaking, education, and advocacy.

We explored how members of the UK public feel about data about them, about having ownership or rights around it, and what kind of control or protection they feel is missing or needs strengthening. We did this over the course of two focus groups and a workshop in London.

To help, we developed a graphic that explains the different types of data about us: '**personal data**', '**sensitive data**', '**behavioural data**', and '**societal data**'. We also tested and developed compelling narratives to help people understand and explore these different types of data in context.

We did this because we saw it can be hard for people to decide how they feel about sharing data about them, without being able to consider the different elements or ways it is used. Sharing sensitive data about us so a company can target us with adverts is different from sharing societal data about us (data which should be aggregated and anonymised) so it can be used to improve public services for everyone.

People told us that they generally feel positive about the benefits brought by the internet and being more connected, but want greater **honesty and transparency**, **agency and control**, **rights and responsibility**, **context and fairness**, and **compliance and enforceability** over how data about them is used.

Ultimately, they want to know that where data is concerned they will be treated as people, not as robots.

This report is part of a range of outputs, including a video, a summary report and a graphic explaining the types of data about us.

We hope this work will help to start a wider conversation between people, governments, businesses, NGOs, interest groups, and think-tanks. Tweet your views using #WeAreNotRobots.

Introduction

Data is everywhere. It is part of the infrastructure of all of our lives, institutions, public organisations and private businesses. We are all involved – often unknowingly – in its creation, management, and use. Just as roads connect us, data connects us. Just as roads generate noise and air pollution, data can create equivalent harms or risks to individuals and communities alike.

What is data? We use the term to cover a range of things. Some data is about us as individuals, some data had at one time been about us but we are no longer recognisable in it, and some data has nothing to do with us. Data can be personal, sensitive, behavioural or societal. It can come in many shapes and sizes, and be accessible to many or only a few. The ODI's Data Spectrum,[1] shows the different levels of data access, from closed, to shared, to open. Closed data is restricted to an organisation; shared data is accessible to specific people or groups; and open data is data anyone can access, use and share.

This project, undertaken by the RSA, the ODI, and Luminate has sought to learn from members of the UK public about their relationship, thoughts, and feelings about data.

We have explored how members of the UK public feel about the idea of having rights over data about them, or ownership over it. We tested people's understanding, reactions and emotions, and sense of responsibility over data about them, and how they felt about it being used for different purposes in society – from private sector companies, to employers, to public authorities.

We also wanted to give people space to explore what kind of control, insight, protection, or security they felt was missing or needed strengthening.

We focused on speaking with people in the UK for this initial tranche of research, but the data rights/ownership debate is not confined to the UK – it is an international conversation and we recognise that the views of people will vary from country to country.

This report explains our research and how we developed different ways of describing the different types of data about us – from personal to sensitive, behavioural to societal – to help people differentiate and explore what types of data about them they felt comfortable or uncomfortable being shared or accessed. We had found that the lack of these clear definitions made it hard for people to decide how they felt.

We have relayed what the public told us, in their words, about what they wanted to see happen next in terms of greater rights, clearer responsibilities, and enhanced protections when it comes to data about us. This report is part of a range of outputs, including a video, a summary report, and a graphic explaining the types of data about us.

1. Open Data Institute (2019) *The Data Spectrum*. [online] Available at: theodi.org/about-the-odi/the-data-spectrum

Background

What is data about us?

We might think of data – particularly personal data – as being solely about us as individuals. The fact that it is referred to as 'personal' means we often refer to it as 'my data' or 'data about me', both of which imply a sense of ownership, individual agency, and control. This language indicates that we have an emotional relationship with the data, and consider it to be something that belongs to us individually and only contains information relevant to us.

In reality, data about us is rarely just about us as individuals. It is almost always about ourselves and others; be it family, friends, colleagues, or people we happen to be in the same place as at any one time.

The data that we hold on our mobile phones, or that we share with any internet connected device – such as a computer, a voice-activated assistant such as an Amazon Echo or Google Home device, a smart thermostat, a connected car, a home surveillance system, even a smart TV – is likely to gather data about others too, from telephone numbers to emails, text messages to photos, documents, voices, images, behaviours and so on. Any device designed to learn from us also learns about the people we know, live with or interact with. It may then combine that data with data from people considered or defined to be 'like us' because they have the interests and behaviours that we are deemed to demonstrate.

Data about us also goes beyond what is collected and shared through connected technologies. DNA and genetic data is often seen as personal to us, but our genetics also contain elements of our family, including people we are distantly related to but may never have met.[2] This can be brilliant for helping to determine familial genetic diseases and for taking action to protect each other. Sharing it can help find medicines and cures for others with similar diseases, and sharing our DNA in relation to diseases can help to identify other people who may be prone to them but completely unsuspecting.[3]

When we are asked whether we would like to share data about us, we are often being asked to make a decision about data about others too.

How data about us can be used

Data alone can tell one story, but data combined can create deeper insights.

Data about us is therefore often collected and combined (by public

2. Erlich, Y., Shor, T., Pe'er, I., and Carmi, S. (2018) *Identity inference of genomic data using long-range familial searches*. Science, Vol. 362, Issue 6415, pp. 690-694. Available at: science.sciencemag.org/content/362/6415/690

3. Hunt, E. (2018) *Your father's not your father': when DNA tests reveal more than you bargained for.* The Guardian. [online]. Available at: www.theguardian.com/lifeandstyle/2018/sep/18/your-fathers-not-your-father-when-dna-tests-reveal-more-than-you-bargained-for

sector organisations and private companies) with data about others, along with non-personal data, to improve services. Sometimes we are aware that data about us is being collected, but sometimes it is far from clear.

Combining data can be done to make decisions that create positive and meaningful outcomes for communities, society, and individuals. But it can also lead to the creation of an intrusive network of unseen organisations using data to make inferences about us, which lead to assumptions that steer us into algorithmic filter or preference bubbles, price discrimination or even denial of services or products.

These inferences are now commonly made by data analysis through 'machine learning'. Machine learning is a set of algorithms that can be used to gather insights and make predictions about data.

A machine learning algorithm uses 'training data' (assumed to be representative of something, such as a history of what we have watched on Netflix or listened to on Spotify) to create a statistical model.[4] This model is used to make predictions about things, based on the training data. These models (and the predictions made by them) can vary greatly in how representative or accurate they are.

Data about us is often used to create insights about, or predict, our behaviour.

For example, when we go online, data may be collected about our behaviour: what websites we look at, how long we spend on them, what device we are using, what we browse, our purchase history and so on. This data is often used to train machines to predict which adverts we are most likely to respond to. The idea is that the more data about us is gathered, the easier it will be to build more accurate machine learning models of our behaviour, and create systems to predict our actions.

This process of things being inferred about us can have a range of effects on our personal agency and more broadly on society as a whole.

Inferences based on our behaviour can help algorithms to determine outcomes that may benefit us individually, for example by steering us towards a relevant piece of information, a news story that we may find interesting, or an advert for a product or service that we may value or may improve our lives. Similarly, the data gathered about the behaviour of groups of people can, when aggregated and stripped of personal identifiers, be used to help with planning of services, provide insights into medical opportunities, establish ways of challenging pollution, deprivation, and creating a better world.

However, just as inferences can create positive outcomes, they can also restrict, block, or prohibit access to alternative products, viewpoints, entertainment, or news stories.

Inferences or assumptions about who we are, based on general analysis of behaviour, habits and browsing history can impact us. For example, it can affect how search outcomes are shown to us, how products are advertised to us, how content such as videos are recommended, and even the news and political campaigns we are shown.

In 2016, the Guardian newspaper revealed how filter bubbles impacted

4. Yu, A. (2019) *How Netflix Uses AI, Data Science, and Machine Learning — From A Product Perspective*. Medium. [online]. Available at: becominghuman.ai/how-netflix-uses-ai-and-machine-learning-a087614630fe

the political content that US voters saw and shared.[5] The power of knowing what people are interested in, and only showing them content based on that interest, is one of the key elements of how online political messaging has become increasingly targeted to individuals, as opposed to geographic areas.[6]

Being targeted as individuals in this way is a process that we have not experienced before. Previously the concept of encouraging people to engage with an idea, campaign, or product was general – all audiences or consumers were shown the same content. The ability to now show us nuanced content – aligned with interests or views that we have been presumed to have – is new to most of us and one that is causing concern across society.

In Eli Pariser's 2011 book, 'The Filter Bubble', he explains that "you may think you are captain of your own destiny, but personalisation can lead you down a road to a kind of informational determinism in which what you've clicked on in the past determines what you see next".

Since this book was published, arguably filtering and preference bubbles have become a standard practice online. As stated in an academic paper from July 2019 for the American Marketing Association:[7] "automated recommendations are now ubiquitous in consumer domains", and that there is "a dangerous risk" that "consumers display overdependence on algorithmic recommendations in a manner that may both reduce their own welfare and propagate biases system-wide".

So how do we, as consumers and service users, feel about these sorts of processes happening? In order to find meaningful ways to consider and answer this, we need to understand what data about us is, how it is used, what it feeds into, how it can be shared, and how it can be protected.

It is for this reason we wanted to test people's awareness of data about them, and develop narratives that could engage people, and help them to decide and express how they felt about data, whether data about them should be something they could own, or what sorts of rights and responsibilities should exist around data.

Data ownership

There has recently been an increase in popular and political commentary around the notion of 'data ownership', in both the UK and international press.

Debates around whether we can have property rights over data have rumbled along for a few years, particularly in relation to the level of control that we can or should have over data about us, and how it can or can't be used.

5. Carrie Wong, J. , Levin, S., Solon, O. (2016) *Bursting the Facebook bubble: we asked voters on the left and right to swap feeds.* The Guardian. [online]. Available at: www.theguardian.com/us-news/2016/nov/16/facebook-bias-bubble-us-election-conservative-liberal-news-feed

6. Smith, A. (2018) *Public Attitudes Toward Computer Algorithms.* [online] Washington D.C. Pew Research Center. Available at: www.pewinternet.org/2018/11/16/public-attitudes-toward-computer-algorithms/

7. Banker, S. and Khetani, S. (2019) *Algorithm Overdependence: How the Use of Algorithmic Recommendation Systems Can Increase Risks to Consumer Well-Being*, Journal of Public Policy & Marketing. doi: 10.1177/0743915619858057.

More recently, the concept of ownership as a determiner of control has been mooted, in reaction to the Facebook/Cambridge Analytica scandal and the widespread misuse of data about us.[8]

Within the corridors of power in the UK the concept of data ownership was raised in relation to rights regarding data back in 2015 by the Liberal Democrats when they proposed a Digital Bill of Rights.[9] The Bill outlined the principle "that personal data belongs by default to the individual to whom it refers; that the individual citizen has a right to access all their own data, in an open digital format; and, where reasonable, individual citizens can decide who else has access to their data".

Then in 2018, prior to the passing of the Data Protection Act 2018 and the implementation of the General Data Protection Regulation in May 2018, the Labour Party briefly proposed an idea for a Digital Bill of Rights.

One of the rights proposed was of ownership. It stated as Article 6 of the Bill the idea that "every data subject has the right to own and control his or her personal data. Every data subject is entitled to proportionate share of income or other benefit from his or her personal data as part of the right to own".

During a Bill debate in the House of Commons, the then Shadow Minister Liam Byrne described the debate about who owns the copyright to data, or how new data could be created by joining data with someone else's data, as "vexed".[10] He also said that "the question of who owns the copyright, and therefore who owns the value of data that is personal in origin, is only going to grow". Very little meat was put on the bones of what was meant by ownership of data. With the Bill failing to go any further, the concept of ownership has taken somewhat of a back seat.

The notion of having property rights over data was also aired by the former Conservative Chancellor of the Exchequer, now editor of the London Evening Standard, George Osborne. In a speech he made in March 2019, Mr Osborne outlined clearly and concisely the way that data about us is used to bring profit to advertising companies, rather than to us as the data providers.[11] His solution was ownership: "say you had the right to take your accumulated data from one producer and share [it] with another that offers you something better in return [...], say social media companies had to pay you for using your data. Say it became an asset, or perhaps even a reward for your labour".

The idea of determining a value for personal data, or developing an approach based on the idea of personal copyright law, was raised by musician and entrepreneur will.i.am in an article in The Economist in 2019.[12]

8. Wikipedia. *Facebook–Cambridge Analytica data scandal* [online]. Available at: en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal

9. Archer, L. (2014) *'Liberal Democrats plan new Digital Bill of Rights'* [online]. Available at: www.libdems.org.uk/new_digital_bill_of_rights

10. Byrne, L. (2018) *Data Protection Bill [Lords]* Deb, 22 March 2018, c301) London: Hansard. Available at: www.theyworkforyou.com/pbc/2017-19/Data_Protection_Bill/07-0_2018-03-22b.297.1?s=%22bill+of+data+rights%22+section%3Auk#g307.2

11. Osborne, G. (2019) *The Politics of Newspapers* www.standard.co.uk/comment/comment/george-osbornes-hugh-cudlipp-lecture-the-politics-of-newspapers-a4085671.html

12. will.i.am. (2019) *We need to own our data as a human right—and be compensated for it* The Economist. [online]. Available at: www.economist.com/open-future/2019/01/21/we-need-to-own-our-data-as-a-human-right-and-be-compensated-for-it

In it, will.i.am suggests that "the ability for people to own and control their data should be considered a central human value. The data should be treated like property and people should be fairly compensated for it". He goes on to say how, as a musician, he benefits from the copyright system and that the same rules should apply to personal data.

While we are mainly interested in the UK and European approaches to data protection and rights, since the majority of big tech companies that we share data with are American, it is worth checking what is being said in the US, in relation to controls over personal data. will.i.am's views, or those that are similar, are also felt by others.

In the wake of Facebook/Cambridge Analytica, recent commentary has focused on the idea of ownership being a way of enhancing data rights. This has come from a range of people, such as Senator John Kennedy, Silicon Valley, and former Facebook investor and advisor Roger McNamee.

Senator Kennedy is one of a number of representatives in the US trying to address data protection concerns. In March 2019 he proposed a three-page Bill: the 'Own Your Own Data Act'.[13] Meanwhile, Senators Mark Warner and Josh Hawley are seeking to make social media companies inform people of the value of personal data collected about them.[14]

In his 2019 book, 'Zucked', Roger McNamee writes about attempts by Senate representatives to get a data privacy bill of rights through Congress that would "actually restore ownership and control to users". He admitted this would be hard but felt that the concept of "owning your data" was a declaration that would promote privacy and freedom. Further proposals in the bill covered the right to "opt in" to data being shared as opposed to being required to opt out, and a range of other rights similar to the EU's General Data Protection Regulation (GDPR) such as the right to obtain, correct or delete personal data; the right to be notified about breaches; and the right to data portability.

While these are just a flavour of the issues people are grappling with, they all acknowledge that the data-exchange we experience is unbalanced. The process of giving up data for free services has, in reality, become an unfair and inequitable value-exchange. But is data ownership desirable, or even possible?

Data rights and responsibilities

The meaning of the term 'rights' depends on its use and context. Broadly speaking, rights can be social, legal or ethical principles. To have a right to something means that we have permission to do it or are entitled to it, and that other people are responsible for enabling us to have it. This is called a 'positive right' because it requires action. We can also have 'negative rights', which require inaction. We might have a positive right to vote, but a negative right not to vote.[15]

Having corresponding responsibilities is something that all of the

13. Kennedy, J. (2019) *Own Your Data Act*. Washington D.C. Available at: www.govtrack.us/congress/bills/116/s806/text

14. Hart, K. (2019) *Scoop: Bipartisan senators want Big Tech to put a price on your data Axios* [online]. Available at: www.axios.com/mark-warner-josh-hawley-dashboard-tech-data-4ee575b4-1706-4d05-83ce-d62621e28ee1.html

15. Wikipedia. *Rights* [online]. Available at: en.wikipedia.org/wiki/Rights

different types of rights have in common. Human rights are specific, apply to everyone, and are defined and protected by law. It is the responsibility of governments to act in certain ways – or to refrain from certain acts – to promote and protect human rights and fundamental freedoms of their citizens.[16] For example, if we have the right to education, as we do under Article 2 of the Human Rights Act, then the State is responsible for providing us with access to education.[17]

Other sorts of rights are more like normative, social rules established by a group, community or society. These are not necessarily protected by law, but they reflect standards that have been agreed on by a group or society, and, in order to make sense as rights, they must have people with the responsibility to protect them.[18]

Like 'rights' in general, the terms 'digital rights' and 'data rights' have been used in many different ways, with many different and overlapping meanings. Many people talk of 'digital rights' or 'data rights' in relation to intellectual property, while others use them in reference to relevant human rights and legal rights that exist to protect people's freedoms to access and use digital media or data.

In this report, we refer to 'data rights' as the rights that we (should) have as individuals or groups around data. They might be rights to access data, withdraw data, or even benefit from, or not be harmed by, data's use or impacts.

The Open Data Institute's theory of change promotes ethical considerations to data collection, management and use, and equity around who accesses, uses and benefits from data. This relies on governments, businesses, civil society, and individuals themselves being responsible for ensuring ethics and equity, through their actions or inactions.[19] This relates to 'data rights' as normative, ethical and legal constructs.

In the UK and Europe, the use of personal data is controlled by data protection law. Since May 2018, the GDPR has been enacted by all EU member states, including the UK, which has enshrined GDPR into law in the Data Protection Act 2018.[20]

The regulation is based on data rights and responsibilities. People have eight data protection rights under the GDPR, they are:

1. The right to be informed – ie be told what is happening with data about you
2. The right to rectification – ie if the data is inaccurate, have the right to ensure it is amended, corrected and made accurate
3. The right to access – ie any organisation must be able to provide you with access to data about you, if you request it
4. The right to restrict processing – ie you have the right to ensure that

16. Un.org. (2019). *Human Rights*. [online] Available at: www.un.org/en/sections/issues-depth/human-rights

17. Equalityhumanrights.com. (2019). *Article 2 of the First Protocol: Right to education | Equality and Human Rights Commission*. [online] Available at: www.equalityhumanrights.com/en/human-rights-act/article-2-first-protocol-right-education

18. Wikipedia. *Rights* [online]. Available at: en.wikipedia.org/wiki/Rights

19. Open Data Institute (2019) *Our Theory of Change* [online] Available at: theodi.org/about-the-odi/our-vision-and-manifesto/our-theory-of-change/#1531394343108-b226e61c-833d

20. UK Parliament (2018) *Data Protection Act 2018* [online] Available at: www.legislation.gov.uk/ukpga/2018/12/contents/enacted

data can be held but not processed any further

5. The right to erasure (also known as the right to be forgotten) – ie the right to request that a company deletes data about you if you ask them to
6. The right to data portability – ie data about you must be provided to you securely, and in a readable and portable format, so you can share it with other organisations if you wish
7. The right to object – any organisation must respond to you if you raise a concern with them about how data about you is being used
8. Rights relating to automated decision making and profiling – you have the right to ask an organisation to restrict the use of data about you for profiling or automated decision making, and to have a human point of view and decision made. It is not always clear however when such activity is taking place.

Organisations using data have the responsibility of determining the lawful basis for processing data. This includes ensuring that they only hold the data they need, that it must be accurate and up to date, that it shouldn't be held for longer than is necessary – particularly if it can identify us – unless it is for scientific, public interest or historical research purposes. It must be secure, and anyone holding the data must be accountable and responsible for demonstrating compliance with the principles. In certain cases, organisations have rights to access personal data if they can demonstrate a legitimate, legal or public need for it.

Since its launch, the GDPR has been seen as a welcome and clearly outlined framework for data protection law. It is not perfect but as the next step of data protection in a data-driven world, it is seen to be a pretty solid foundation.

There are widespread calls for GDPR-type protections to be adopted internationally.[21] Already, the GDPR has been used as the framework by countries outside of Europe to build their own data protection laws around. Brazil, for example, signed off on a General Data Protection Law at the end of 2018,[22] which adopts the GDPR concepts of data subjects, data controllers, data processors, and it also develops standards for consent.[23] California has recently passed the Consumer Privacy Act 2018, which will become law in 2020.[24] The Act will offer a range of GDPR principles, such as the right to data deletion, transparency of how data is used, and the right to tell a business not to sell personal data to a third party including as an opt out. In India, the Personal Data Protection Bill has been proposed as a legal framework to establish limits on how data can be collected and processed with regards to necessity, proportionality, and fairness.[25]

21. Cook, T. (2019) *You Deserve Privacy Online. Here's How You Could Actually Get It* Time [online]. Available at: time.com/collection/davos-2019/5502591/tim-cook-data-privacy/

22. Brazil National Congress (2018) *the "Brazilian Internet Law"*. [online] Available at: www.pnm.adv.br/wp-content/uploads/2018/08/Brazilian-General-Data-Protection-Law.pdf

23. DLA Piper. *Data Protection Laws of the World*. [online]. Available at: www.dlapiperdataprotection.com/index.html?t=law&c=BR&c2=

24. SB 1121, Dodd. *California Consumer Privacy Act of 2018*. [online].leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1121

25. Srikrishna Committee (2018), The Personal Data Protection Bill, 2018. [online]. Available at: dataprotectionactindia.com/wp-content/uploads/2019/01/Personal-Data-Protection-Bill-India.pdf

What we did

Over the spring and early summer of 2019 we convened two focus groups and one participatory workshop in London. Our aim was to understand public feelings towards data and their reasoning behind it.

We wanted to understand how people responded to and engaged with the concepts of 'data ownership' and 'data rights', and find narratives that would help people to understand, decide, and express the relationship they have with data about them. We chose to use research methods that would allow us to capture qualitative data, holding conversations with small groups of people, listening to and learning about people's personal experiences and insights.

We worked with an independent research company to recruit people to take part in these sessions. The sessions were held at the weekend and participants were paid for their time in order to minimise barriers to participation.

To ensure a diversity of views, the participants were selected to represent a range of ages, ethnicities, abilities, and socioeconomic backgrounds. Participants were selected to ensure that the group had varied attitudes, based on their answers to the questions: 'Do you think the internet is a positive thing or not?' 'Where do you place yourself on the political spectrum left to right?'

The focus groups took place over a day in April 2019. There were two groups, each attended by 15 participants. In these, we tested people's understanding of data, of ownership and of rights, and whether data ownership or data rights appealed to them as concepts.

The workshop took place over a day in June 2019. It was attended by 13 participants. In the morning, we tested three stories about data about us being used in different ways – in particular, how engaging the stories were, whether they were informative or made the participants feel differently about it. In the afternoon they were asked to develop the stories in groups. They were then joined by four more participants who were tasked to be non-biased 'judges', and provide feedback on their stories – how clear, engaging, and informative they were.

One limitation of our research was its geographic focus on Greater London. It would be good to develop the work to explore what attitudes are felt about data about us in other regions of the UK and abroad.

The focus group

What we tested and learned

Ownership and control

We split the two focus groups: one to focus on exploring the concept of data ownership, and one to explore the concept of data rights.

During the data ownership focus group we asked participants what ownership meant to them as a concept, and then asked them to describe different types of data. We then paired 'ownership' and 'data' together to understand whether the idea of 'data ownership' resonated with our participants.

We conducted the same process with the second focus group, replacing the concept of 'ownership' with 'rights'.

The participants had broad takes on what data is. One person explained to us that:

> "Data is now literally anything you can put into a computer. So it can be it can be your words, it can be your image, it can be your opinion, numbers. Yeah, I think that's maybe what some people have forgotten that as soon as your image becomes data it can be transferable manipulated, whatever. Yeah, it can be stored".

Participants told us that ownership meant control, "freedom to choose" and "decision making rights". They suggested we can own not only physical goods, but also intangible things like opinions:

> "Yes. I own my vote, no one can take it away from me".

In some cases people said that ownership gives us the ability to exclude others, or to collect revenue from people if they use things that we own. Overall, we learned that ownership provided people with a sense of security and freedom of choice. Some people highlighted that ownership is a privilege that we don't all get to benefit from.

When it came to the concept of data ownership, one person expressed that data ownership seemed like a "myth". They said:

> "I don't think we'll be able to own our own data, we're way too far gone. It's a myth. They've got all of our data, and maybe only if you haven't been born yet. I think it depends how much data they have on you, I think it's very intrusive. And if I had my own way, I'd have all my data back".

They suggested that we could only own the things we have control over, and that once we share data we lose ownership over it:

> "You have a choice of how many physical opinions you share, your life is broken down into core data I guess, like date of birth, name, etc. Now I imagine most people only want to share the bare minimum, and I guess that way, the more you share, the more you lose ownership". "I think ownership is a very strong word to use in a very wooly conversation. I think we have all agreed we don't own our data. Do we even want to own our data?"

One person said that that they didn't believe that we even own virtual items that we buy online:

> "I think ownership is becoming a blurred term now. People have physical objects that they physically own, but my son buys purely virtual items, ie guns on his video game. But he has all these things that he thinks he owns, a catalogue of things he owns, but it's virtual, intangible object and he's probably just bought a licence".

This point was reiterated by another person, who said:

> "I think everyone else defines the terms on which you can own something [...] so I had a film with a streaming service, but they shut down the service, I no longer have access to that film. So I'm told I own my data, but realistically, I don't own it on my terms, I own it on someone else's terms".

Some people said they thought that, because of the way terms and conditions are written, ownership now belongs with companies not with individuals. For example, we were told:

> "I now know that whatever I post on Facebook or Instagram belongs to them and that includes my face [...] I think there is a clause written somewhere deep down where I've signed up to that".

Other people said they thought the words '**possess**' or '**access**' were better for describing the relationship that these commercial entities had with data about us.

There was an assumption held by some that the companies who hold data about us have a responsibility to ensure it is used correctly and held safely and securely.

Just as will.i.am raised copyright as a way of having control over how data about us is used, people in the focus groups explored the same idea. There was agreement that intellectual property and copyright was an interesting idea. But in the online world the issue of copyright is controversial.[26] The creative freedom for people to make new user generated online content based on content created by others (which may have their copyright attached to it) is a sensitive issue and gets to the heart of ownership and permission to use or even to sell data.

Ultimately, people felt strongly about wanting to control data about

26. Reynolds, M. (2019) *What is Article 13? The EU's divisive new copyright plan explained*. Wired [online]. Available at: www.wired.co.uk/article/what-is-article-13-article-11-european-directive-on-copyright-explained-meme-ban

them, but they did not consider 'data ownership' to be a realistic concept or one that inspired feelings of confidence or safety:

One person said:

> "We don't have the control we think we've got, because people are able to use it even if we don't give them permission".

Another said:

> "Once it's online you don't have control. You can strip all the data, but once you share something it can be screenshotted and passed around. Once you put something out there, you have no control what happens to that photo anymore".

Another said:

> "The only point we own data is when its within ourselves, and as soon as your reveal it, you don't own it".

Part of why people at the focus groups were sceptical of data ownership was the issue of how to value data about us. The value that data has, after all, depends on many things and is an ever-changing concept. What has certain value to one person or entity has a different or no value to another.

One person said that:

> "Every bit of data has a value, it just depends on how you segment it, who's looking at it, what people want to achieve what that data, whether it's to give a more personalised service, understanding your political or financial services. To a financial service my data is valuable, but the same data is not valuable to my friends".

It is hard to disagree with this statement. Who, what and how we could define the value of data are the initial questions to be asked. They were not ones we answered in this project but are being asked more broadly by economists, policymakers and businesses.

It was this level of complexity that led the members of the public we spoke with to decide that ownership over data about us is not something they found to be logical or desirable. However, they emphasised that they wanted data about them to be used responsibly, and some suggested they wanted to have more choice, control and agency in how they share data about themselves.

Rights and responsibilities

When we discussed the overall concept of rights (rather than specific rights and responsibilities) with the second focus group, people said they felt that having rights meant being safe, having freedom within reason, having control, and that rights were an ownership of some kind. Participants agreed that there were different kinds of rights – some legally protected and others just perceived, like the "right to be offended".

They raised many similar themes as the first focus group did around

ownership, but focusing on rights more quickly inspired the group to start talking about frameworks and responsibility.

Many people said they saw rights as generally positive things, and only negative if they impinge on other people's rights. They said that the rights of different people could come into conflict due to changing norms, and that new legislation is needed to protect those rights.

Everyone we spoke to in these focus groups, and in the later workshops (set out below) had heard of the GDPR and could outline the regulation's basic premise. For example, we were told:

> "Under data protection legislation it's all written down isn't it? What the responsibilities are, and who has it. Whether anybody actually complies with them is another matter, isn't it?"

How the GDPR was implemented was an issue that a number of people raised. One person summed it up by saying:

> "Some companies did a much better job than others, some said: 'look we've got that information and we're going to hang onto it unless you tell us otherwise', some people went on and on and on [...] eventually you could just be bamboozled with information so that you can't make a decision".

Some could tell us one or more of the rights that the GDPR has given us. For example, one person said they had undertaken a subject-access request:

> "I did a data-access request for what a company had on me, and I was really shocked at what they had on me. It was about me personally [...] 99% of it was banal, functional information, but there's 1% of it that was outrageous".

Another person told us that:

> "I'm a fan of GDPR. I've used it and had stuff moved".

Some people we spoke with said they had engaged with the GDPR as part of their work or business practice. Their views very much depended on the jobs they had. For example, we heard from one person that:

> "GDPR was a pain in the arse, but it did some good".

Another person said:

> "No one understood it for ages, but I think it can be good and it can be bad". Someone else said: "It negatively affects business because [business] is all about seamless service, where they know everything about you. Whereas now they have got to ask".

A conversation took place around how rights are drafted, legislated and enforced in order to ensure fairness. Some participants said that a broad spectrum of people needed to be involved in crafting them, and that it should be a democratic process. Others were keen to ensure that people who understand data are involved. One person said that:

> "With legislation going forward, there's going to be a lot of tightening up around data, and I hope that is done by people who understand data, rather than politicians who have been pushed into a certain office and have no understanding of it, genuinely, or the implications of it. I think irrelevant of political parties, laws are passed by people who don't understand the details in the first place".

New rights come with new responsibilities. The group felt that legislation needed to be enforceable, with one participant summing up that:

> "Rights have got to be enforceable. Otherwise they're not rights, they're just wishing something".

The group discussed the difficulties associated with trying to enforce rights on a global level. This is an interesting point in relation to whether the internet has borders, and whether an international approach to rights for data can be put in place. Determining an international treaty of data rights was considered by people in the focus groups to be ideal but too complex. One participant said:

> "data doesn't respect boundaries".

When asked what they thought of the term 'data rights', some people first said that it implied protection:

> "Part of data rights is to have your data protected. Like your bank information, your medical information, that kind of thing".

There was then a long discussion about protection, with many sharing a view that organisations asking for data about us had a responsibility to protect it, but that it was often unclear whether it happened in practice. One person said:

> "So, I don't know how it works. And they know I don't know how it works. We're trusting them. And when I say I agree at the bottom of their things I don't read it".

On terms and conditions, another person said that:

> "Nobody reads it. Apparently you gave Facebook permission to have your face, which I never knew".

The concept of protection led the group to discuss the concept of control – in particular how out of control they felt when it came to knowing or choosing how data about them was to be used. Some people felt

that the very notion of being asked to share data about them with services online meant losing control:

> "I think once it's out there the notion of having control at all. I think it's gone completely, really. And truly. I think that's the bottom line. I hate to sound pessimistic".

Another person described a sense of resignation about the control they have over data about them they have shared online. They said they felt that

> "trying to get something back off the internet is like trying to take the piss back out of the swimming pool".

Another person said:

> "I feel like we need to be educated. Like, when I sign up to a website, when I do my banking, when I do anything on the internet, and then they ask me 'Do you accept these terms and conditions? Give me your name', I have no idea where all that's going. I just feel like I can't do anything unless I give my data away".

The group generally agreed that they felt a loss of control and that they were not happy with the current situation.

One person said they would like to have a 'dial' that we could turn in order to choose how much data about us is shared with a device or a website:

> "I think we need a dial. One hundred percent let it all go you can know everything about me, or zero percent, you can know nothing about me. You don't even have to type it in, just give it literally a control".

This idea was leapt on by others in the group who saw it as an opportunity, with one person summing up the view of the table as being:

> "We decide what we want to share".

We also learned that people thought it would be useful to be given a moment to decide whether they really wanted to proceed. One person said that they liked being asked, 'Do you want to delete this?' but they would also like to be asked, 'Do you really want to post this?' before pressing send:

> "I do think we can make it very clear that we have to be careful and cautious. This 'are you sure' question that comes up when you're deleting something? It should come up when you're putting something in as well, shouldn't it?"

One person said that overall:

> "We've got to learn to be responsible online and not be so impulsive.

There's a lot of impulsiveness. And also corporations have got to learn to be responsible and open about what they're doing about it".

There were some strong opinions about companies collecting vast quantities of data about us, and how they were using it. There was discussion about the positives that came with more personalised services, but more of the discussion seemed to focus on people's concerns about personalised services with a distinction being made between having the choice to sign up for a service, and being shown unsolicited targeted advertising. As three people said, respectively:

"Profit. I mean, in the end we're being targeted".

"They say that if something's free to you, then you're the product".

"They're constantly coming to you, you're not going to them [...]Yes my privacy has been infringed. It's in my home, I'm on my phone or whatever. I haven't chosen to have these people talk to me".

Concerns about how responsible companies are with the data they hold were also raised, with one person saying:

"We're relying on the companies that we use to stay credible and use it with integrity. If that's the case, that's why we sign up, cause we're assuming they're going to be responsible. If we knew they wouldn't be, I bet none of us would say yeah".

As with the first focus group, participants cared about how data about them was being used. Overall they were more open to sharing data about themselves if they could do so anonymously and it was being used for public benefit. Some people suggested they would be happier to share data about them if it could be guaranteed that it wouldn't be shared with a third party. We were told that, most important of all, people wanted to have a choice in how data about them was shared and used.

One person said they didn't want to own data about them necessarily,

"but I do want someone to ensure that it's not misused, or only used in a positive way. I want it regulated".

What we developed after the focus groups

We are often led to believe, by the press, parliament and in published surveys, that the UK public's knowledge and understanding of data protection is low. However, the discussions had between people in our focus groups suggested otherwise.

While people showed a broad understanding of personal data, we found that they wanted to be able to better describe when and why they were comfortable, or uncomfortable, sharing data about them online.

One person said:

"We have to differentiate between what's personal, what's sensitive, and

> what's open. Because open it seems like none of us care [about] [...] but what's personal, and sensitive, that's the question mark, and how can we control it?"

Some said they felt greater education was needed. While the GDPR is seen to have been helpful in outlining personal and sensitive data, it has been, for some, less good at helping people to understand that data about them can have other uses, and can provide deeper nuances or insights into them as individuals, groups or communities, and society as a whole.

For this reason, we sought to find a way to clearly differentiate between the different types of data about us.

We developed a graphic setting out these different types of data, with examples, in order to help people see how it can have different elements. Some data about us is deeply personal, and can be used to create insights and inferences. Some data about us is personal but, if stripped back and aggregated with data about other people, less individual in nature, and helpful for wider decision making.

This tool was designed to help make people aware of what kind of data about themselves they are being asked to share, what data they have to share and why, and what data they would prefer not to share. We felt this would help people better understand what they are being asked to consent to and assist with improving education.

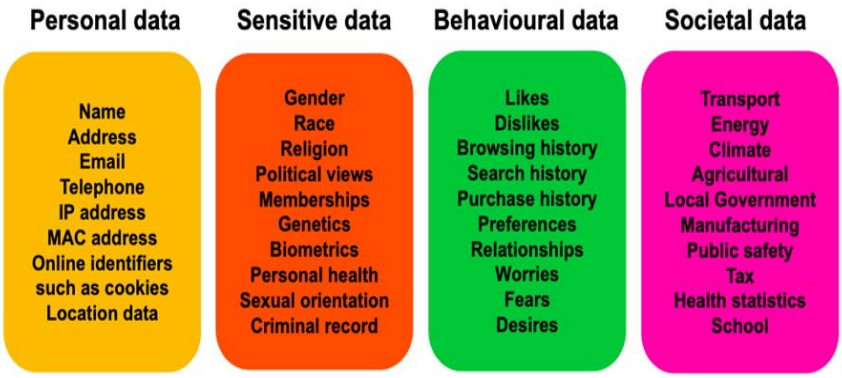
We tested this tool in the workshop, and we have iterated on it since for more clarity around societal data. Both these graphics are featured below.

Data about us: the four categories

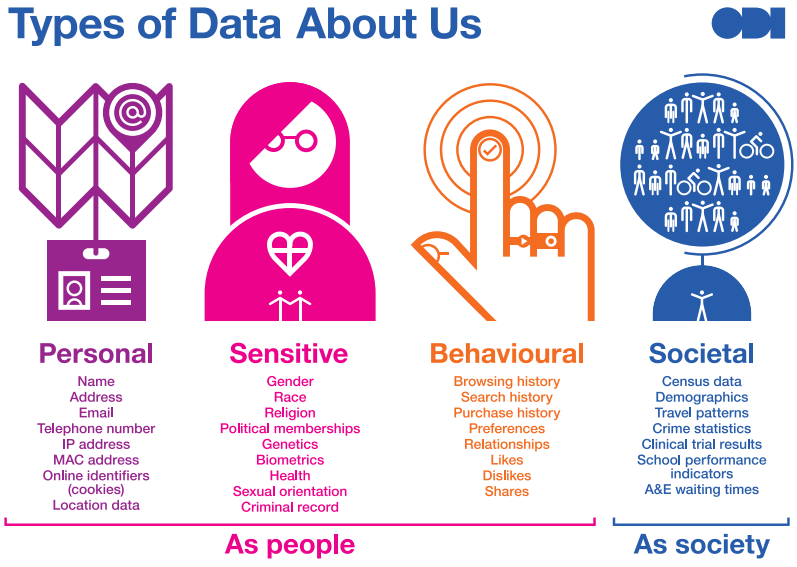
As we've explained, in this graphic tool we have sought to differentiate between and define the different types of personal data about us.

The first two boxes from the left, 'personal data' and 'sensitive data', are clearly defined in the GDPR – the second two to the right, 'behavioural data' and 'societal data', less so.

We presented our initial version (featured below) during a short presentation to the workshop groups.



We have subsequently iterated in this graphic to provide more clarity around societal data.



What are the different types of data about us, and how are they used?

Personal data as defined by the GDPR is any information relating to an identified or identifiable natural person. Some of it is data that we are familiar with sharing in order to be identified: our name, address, date of birth, telephone number, email address.

In a connected world there are increasing technical identifiers linked to our personal data: these are MAC addresses and IP (internet protocol) addresses. These can be used to identify the devices that individuals use, and hence the individuals themselves.

All of these identifiers link directly to individuals but may also link to others. For example, we may tell other people our name, we may share our birthday with people we don't know, our devices are shared amongst family members and our location is, depending on where we are, an identifier we share with everyone else at that location at that time.

The GDPR lists 'cookies' as an online identifier. Cookies are complex things. They come in a range of different flavours. Functional cookies, authentication cookies and unique identification cookies are cookies that act like a memory recalling where we have been online and what information we have given to a website. These make pages on the internet work, they leave a crumb enabling the website to remember that we have previously visited it. These cookies tend to be persistent – they embed themselves in our browsers. Sometimes they are used to mirror us between devices so that browsers remember us across a phone, laptop, or desktop. These cookies are focused on recognising it's us and remembering the information that we have given to the website before, such as a password, our bank details, etc.

Sensitive personal data is data that describes integral features of who we are: our ethnicity, gender, genome, biometrics (such as our fingerprint, DNA, facial biometric, voice and gait – all of which are completely unique to us as individuals), sexual orientation, and sex life. It also includes health data, educational data, employment history, criminal convictions, political party membership, amongst others.

Depending on the circumstances, making someone's sensitive personal data public, or misusing it in some way, could cause them serious harm. This data explicitly provides more detail about us, and deeper insights into us. It's data that, if misused, could lead to bias, discrimination or other harmful situations.

There are a range of laws and rights to protect us from such harms, such as the Human Rights Act and the Equality Act, which protect a range of human characteristics.[27] The GDPR similarly makes clear that, because this type of data is more focused on specific aspects of a person, special protection is needed to ensure that organisations wanting to ask for this data, use it, share it, store it etc have a specific need and clear reason as to why.

It is worth noting again that while this data is specific to individuals, it also describes people around them. For example, we share our genome with members of our family (even distant members or people we have

27. Equality and Human Rights Commission (2019). *Protected characteristics* [online]. Available at: www.equalityhumanrights.com/en/equality-act/protected-characteristics

never met). If we are asked to make a decision about how this data about us is used, we may find that we have to make a conscious decision about how its use might impact others.

Behavioural data is not specifically defined by the GDPR in the same way as personal or sensitive personal data. It includes data about an individual's behaviour and data that organisations use to infer, or guess, how we will behave.

Behavioural data tends to be collected via lots of different types of cookies, namely: performance, tracking, third-party, targeting, advertising, and social media cookies. The focus of these cookies is to track, monitor and analyse the behaviours we demonstrate when we are online: the what, when, why, and how of our online activity.

Behavioural data is also collected when we buy things with credit cards, store cards or, in certain cities, travelcards.

The adverts and content that we see online are often sold by AdTech and social media companies using cookies. These cookies are able to identify us, so the GDPR offers some protection to us in terms of personal data, but currently the companies that use these types of cookies are paying no attention to the law, a problem highlighted in a recent report from the Information Commissioner's Office (ICO) detailing the AdTech industry's use of cookies.[28]

Behavioural data is a deeper, more complex type of data about us. It is not just used to identify us, but also to understand us from watching and analysing our behaviours. This can be done both on- and offline.

Take location data. Our specific location when we stand in the street is an 'identifier' pinpointing us to a certain location, but if that identifier stays on and moves with us it creates a constant real-time stream of location data about us, giving an insight into our behaviour, not just a way of identifying where we are. Combine real-time location data with data about purchases we've made on our contactless bank card, while on that journey, and a more nuanced and detailed picture is created about us.

The same applies to a journey that we may take online. The website we visit is the location, but what we browse gives an insight into our personality – taste, mood, preferences etc. The insight of one website by itself is far from nuanced or accurate, which is why these cookies stick with us as we go from one website to another.

This 'behavioural data', as we've defined it, is data that historically has not been captured before in such a constant way. It is this data about us that seems to cause the most concern and anxiety for people. This is the data that people want to have more control over, which is why we have tried to define it more clearly, and what it can be made up of, so people can understand what is being gathered when they are presented with a choice of cookies, to accept or deny.

While cookies are covered in the GDPR, overall regulation for cookies comes from the Privacy and Electronic Communications Regulation (PECR).[29]

28. Information Commissioner's Office (2019). *Update report into adtech and real time bidding*. London: Information Commissioner's Office. Available at: ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf

29. Information Commissioner's Office (2019). *What are the Privacy and Electronic Communications Regulations?* [online]. Available at: ico.org.uk/for-organisations/guide-to-pecr/what-are-pecr

Societal data is data about us that incorporates elements of personal data (for example our location, our health, our energy use data) but does not need to include personal identifiers, that link back to us as individuals, in order to be useful.

Societal data forms part of important data infrastructure, which is crucial to understanding and meeting our society's needs. Examples of societal data are road and rail traffic, footfall in cities, disease statistics, and school attendance.

If handled correctly, societal data should be made to be as open as possible, with the data being open for anyone to access, use or share.[30] Open data can help governments, businesses and communities make decisions that benefit society as a whole, create products or tools such as maps or transport alerts, spot patterns in healthcare, or determine where local services are needed most. For these uses, societal data about us is often combined with non-personal data, like the names of towns, the location of bus stops, or the temperature at a particular time of day.

Even though societal data is often about us, in that it might include data about our commutes to work, in order to bring value it does not require anyone using it to know precisely who we are. Often this data is aggregated, focusing on total numbers of people rather than individuals.

The potential for us to be identified, or someone to learn more about us as individuals from aggregated data, will depend on how many questions are asked of the data, and how much the person looking already knows or can find out about us from other sources. There is with any dataset the risk that a person can be identified or re-identified, even if we are told data about us is anonymised.

Of course some societal data will include, or generate, behavioural data. Knowing how many people are getting the 5:15pm train out of Manchester Piccadilly station, for example, gives an insight into people's travel behaviour. Understanding how many people have contracted the flu in Plymouth in the month of March gives insights into people's health. An organisation might use this aggregated data to infer something about the behaviour of either specific individuals or groups of people.

The GDPR requires that societal data, as we have defined it, to be handled in clearly defined, specific ways. Just as with all data it must be handled lawfully, fairly and transparently. Anyone intending to process societal data should ensure they determine the appropriate lawful basis, be it consent, contract, legal obligation, vital interest, public task, or legitimate interest.

While societal data can bring opportunities, we must not be complacent and assume that all opportunities will benefit all of us fairly. A democratic debate about the use of data about us for societal decision making, and how the benefits are shared, is necessary. Just as with all data use, ensuring that the purposes of the use are legitimate, necessary, and proportionate is critical.

30. The Open Data Institute (2019). *What is open data and why should we care?* [online]. Available at: theodi.org/article/what-is-open-data-and-why-should-we-care

The workshop

What we tested and what we learned

Based on the conversations and findings from the two focus groups in April, we felt the next step should be to test narratives about data. We decided to do this within a workshop environment, so people could work together in groups to construct stories and test their knowledge.

Before the workshop, we write three stories about data to test. Each story was presented as a scripted conversation between two people. We included a small context-setting section before introducing a dialogue between two people talking about a real-life situation in which data about them was being used.

We asked participants to capture their initial reactions to each story in a feedback table before opening up to table conversations facilitated by a moderator and a note-taker.

We wanted to learn if talking about different uses of data make people:

- More aware of the various types of data about them which are used within society .
- Increase their interest in and demand for rights and responsibilities around data.

We wanted to learn what feelings different situations and examples of data use generated amongst the participants. Did they find the story appealing, reassuring, disgusting, or annoying? We also wanted to explore concepts of fairness and what people saw as a good or a bad use of data about them.

Within three table groups of four to five, participants were asked to engage with the stories one at a time, allowing notetakers to capture responses to each story. Table facilitators introduced each of the stories using the context-setting section, before inviting them to role-play in pairs with the scripted conversations. They were then asked to discuss what they thought or felt, and how they interpreted the stories.

After all the tables had fed back with their responses to the stories, we presented and discussed the different 'types of data about us' graphic [as explained in the previous section].

After lunch, we invited each table to work on and present a way of retelling one of the stories. This could be a poster, a play, or a presentation. We provided materials to help with generating something visually interesting.

They each presented their new stories to the group, along with new participants who were brought in as 'judges'. Not having heard the stories before these judges were intended to offer fresh, balanced judgement around the stories that most resonated with them, and increased their

demands for data rights and responsibilities.

Story 1

Story 1 was a made up conversation between two people about how data about us can be used by our employers to make decisions about us in the workplace. This was described by presenting the scenario of a woman not getting a job based on the interpretation of a photo of her that was posted by friends online, and a scenario of data collected from someone's 'wearable' worn at work being used to monitor their productivity.

The intention of the story was to test how people felt about the ways in which data about us can be interpreted, and to test concepts of fairness around automated decision-making within employment practices.

This story had the most visceral reaction amongst participants, with people expressing indignation, disgust, disbelief, discomfort, anger, a sense of their privacy being invaded, and that the situations were "unfair", "wrong", and "sinister".

The initial idea of a person being defined by a photograph of them online without consideration of context, or a conversation with them, was of concern to the participants. It led to a conversation about the responsibility we have towards friends and family when we post things about them online.

One person told us:

> "I've got one friend who says, 'don't put any pictures of me up on Facebook, Instagram [...] Don't put anything on it!'"

Another said:

> "It happened to my niece, someone put a photo on[line] and she didn't want it on[line], and it caused a lot of problems. And in that case, you should ask, you shouldn't just presume they want the photo to go up".

Generally people told us that they thought individuals should take initiative in ensuring their safety and privacy, and take responsibility for how and what data they were sharing online. But they also recognised that people need a basic level of education and awareness of how social media works, and how it can affect privacy.

> "I think people need to take a bit of responsibility for themselves in using those platforms, but it goes hand-in-hand with the company telling you honestly as well, because there's also a bit of that Cambridge Analytica and Facebook using your data and people not being aware of what it was being used for. I think it goes both ways [...] people have to take responsibility for using these platforms themselves, but I think the company has to tell you".

The group discussed the importance of boundaries between personal and work life, one person said they thought that monitoring via wearables is a misuse of data, and that interpretations or judgements based on data out of context are "wrong" and undermine privacy. There was a sense of

agreement that people shouldn't be penalised for behaviour within their personal lives that doesn't have an impact on their work:

> "Why should an individual in their own spare time be penalised for what they do outside their work? It's nothing to do with in the workplace".

The group suggested that without full context, assumptions can be made about individuals that aren't fair:

> "Computers are just brutal aren't they? Whereas I suppose if you had a boss who was a human, they might [think], 'Oh she's going to the loo, but I know her work, she's a really hard worker, she's not lazy".

Another felt that:

> "If you've got an algorithm looking at somebody, then the algorithm should be challenged […] it's not necessarily true".

Another person summed it up by saying that:

> "Data is not black and white. It's how it's read and it's how it's interpreted [...] We're not all robots [...] Everybody's got strengths and weaknesses".

The concept of data not being straightforward ("not black and white") and the need for human interpretation and consideration was raised a number of times across the workshop as a whole.

In the scenario, when the 'opt-out' option was presented as the decision to quit a job (to escape its work-place monitoring), the group agreed it was only a viable choice if other employment was available.

One said:

> "If it was every single company, [so] you had to give away your privacy to get a job at all, then that would just be completely... just so alarming. That's an abuse of human rights".

The same participant went on to suggest that more should be done to help people opt out of that kind of treatment, especially as some groups are more vulnerable than others:

> "What about the people who really don't understand it at all, [if] they would like to opt out but they have no data literacy, they don't understand the tech, they don't understand all of those terms and conditions [that are] hard to read".

Story 2

Story 2 was a conversation between two people describing how data about us (when anonymised, de-identified and aggregated) can be used for decision making for public services. The example given was wi-fi connections on people's personal connected devices being tracked in order to track and analyse the flow of people around a transport network. This story

was based in part on the tracking of wi-fi connections on the London Underground.

The intention of the story was to test how people feel about data about us being used for decisions that aren't about the individual, rather about benefits to wider society. We wanted to understand what feelings this raised, and hear what protections (if any) people wanted to see with the data.

This story was found overall to be the least contentious. One participant said they felt

> "more at ease because it's not as intrusive".

Another said:

> "Yeah, they can know that for society and stuff, but not to just get more money out of me and sell me stuff. I feel like for society for transport and energy, I don't mind them knowing".

While people were accepting and felt positive about data about them being used to make services better, they had some expectations about how it should be handled in order for them to feel completely comfortable.

They said the need for transparency is critical, and that they expect organisations or companies to be clear about what is being done with the data. There was a feeling that if we are not told what is being done then the purpose may be harmful or unfair to us.

> "They should tell you. They should be transparent. And if they're doing it covertly, what else are they doing with that information?"

How the 'benefit to society' would be defined was also an issue. In relation to the use of data about us in the transport story, one person asked:

> "Are they looking for where to make investment or looking for where to make cuts?" Another said: "If it [is] about knowing where people [are] to improve services that's one thing, but we shouldn't have to help them decide [how] they get most money by where people stand […] that's another thing, and I wouldn't agree to that".

Improvements and efficiencies need to be visible and transparent for public buy-in. One participant said that:

> "The thing I'm worried about is permissions". Another said: "I would be worried about vulnerable people or children who struggle to understand what's going on or what they're giving permission for. I like the idea of an independent body overseeing what's going on".

Participants then suggested that there should be clear tracking notifications or 'requests' that pop up on people's phones; the ability to remove data about them at a later point; an update after six months explaining what has been done with the data and any money generated if the data is sold for any purpose.

Storage and future use of the data was also raised as a potential concern. One person said:

> "It's fantastic to be planning and making services better and having more [trains] at particular times. All that is absolutely fantastic. But once they have this information, things might change. And I don't like the thought that data like this can be stored for years and years".

Story 3

Story 3 presented a conversation between a couple who are planning a holiday. The scenario presented them searching online for the same flights and hotels, one using their phone, the other a laptop. The story sought to explain how our behaviour, the devices we use, and the way that data about us is collected and analysed by online trackers such as cookies, can have a direct impact on the content, offers, and prices that we are shown as individuals.

The intention of the story was to test whether people were aware that companies collect behavioural data about us along with personal data about us, and that this can directly affect what we are shown online. We wanted to find out how that made people feel, and whether they found it fair or unfair, or were indifferent about it.

People's initial reactions to the scenario were that they were confused, surprised and curious.

> "I felt stupid, I didn't realise what the companies do", one person said. Some participants weren't aware that companies may personalise pricing or content based on people's profiles. One person said: "I'm astonished that companies [...] are allowed to do stuff like that – purely based on [...] someone's perceived wealth".

Some were confused as to exactly how the algorithms behind websites work to create this personalisation.

Several people related this story to analogies of price discrimination in the non-digital world targetting people perceived as vulnerable or ignorant, such as

> "Rogue traders [going] to an elderly person's house and [saying] 'Oh you're guttering's broken, it's going to cost £800,'" or the idea of a car mechanic charging a woman more because "He's thinking, 'oh, I can take advantage of her because women don't tend to know as much about cars as men'".

The group discussed the difference between usual price-discrimination through peak supply-and-demand times (such as summer holiday prices unfairly penalising parents) and this situation, which they considered a greater injustice. Price discrimination based on how far in advance someone booked a holiday wasn't seen as controversial, as that was considered a personal decision, but personalised tailoring of prices was considered to be unfair as the individual had no control over it:

"Why is it more expensive? You're obviously being watched on different websites".

In particular, they thought this type of activity targeted the vulnerable or ignorant. A couple of people were surprised that the type of device they use can have an influence:

"I wasn't aware that not only [do] they take into account what websites you've been on and what products you buy, but also the device you're using and everything. I didn't know that".

Adverts were contentious. One person said:

"I don't mind being recommended a different book, but I hate millions of adverts". Another said: "You can't tell companies not to target ads at you".

This led to a conversation "the right to choose", wanting to be able to give a shop or service the ability to advertise to you. One person expressed concern about her 12-year-old daughter being targeted by adverts, saying the advertisers had no right to do that. In response, another person said:

"We're starting from the assumption that companies can do this, but [...] none of us have given permission". Another person's suggestion was that "Sometimes [if I'm] sick of the sight of the same three t-shirts I've looked at or something, I'd rather have random adverts".

The idea of context being important – not just in terms of accuracy, but also fairness – was raised:

"It's not even accurate. They've got it wrong. Why can't I challenge this? Oh it's so frustrating, they're making guesses about me".

One person said:

"There's loads of people watching everything [I] do, and they're all making a guess about what I am like [...] I just like to browse it. Sometimes I'm not even browsing for me. It's for someone else". Another said: "They need a change in their business model. Right now, the more they know the more they can make money second guessing you".

Another told us:

"I feel like I've got more of an issue with this one because it's just quite inaccurate. So it could be offensive [...] it's unfair because it's inaccurate. So it's just bad use of data in this sense".

The concern about how we are "judged" based on data about us was extended to the physical world, with one person saying, in relation to getting a visa to move to the USA, that

"They will check your political affiliation or your [online] posts. It's just nuts to me".

As to rights around data, one participant said:

"I should have the right to decide whether my data is being taken and used". Others outlined that the rights that exist already – such as opting out – don't always seem to be working: "Not all cookies allow you to opt out of it. Not all websites are clear about cookies".

Feelings of resignation, that 'this is just the way it is', came up with some participants. One said:

"I actually think a lot of things go on around us that we're not even aware of".

Another said:

"I wouldn't know what to do or what to change about my behaviour online because there isn't a control for the individual at the moment".

Another said:

"There's no manual [for how to use the internet]. It's literally trial and error".

Reaction to the types of data about us

When we shared the four types data about us with participants, they all said the different categories ('personal data', 'sensitive data', 'behavioural data', and 'societal data') were very useful in helping them understand their relationship with data, and place themselves within the three stories we had talked about. One person said:

"I've understood it more in that one slide than I've probably [ever] read about [data] online, so that's very good".

Several people picked up on the blurred lines between 'sensitive data' and 'behavioural data', and how one could be inferred from the other. One person explained this by saying

"They can make a guess on maybe different cultural events that you [go] to, or holidays that you [celebrate]. If you were [to] Google Christmas presents, or Google something going on for Ramadan or something".

People said they wanted greater protections or a new social contract. One said:

"I think societal data is something separate, but the sort of blurring between those different categories, particularly sensitive and behavioural data, I mean, I get that one is protected under GDPR but like how... how

do you define and pull apart those things? Because you could infer someone's political views from things they like on Facebook, and that is what political parties and lobbyists do".

As a group, they agreed that societal data is useful and that they were happy for data about them to be used if it could be de-identified with a guarantee that there would be transparency of how the data would be used, what the benefits were, and how it would be stored. One person said:

"Yeah, they can know that for society and stuff, but not to just get more money out of me and sell me stuff. I feel like society for transport and energy, I don't mind them knowing".

Presentations and feedback from the panel

We wanted to test what participants had learned during the course of the day, whether the stories and the explanation of the different types of data about us had had an impact, and what they considered important to share with others. Therefore, in the last session of the workshop, we brought in a 'panel' of four new members of the public, to listen to the stories that the other participants had developed in tables, and give their reactions.

Table 1

Table 1 created a poster with presentation in relation to 'Story 3': inferences based on data about us as consumers.

The group had categorised factors about the story into a 'traffic light system' from green to red. Green represented convenience, amber represented data protection and red represented the tracking of people.

The panel felt the presentation was informative and liked the balance provided by the traffic light system, but felt it lacked emotion and context explaining why the group had chosen the factors they had.

When asked what feelings the story inspired and values it captured, the panel said it made them realise how unclear it is how data is used and whether it is done in a responsible way. One panellist said:

"It seems that we don't really have full control of how our data is used. We are manipulated in a subtle manner by corporations".

A few panelists thought that some of the examples in the story were "random" and the approach didn't tell a cohesive story. Instead they wanted to see more detail about who is responsible for governing data. One said in relation to whether it is companies or people who are responsible:

"Do we have responsibilities as well? Or maybe, because of the flexibility of online services, we've just handed the responsibility to the organisations?"

Our learning: the scenario of inferences being made about us based on data about us as consumers was found easily relatable by both the morning group and the afternoon panel. The scenario prompted discussion

about the level of comfort and control people have about how data about us is used by companies. It also prompted conversation about what responsibilities should be considered. The response of the panel revealed that, in terms of capturing people's imagination about how data about us is used, people wanted to see a story with characters and relationships rather than a presentation to help them engage more deeply

Table 2

Table 2 created a poster and gave a group-presentation in relation to 'Story 1': automated decision making in the workplace.

The group used a poster to outline a scenario where people were being monitored and automated decisions were being made in a fictional company. The poster and verbal presentation was designed to act as a call-to-action to employees to stand together and challenge the company's approach to the handling of data about staff.

The panel felt that this story as the most emotive. They suggested it felt as though the group were calling for a revolt. They liked the fact the group had used pictures and felt the poster approach engaging. Some of the feelings they said it raised were:

> "[It] makes me feel like information and data has been used in a deceitful way, and that there has been a betrayal of trust".

> "[It] reminds me that data capturing can be both positive and negative – if not used with the right context".

One panelist said:

> "It made me feel that there is almost a data battle between modern management and staff".

Whilst the story had been emotive, it was felt by the panel that the group hadn't resolved the issues raised, and felt it was far from clear what action should be taken, or where the issue of rights should fall. A couple of people commented that they wanted to see a more collaborative approach taken between employees and employers rather than the "us versus them" framing.

Our learning: all the panelists agreed the poster presented had been helpful, and that it represented a clear sense of the story. The emotion of the story was referred to with panelists repeating the groups' initial reactions of anger, their sense of deception and compromised personal agency, due to the ways that technology and automated decision-making were used by the employer in the scenarios presented.

Table 3

Table 3 wrote a short play featuring characters represented by each member of the group in relation to 'Story 2': use of data about us to help make societal decisions. The group incorporated the Types of Data About Us graphic that we had previously shown to them, as part of their performance.

The group drew on elements from all three of the stories that we had presented earlier. They modified the dialogue from 'Story 2' to incorporate some of their concerns about being tracked online, and somewhat humorously also referred back to 'Story 1', with some actors wearing fake wearable wristbands they had made.

The panel: As a whole, the panel agreed this was the strongest, most engaging, informative, and structured story of the three. The context and characters were felt to be engaging and they said that using Types of Data About Us graphic helped give them a greater level of understanding. On the graphic, they said:

> "[It gave the] best clarity so far, [about] what can and cannot be used. This should be more public".

> "I think the presentation slide was excellent".

Similar to the morning groups, panelists were supportive of the idea that data could be anonymised and used in ways to benefit society, although some also questioned whether the real reasons for collecting data would be beneficial to the public:

> "Companies are thinking about improving services, that's the reason why they seem to be collecting, but the concern is what else is the data used for".

A few people highlighted a need for people to have a choice to opt out. Another suggested we should have the right to delete all our activity at the end of an online interaction: "Every time you connect to it, it should come up 'do you want your data', you know, ask the question. And then you have the choice to say yes or no to it, and explain: 'look, the data may be used just for services and you may feel that's safer, but be informative about and let people know what you're doing'".

Our learning: compared to the other two presentations, panelists felt they had a deeper level of understanding of the types of data about them. They attributed this to the graphic the group had chosen to put in their presentation and the nuances of presenting both positive and negative uses of data.

Conclusion

Our data lives are complicated. Never before has data played such an integral and granular role in how we live.

We all take different approaches to our digital lives. Some people are cautious about how data about them is used. Others are willing to share access to data about them with everyone and anyone. But, as we've found in the course of our research, most people want to make a choice based on how they feel at a moment in time, and be able to change their minds when they feel differently.

This nuance is often ignored or misinterpreted, particularly in quantitative research that is focused on people's perceptions of data and the value-exchanges around it.

Everyone we spoke with in our focus groups and workshop made clear that they had accepted the internet and their connected lives. Everyone said it gave them positive and beneficial experiences, be that more choice, better connectivity, ways of keeping in touch, sharing, engaging, learning, and teaching. People liked to be able to shop online, and download music, films, and games. Some said how mobile technology helped their children not to get lost; others said how it benefits their work and their hobbies.

However, people also expressed discontent, worry, and feelings of resignation. They worried about how much they understand, how well they are educated in using connected technologies, how safe they are, and their lack of control over how data about them is used. They also worried about being subject to organisations – both private and public – making decisions about them and decisions about society, which they fear might not actually benefit them.

It was a commonly felt concern among the people we spoke with that we are being misrepresented, misread, misinterpreted, and misunderstood because of algorithms, automated decision making, and a simple lack of human engagement.

Participants didn't want to have assumptions made about them. They raised concern about the impact this can have on our mental health, our children's mental health, and our general wellbeing. They said that they didn't want to be judged by how far they deviate from a statistic of what is 'normal' in society, but be allowed to be the complex human, not a robot.

The overall sense of mistrust was strong and people expressed the feeling that power is sitting in the wrong place, used for financial gain and not sensitive to the impact it has on people.

Most interesting of all was the widespread engagement with data protection. Everyone had heard of the GDPR. Knowledge of the detail was varied, but it was legislation that people were familiar with and they spoke of it with some confidence.

We posed general questions and presented basic stories to people to

test their feelings, knowledge, understanding, and thoughts. What we heard in return was a desire for clarity, knowledge, choice, and control.

Suggestions from participants

The aim of the work was not to present a series of formal recommendations. We wanted the people to speak for themselves and to make their suggestions.

Some of the suggestions they made are things that already exist but clearly require greater signposting, strengthening, and improvement. Some are impractical but deserve consideration in terms of how the data world is seen by people and how government and business can approach transparency and engagement in the future. Some are simply demonstrations of how people want to be treated in their connected lives.

The following suggestions, which we have set out in bullet points under five categories for ease, are a selection of the collated suggestions made by the people we spoke to in our focus groups and workshop.

Honesty and transparency

We were told that:

- People want to have continued and improved transparency and information about what data about them is being used, how, when, and for what purpose by all organisations both public and private.
- People want honesty about how long data about them is being kept for and what it is being used for. In the focus groups we held it was felt by participants that if a company couldn't tell them how data was being used then it was up to no good.
- People want to be clearly told when tracking of them is taking place. Both on- and offline and to be asked to opt in rather than opt out.

Agency and control

We were told that:

- People want to see a more unified approach to how cookie consents are displayed and handled – they don't what to go hunting around a website and want better, clearer, understandable communication and explanation about what it is they are having to decide.
- People want to see wider use of opt in rather than opt out. Opting in was seen to give more agency over making decisions about the use of data about us.
- People want to be asked clearly for their permission to share data for societal purposes – not for it to be automatically assumed, and have it explained why data about them is needed.
- Some people want the opportunity to choose which adverts they

want to see, and from which shops and companies, as opposed to being served adverts based on assumption or inference from behavioural data.
- Parents want to have the opportunity to restrict and control what adverts are shown to their children.
- Some people want to be able to choose which services are personalised to them – rather than have organisations decide that for them based on their behaviour.
- People want to have continued access to education and more choice of services that strengthen privacy and security – examples given were encryption, virtual private networks (VPNs), and search engines that don't track people.
- Some people wanted to see more friction in place online. They want to be asked "do you want to share?", "do you want to upload this photo?", "do you want to tag this person?", "do you want to delete your data now?" because it would enable a pause for thought, and help them decide what they want. We were told that even if this was irritating it might help people engage with what was happening.
- People want to have clearer signposting, such as a pop-up alert, with clear instructions how to delete cookies at the end of a session on a device.

Rights and responsibility

We were told that:

- Some people feel that individuals need to take responsibility for communicating to others what they are comfortable with regarding posting of photos, comments, and other content about them on the internet. Others felt that there is also a place to ask before uploading or sharing data about others.
- People want to see companies take greater responsibility in their role of communicating what is happening with data about us online. This includes being clearer about how data about us is shared, sold, stored, and used to make decisions about us. This will enable us to control how we are tracked and how data about us is used to make decisions about us.
- People want to see government regulate companies to do things properly, some want this regulation to be "light touch", some want regulation to be a level playing field where all companies have to conform.
- People want independent oversight: such as commissioners, ombudsman or independent bodies to oversee the enforcement of many legislative or regulatory moves made to improve or protect how data about them is to be used.
- People want government draft legislation with people who understand data and the implications of it rather than politicians.
- People don't want regulation or legislation to be influenced by financial motive.

Context and fairness

We were told that:

- People want to see legislation and governance that would prevent prejudices and bias from being replicated and scaled through biased datasets.
- Some people felt strongly that news should be general and not personalised to them based on inference based on data about them – personalisation of news was considered unacceptable.
- People want to have clearer signposting and meaningful ways to prevent automated decisions from being made about them.
- People want the right to stop automated inferences from being made about them.
- People want an end to the idea that the value-exchange between data about them for deals, convenience, or nuanced recommendations is the preferred approach. People were clear that it's not a one rule fits all and that decisions are made about comfort levels of access to data about them in that moment. A decision on one day may not be the same decision the following day.
- People want to see stronger and clearer reference to how data about them will be anonymised and how they won't be easily linked/tracked/reidentified.
- People want to see companies who use data about them to contribute back into society by paying their taxes.

Compliance and enforceability

We were told that:

- People want companies to clearly tell them why an advert is being served to them, without the need to read lengthy explanations.
- People want to be given clear explanations of why data about them is being used, what the exact benefit is, and who has determined the benefit.
- People want to have improved communication from a company that they have been fully unsubscribed from a mailing list – people raised concern that they still received rogue emails or have to undertake lengthy unsubscribe processes.
- People want to have improved communication that data has been deleted.
- People want an end to data about them being sold to third-parties.

What next?

Society is still finding its feet in this relatively new data-driven and connected world.

Governments and businesses make decisions as to how the world should look all the time. Sometimes these decisions provide immediate benefit and little risk, sometimes however they are based on the principle

of ‘move fast and break things’. Both approaches have their benefits, but as we become more familiar with the role data plays in every facet of our lives, both online and offline, moving fast and breaking things is becoming far less tolerated legally and societally. How people respond, what they will tolerate, and what they will simply oppose, are likely to become more nuanced as time goes on.

It has been a privilege for us to sit down and speak to people in the UK about their feelings about their data lives, and the impact that the use of data about them have on them individually, as part of a larger group and for society as a whole.

What we were told was eye-opening. It made clear that people had much more awareness and understanding than ‘the UK public’ had been given credit for, largely by politicians and press. It also showed us that people were keen to express their wants, needs, and ideas for how data about them should be protected.

People are not naive or ignorant about data. We all understand – to a greater or lesser degree – its impact, role, and importance. Give people the chance to talk and they will engage in ways that will bring meaningful insight into the development of future rights, responsibilities, regulations, policies, and products.

This body of work should be the start of a wider conversation between people, governments, and businesses, as well as the more commonly held conversations between NGOs, interest groups, and think-tanks.

We’d like to hear what you think about data about us, about the rights we have, about the responsibilities we should have, that governments should have, and that businesses should have to strengthen our data rights.

Tweet your views using #WeAreNotRobots.

The RSA (Royal Society for the encouragement of Arts, Manufactures and Commerce) believes in a world where everyone is able to participate in creating a better future. Through our ideas, research and a 30,000 strong Fellowship we are a global community of proactive problem solvers, sharing powerful ideas, carrying out cutting-edge research and building networks and opportunities for people to collaborate, influence and demonstrate practical solutions to realise change.

The RSA has been at the forefront of social change for over 260 years. Today our work focuses on supporting innovation in three major areas; creative learning and development, public services and communities and economy, enterprise and manufacturing.

Central to the RSA's current work are the concepts of 'convening' and 'change making'. The RSA has also developed a distinctive approach to change: 'Think like a system, act like an entrepreneur' which now runs through our projects.

The RSA: uniting people and ideas to resolve the challenges of our time.

RSA
Action and Research Centre

8 John Adam Street
London WC2N 6EZ
+44 (0)20 7930 5115

www.thersa.org

ISBN 978-0-901469-78-6