Data governance: public engagement review

June 2017

THE
**ROYAL
SOCIETY**

BRITISH
ACADEMY
for the humanities and social sciences

Contents

Introduction

This review was undertaken as part of joint work between the British Academy and the Royal Society on data governance. The final report *Data management and use: Governance in the 21st century* was published in June 2017.

This review does not represent the views of either of the Academies.

The accelerating exchange and use of data is impacting everyday lives, activities and communities in new and unexpected ways. Recognising the new governance challenges posed by a changing data environment, the Royal Society and British Academy initiated a review of data governance, seeking to characterise and illustrate some of the changes that expanding data use has brought about, the tensions arising from these changes, and the ways in which a principle-based approach to data governance can provide direction and stewardship during a potentially disruptive period of transition.

To be effective, the governance of data and its use needs to be grounded in engagement. Such engagement needs to include activities which consider the context-specific nature of data use, seek thoroughly considered and representative viewpoints, and engage deeply with the complex social and technical issues that sit at the heart of these challenges.

Substantive public engagement can contribute to better decision-making and create more socially robust scientific and technological solutions[1]. Technological developments and dialogue need to happen in parallel.

As a starting point, this paper summarises key findings from an initial literature review of past public opinion surveys and qualitative workshops on the theme of the collection, sharing and use of data, and its governance. It identifies common themes and reveals aspects which have not been substantially addressed in previous studies, and which may help inform future public engagement.

This review was part of the evidence gathering process for the two Academies' data governance work, and helped identify key areas for attention. On this basis, the final report *Data management and use: Governance in the 21st century* looks at the overarching reasons for concern and what specific tensions arise in the current data governance landscape. The report considers the interconnected nature of data processes, the social and ethical opportunities and challenges that arise, and sets out actions needed to establish a governance framework that is fit for the 21st century.

1   Wilsdon J and Willis R. 2004 See-through Science: Why public engagement needs to move upstream.
    (see https://www.demos.co.uk/files/Seethroughsciencefinal.pdf, accessed 12 June 2017)

Summary of current findings and future needs for public engagement

Over the period 2009-17, a number of surveys and public dialogue activities have looked at the views of society on the collection, sharing and use of personal data for different purposes. This review identifies trends as well as gaps.

This review finds:

- Individuals tend to underestimate the volume and rate at which they generate data, and have limited knowledge about the data that organisations hold about them. Many report a sentiment of loss of control over data about them and its use.

- People who have a greater awareness about data collection, sharing and use tend to adapt their behaviours and protect themselves.

- The awareness of new uses of data, such as machine learning, is low.

- The language and framing of questions matter and can influence the outcome of public dialogue. For example, the phrase 'personal data[2]', used in a number of studies, may be confusing, because people often do not have a clear understanding of what data is technically considered 'personal data'.

- There is a discrepancy between people's concerns and their behaviours. For example, many use store loyalty cards despite concerns about the use of the data collected by retailers.

- People's views on data collection, sharing and use depend on purpose and context. For example, people take into account what the benefit(s) would be and who would receive those benefits.

- People want uses of data, either by public or private organisations, to benefit them, personally or collectively.

- Health and medical data are often seen as more sensitive and confidential, but they are also considered differently because of the perceived immediate benefit of their use to advance research and care. In contrast, the benefits of using administrative data to advance social research was not as immediately obvious. In addition, there seems to be an opposition between a relatively un-engaged attitude to data use by commercial organisations and a more complex set of responses to data when it comes to the use of that data by public sector organisations like the health service.

- Attitudes depend on previous exposure and knowledge. If people knew about a specific data use, they had a better appreciation of benefits and risks involved.

- Actions contributing to trust include being transparent by having continuous communication and providing evidence of the secure storage and protection of personal data. Clarity about what the data will be used for and how it will be shared are also essential. There is evidence that the awareness of regulation and safeguards can increase trust. For example, the existence of the Freedom of Information Act (FOIA) had effectively increased confidence and trust in public organisations.

- People expressed a desire for an impartial and independent oversight of new uses of data – it would ensure that the technology was not being abused and would guard against it being portrayed as accurate if it was not.

---

2   The Understanding Patient Data initiative has been investigating the language that can support better conversations about health data; Understanding Patient Data 2017: What are the best words to use when talking about data? (see https://understandingpatientdata.org.uk/what-are-best-words-use-when-talking-about-data, accessed 12 June 2017)

The review identifies the following gaps:

- There is a need for more engagement and communication about data and its use.

- As methodologies for the collection and analysis of data, and the technology context in which they are applied, evolve rapidly, there is a need to revisit certain questions over time. Only very few studies investigated attitudes to new and future uses of data. In addition, while some studies have explored potential near-term applications of data technologies, none so far have looked into future worlds enabled by data.

- Studies also highlighted that some groups within society could find it difficult to assess the benefits and risks of data uses. While several studies have looked into what criteria people use to define what is considered a valuable and beneficial output of data, they have not looked in depth at the social and ethical values at stake nor at the tensions between public good and personal risk. Similarly, our understanding of what concerns people most (for example security versus discrimination) is limited.

- Some studies also pointed to the fact that different generations have distinct relationships to data and its applications. It would also be useful to gain a better understanding of such generational differences, and in particular how they might derive from differences in experience and engagement rather than age specifically. Millennials also tended to be less aware of their rights regarding data held about them, and it would be interesting to explore how the level of awareness might affect their behaviour. The attitudes to data of 'digital natives' have not been explored in depth so far, though young people have been included in broader groups in many studies. Building on the major dialogue exercise it conducted in 2016, the Royal Society has carried out public dialogues on the views of digital natives on machine learning and its applications.

Findings from past public engagement studies on data and its use

We have reviewed a number of public engagement studies on data and its use to identify trends in public opinion to data over the period 2009-17, as well as any gaps. Some of these studies were analysed in a 2014 *Big Data* review by Sciencewise[3].

The Sciencewise review showed that people wanted to have more control over the use of their data, and wanted stronger safeguards as well as more information about how organisations collect, share and use data about them. The review also reported a discrepancy between people's concerns and their behaviour. Our review confirms these points. In addition, the Sciencewise report found that, while personal benefit was the strongest incentive for people to agree with data collection and use, ahead of public goods, the public saw little benefit themselves and had little hope to benefit from data use in the future.

Finally, the 2014 review stressed the need to gather more evidence around how public views change over time, what public views are on specific data technologies and what factors affect how the public makes trade-offs. We have included in our review several studies which have subsequently brought some clarity to these points.

The sections below summarise findings for a number of recurrent themes in the 2009-17 studies reviewed: awareness, purpose of data use, trust, new applications of data, ethical and social challenges, regulation and governance, and segmentation in public attitudes and opinions.

A. Awareness

A number of studies have investigated how aware the general public is of current data collection, storage, security, regulatory frameworks and how long the data is stored for.

The language used in public dialogue about data influences the answers and discussions, as found by the study *Understanding Patient Data*[4]. For instance, 'anonymous' and 'anonymised' were among phrases that were the most confusing – this is important because one of the biggest concerns people have about the use of data is whether the information could be traced back to them personally. 'Personal data' can also be a confusing phrase, because people think they understand the term, but may not have a clear grasp of its technical definition. This may explain a finding from a previous Digital Catapult study where participants were asked if they could define 'personal data' – while 96% claimed that they could, there was no agreement on the definition, with 64% defining it as 'all information about me in existence[5,6]'. Similarly, a 2012 Demos study found that the public did not have a clear understanding of how personal data or information was defined[7].

3   Sciencewise 2014 Big Data: Public views on the collection, sharing and use of personal data by government and companies. (see http://www.sciencewise-erc.org.uk/cms/assets/Uploads/SocialIntelligenceBigData.pdf, accessed 12 June 2017)

4   Op. cit. 2

5   Digital Catapult 2015 Trust in personal data: a UK review. (see http://www.digitalcatapultcentre.org.uk/wp-content/uploads/2015/07/Trust-in-Personal-Data-A-UK-Review.pdf, accessed 12 June 2017)

6   The other options of the multiple choice question were 'all online data', 'all data collected about me by organisations' and 'all data shared by me with organisations'.

7   Demos 2012 The Data Dialogue. (see https://www.demos.co.uk/files/The_Data_Dialogue.pdf?1347544233, accessed 12 June 2017)

Awareness of data collection, storage and access

The public engagement studies we reviewed revealed limited awareness amongst participants of how different organisations collect and use data about individuals. People often struggled to estimate the volume and rate at which they generate data, and had limited knowledge about the data that organisations hold about them.

A recurrent theme in a number of public engagement studies was the belief that too much data was being collected by various organisations on individuals. One study[8] found that 76% of participants believed that too much data was collected about them. This sentiment was reflected in the ESRC *Big Data* report[9] where participants felt they were no longer in control or could keep track of their own data. When participants of a Digital Catapult study[10] were asked if they wanted greater control over their own data, 94% said yes.

Studies have shown that people know about active data collection methods[11] and are generally less aware of passive data collection methods[12,13]. For example, in a Wellcome Trust study, few participants were aware of social network sites monitoring and analysing information from personal posts[14]. An EU Barometer survey on *e-Privacy* found that fewer than half of UK respondents knew it is false that instant messaging and online voice conversations are confidential and cannot be accessed without permission[15]. An Ipsos MORI *Global Trends* survey found that 83% of the UK respondents were unsure what information companies had on them[16].

Some studies have revealed that the more aware consumers are, the more they tend to adapt their behaviours. People who have a greater awareness that information about them is being collected are more likely to protect themselves, according to a 2012 Deloitte and Ipsos MORI study[17]. Similarly, the EU Barometer *e-Privacy* study found that 6 in 10 UK respondents had changed the privacy settings on their internet browser (e.g. to delete browsing history or delete cookies).

Beyond data collection, studies have shown low awareness of data storage and access. The participants of an ESRC *Big Data* report[18] knew little about digital storage practices and believed that physical storage is safer than digital storage. The same *Big Data* study also found participants knew little about how data about them is accessed by organisations.

---

8   DATA-PSST and DCSS 2015 Public Feeling on Privacy, Security and Surveillance (see https://sites.cardiff.ac.uk/dcssproject/files/2015/11/Public-Feeling-on-Privacy-Security-Surveillance-DATAPSST-DCSS-Nov2015.pdf, accessed 12 June 2017)

9   Hopkins Van Mil: Creating Connections Ltd (research sponsored by ESRC) 2014 Big Data: Public views on the collection, sharing and use of personal data by government and companies. (see http://www.esrc.ac.uk/files/public-engagement/public-dialogues/public-dialogues-on-the-re-use-of-private-sector-data-for-social-research-report/, accessed 12 June 2017)

10   Op. cit. 5

11   Active methods of data collection mentioned: filling out forms.

12   Passive methods of data collection included: data from cookies, travel and purchasing patterns.

13   Ipsos MORI (research sponsored by the Economic and Social Research Council, and Office for National Statistics) 2014 Dialogue on data: Exploring the public's views on using linked administrative data for research purposes (see http://www.esrc.ac.uk/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/, accessed 12 June 2017)

14   Ipsos MORI (research sponsored by Wellcome Trust) 2016 The One-Way Mirror: Public attitudes to commercial access to health data. (see https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf, accessed 12 June 2017)

15   European Commission 2016 Flash EU Barometer 443: e-Privacy. (see http://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/Survey/getSurveyDetail/search/e-privacy/surveyKy/2124 , accessed 12 June 2017)

16   Ipsos MORI 2016 Global Trends. (see https://www.ipsosglobaltrends.com/awareness-of-personal-information-held-by-companies/, accessed 12 June 2017)

17   Ipsos MORI (research sponsored by Deloitte) 2012 Data Nation 2012: Our lives in data. (see https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/data-nation-2012-our-lives-in-data.pdf, accessed 12 June 2017)

18   Op. cit. 9

Awareness of data use

Studies have shown a relatively low awareness of the uses of data, including health data. Participants of the Wellcome Trust and Ipsos MORI 2016 *One-Way Mirror* public dialogue[19] had little knowledge of how health care data and medical data were being used by the NHS, commercial organisations and academia (33% were aware of use of data by NHS; 16% by commercial organisations; and 18% by academia). This was reflected in a 2010 Royal Academy of Engineering study[20] that found a low awareness among younger people of use of electronic patient records in medical research. A common concern for participants of research studies is that data about them would be used at an individual level and could be traced back to them; most participants were much more comfortable when they were told the data was anonymised and aggregated[21,22].

Some studies have shown that members of the public were aware of data use by commercial organisations. For example, this includes the use of personal data for marketing purposes by organisations to better target their consumers. Participants also expected benefits in return for this collection and use[23,24].

When asked about data sharing, participants in a 2016 Government Data Science (GDS) Partnership and Ipsos MORI study[25] were confident in what data sharing was, and were surprised by how much data was shared in government. In contrast, participants of a 2014 Ipsos MORI *Dialogue on Data*[26] expected that data would be more linked across government departments than it was – most found it reassuring when they realised that there were multiple barriers to sharing data.

The GDS Ipsos MORI study showed there was little awareness of data science, from knowledge about current research and innovation to the value of data science[27]. At dialogue sessions run by Ipsos MORI on behalf of the Royal Society[28], the vast majority of participants knew very little about machine learning itself even though they had in fact already come across programs or applications that use machine learning.

19   Op. cit. 14

20   The Royal Academy of Engineering 2010 Privacy and prejudice: Young people's views on the development and use of Electronic Patient Records. (see http://www.raeng.org.uk/publications/reports/privacy-and-prejudice-views, accessed 12 June 2017)

21   Op. cit. 13

22   Ipsos MORI (research sponsored by Government Data Science Partnership, and Sciencewise) 2016 Public dialogue on the ethics of data science in government. (see http://www.sciencewise-erc.org.uk/cms/assets/Uploads/data-science-ethics-in-government.pdf, accessed 12 June 2017)

23   SAS 2015 Finding the Right Balance Between Personalization and Privacy. (see https://www.sas.com/content/dam/SAS/en_us/doc/research1/balance-between-personalization-privacy-107399.pdf, accessed 12 June 2017)

24   Op. cit. 9

25   Op. cit. 22

26   Op. cit. 13

27   Op. cit. 22

28   Ipsos MORI (research sponsored by the Royal Society) 2017 Public views of Machine Learning. (see https://royalsociety.org/~/media/policy/projects/machine-learning/publications/public-views-of-machine-learning-ipsos-mori.pdf, accessed 12 June 2017)

Awareness of data regulation and safeguards

Studies generally showed there was a low awareness of data regulations and safeguards. Participants in a Wellcome Trust and Ipsos MORI study[29] asked for more storage and access safeguards such as sanctions. However, they admitted that they knew little about how such safeguards would actually work, or how data was stored or accessed, but the idea that regulations were in place felt comforting to them.

Uncertainty was also recorded regarding the public's rights over how their personal data is handled by companies (71% of UK respondents were unsure)[30]. According to surveys sponsored by the Information Commissioner's Office (ICO), spontaneous awareness of the Data Protection Act (DPA)[31] was at its highest in 2007 at 45%, dropping in subsequent years before it rose again to 38% in 2011. However, prompted awareness was high (97% in 2016) – in comparison, only 16% reported they had heard about the EU General Data Protection Regulation (GDPR) when prompted[32].

Like the DPA, spontaneous awareness of the Freedom of Information Act (FOIA)[33] was generally low. In 2012-2013, it was at 25%, whereas prompted awareness was at 86% and participants had an overall high awareness of the level of rights under the act.

B. Purpose of data use

Public attitudes towards the use, sharing and linking of personal datasets were context-dependent. Generally, many studies concluded that if there was a clear personal, local or societal benefit, then individuals were more inclined to consent to it[34,35,36]. For example, the Royal Statistical Society and Ipsos MORI 2014 survey had a question about to what extent people agreed with government sharing data about them with different organisations: 50% supported sharing with universities and similar research organisations, whereas only 27% agreed with sharing data with companies to help them improve their products or services[37].

In addition, studies have shown that people are generally much more comfortable when data is anonymised and aggregated, and cannot be traced back to them or used to target them[38,39].

Data use by organisations can be broadly split into two categories: use of data for research and use of data for non-research applications. In this section we review both these uses of data and their public acceptability.

29 Op. cit. 14

30 Ipsos MORI 2016 Global Trends. (see https://www.ipsosglobaltrends.com/awareness-of-rights-over-companies-use-of-personal-information/, accessed 12 June 2017)

31 Opinion Leader (research sponsored by the Information Commissioner's Office) 2013 Annual Track. (see https://ico.org.uk/media/about-the-ico/documents/1042195/annual-track-2012-individuals.pdf, accessed 12 June 2017)

32 CitizenMe (research sponsored by the Information Commissioner's Office) 2016 Annual Track. (see https://ico.org.uk/about-the-ico/our-information/research-and-reports/information-rights-research/, accessed 12 June 2017)

33 Op. cit. 31

34 Ipsos Mori (research sponsored by the Royal Statistical Society) 2014 Public attitudes to the use and sharing of their data. (see https://www.statslife.org.uk/news/1672-new-rss-research-finds-data-trust-deficit-with-lessons-for-policymakers, accessed 12 June 2017)

35 Aitken M, de St Jorre J, Pagliari C, Jepson R and Cunningham-Burley S. 2016 Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Medical Ethics 17. (DOI: 10.1186/s12910-016-0153-x)

36 Citizens' Juries c.i.c. (research sponsored by Connected Health Cities) 2017. Connected Health Cities Citizens' Juries Report: A report of two citizens' juries designed to explore whether the planned and potential uses of health data by Connected Health Cities are acceptable to the public. (see https://www.connectedhealthcities.org/get-involved/citizens-juries/, accessed 12 June 2017)

37 Op. cit. 34

38 CM Insight (research sponsored by the Wellcome Trust) 2013 Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data. (see https://wellcome.ac.uk/sites/default/files/wtp053205_0.pdf, accessed 12 June 2017)

39 Wellcome Trust 2017 Understanding Patient Data: What are the best words to use when talking about data? (see https://understandingpatientdata.org.uk/what-are-best-words-use-when-talking-about-data, accessed 24 May 2017)

Research

Research is carried out by various sectors spanning academia, private organisations and public bodies. Public dialogues have looked at different research themes, and use of data sharing and linking, to understand their acceptability with the public. Researchers behind these studies are keen to use this information to improve their work, increase participation, and assess the acceptability of their research.

Studies showed that greater acceptance and participation in research studies can be gained if there is an output which leads to perceived benefits at the individual, local, regional or national level.

A Wellcome Trust 2013 (pre care.data) study found that health and care data are perceived somewhat differently to other data, because there is thought to be an unquestionable benefit to people in terms of experts having information about their health, in relation to treating illness or avoiding it[40].

Compared to medical research, social research and its outputs are less well known[41]. When social research was explained with examples of the types of research that have been conducted, the *Dialogue on Data* participants responded more positively, particularly when they could see a clear social benefit, such as improving national security and future planning. A similar response was also seen in a different study looking at socio-economic research[42]. After experts explained how they used data for social research, participants felt more reassured of the safety of the data held about them, due to the demonstrable expertise of the researchers. Participants suggested that researchers should be vetted before carrying out the work, with penalties if any regulations were broken. Participants also felt a risk that if social research was carried out by commercial organisations only, the potential benefits would be unlikely to be felt by society.

A Scottish *Cross-Sectoral Data Linkage* dialogue found that people were supportive of cross-sectoral data linkages if they could see benefits for the community[43].

The public tend to be more accepting and consent to use of medical records and health data in an anonymised form due to the benefits of research spanning the individual, community, society and national scale. A 2014 *Public Attitudes to Science* survey found that 61% of adults did not mind data about them being used as long as it was anonymised[44]. In a Wellcome Trust track survey[45,46], 77% of participants in 2016 said they were willing to share anonymised medical records for research. When asked about the use of personal data, participants in the Ipsos MORI *Dialogue on Data* stated that no identifiable information should be used (name, National Insurance number) and other private information such as HIV status should be removed.

---

40 Op. cit. 38

41 Op. cit. 13

42 Op. cit. 9

43 Scottish Government Social Research 2012 Public Acceptability of Cross-Sectoral Data Linkage: Deliberative Research Findings. (see http://www.gov.scot/Resource/0040/00400976.pdf, accessed 12 June 2017)

44 Ipsos MORI (research conducted in partnership with the British Science Association, sponsored by the Department for Business, Innovation and Skills) 2014 Public Attitudes to Science. (see https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/348830/bis-14-p111-public-attitudes-to-science-2014-main.pdf, accessed 12 June 2017)

45 Wellcome Trust 2013 Wellcome Trust Monitor Wave 2: Tracking public views on science, biomedical research and science education. (see https://wellcome.ac.uk/sites/default/files/monitor-wave2-full-wellcome-may13.pdf, accessed 12 June 2017)

46 Wellcome Trust 2016 Wellcome Trust Monitor Summary Report Wave 3 Tracking public views on science and biomedical research. (see https://wellcome.ac.uk/sites/default/files/monitor-wave3-full-wellcome-apr16.pdf, accessed 12 June 2017)

Medical research carried out by commercial organisations was not as supported but still had high approval[47]. 54% of the public involved in the Ipsos MORI and Wellcome Trust *One-Way Mirror* study were supportive of the sharing of anonymised health care data with commercial organisations for medical research purposes, with only 26% opposed. Acceptance increased to 61% if this was the only way this certain type of research could be carried out, but a quarter would still rather the research not be carried out if it involved sharing data with a commercial organisation.

Non-research handling of data
There are many uses of data that do not relate to research. This section reviews information on public acceptance of such uses by various organisations. These include: uses of data for government such as for public services; surveillance; and companies using data to improve their products and services.

A number of studies looked at various possible uses of medical data. Participants in a Royal Statistical Society and Ipsos MORI study[48] were interested (77%) in medical records being shared with GPs so that they can provide appropriate care. A different study by the Royal Academy of Engineering[49] found that younger people wanted to be informed and engaged in the governance of electronic patient records. This dialogue and a Wellcome Trust 2016[50] study found that the sharing of medical records with insurance companies or private health care organisations was a red line.

Generally, studies have found that people were supportive of applications of data that could improve public services, not only in health, but also in other sectors such as transport, education, or to tackle crime[51,52,53,54,55]. However, many participants of an ESRC study[56] were concerned that people did not know who was holding data on them. Concerns raised at an Ipsos MORI *Dialogue on Data* included organisations holding data for longer than they should, inaccurate data, the use of linked administrative data to justify controversial policies such as the Bedroom Tax, and questions around the interplay of ideology and data use[57].

Mixed views were recorded on the idea of surveillance of personal data, including emails, text, phone and browsing history. In one study, 90% of UK participants were comfortable with surveillance being carried out to help national security[58]. However, mass surveillance justified by national security was less acceptable (56%) than targeted surveillance (88%). In an Ipsos *Global Trends* 2016 survey, half of the Britons surveyed found it unacceptable for government to use surveillance of their communications without consent, in the context of the immediate threat of a terrorist attack (only one third found it acceptable)[59,60].

47 Op. cit. 14
48 Op. cit. 34
49 Op. cit. 20
50 Op. cit. 14
51 Davidson S, et al. (research sponsored by Scottish Government Social Research) 2013 Public Acceptability of Data Sharing Between Public, Private and Third Sectors for Research Purposes. (see http://www.gov.scot/Publications/2013/10/1304/0, accessed 12 June 2017)
52 Op. cit. 9
53 Op. cit. 13
54 Op. cit. 22
55 Op. cit. 28
56 Op. cit. 9
57 Op. cit. 13
58 Op. cit. 8
59 Survey respondents were asked about the monitoring of: phone calls, text messages, emails and internet use.
60 Ipsos MORI 2016 Global Trends. (see https://www.ipsosglobaltrends.com/government-surveillance-of-internet-use/, accessed 12 June 2017)

There are varying opinions on the use of data in the commercial sector. A study by the Royal Statistical Society and Ipsos MORI[61] found that 7 in 10 respondents did not like online retailers using browsing history to target adverts. Another[62] found that the majority of participants had an understanding that if companies collected personal data on them, it might enable their organisation to run more effectively and efficiently, which would benefit consumers. When evaluating different applications of the technology, participants of the Royal Society public dialogue on machine learning[63] used several criteria, one of which was about who the beneficiaries would be. Where the benefit was felt to be more universal, such as in healthcare or education, views were more positive. If the sole purpose was making money for companies, then the application tended to be considered less valuable.

C. Trust in organisations to protect and use data

Trust in public bodies or other organisations was an important theme in public dialogues on data. Public dialogue activities have sought in particular to identify actions and processes carried out by organisations and public bodies that lead to a gain or loss of trust.

A Digital Catapult survey[64] asked *'who most beneficially uses your data?'* and 45% responded public services, 16% financial services, and 11% retail. Only 8% believed that the public sector used data about them without being clear they were doing so; compared with 30% for retail and 29% for media organisations. The Royal Statistical Society and Ipsos MORI[65] survey found only 4 to 7% respondents reporting a high level of trust with media, internet, telecommunications and insurance companies compared to 36% for NHS.

Two studies[66,67] found that participants fear that no data is safe following leaks of top secret government data in the US, both from an insider and outside threats. 64% of respondents to an EU Barometer survey on cybersecurity[68] agreed that they were concerned that information was not kept secure by public authorities; 70% agreed that they were concerned that their online personal information was not kept secure by websites more generally. Another study[69] found participants had concerns over the fallibility of humans and IT systems as a whole, and that they would not be able to keep the data they held about them secure. 76% of respondents to the EU Barometer study on cybersecurity agreed that the risk of becoming a victim of cybercrime had increased in the past year. While these concerns relating to security are informative in themselves, it is unclear how they would rank compared with, for example, more systemic forms of discrimination or political manipulation.

Trustworthy actions

Trusted actions included being transparent by having continuous communication and providing evidence of the secure storage and protection of personal data.

A 2013 Ipsos MORI and Deloitte study highlighted that people who were confident that companies tell them what data was collected, and how it was used, were between two and three times more likely to also be confident that companies handled, shared and used data about them to deliver personal benefits[70].

A 2015 study[71] found that the decision to consent to share personal information with an organisation was greatly affected by the ability of that organisation to keep the data secure (63% would consider providing information on the basis of their level of trust that the data would be secure).

61   Op. cit. 34
62   Op. cit. 9
63   Op. cit. 28
64   Op. cit. 5
65   Op. cit. 34
66   Op. cit. 13
67   Op. cit. 43
68   European Commission 2013 Special EU Barometer 404 Cyber Security.
     (see http://ec.europa.eu/public_opinion/archives/ebs/ebs_404_en.pdf , accessed 12 June 2017)
69   Op. cit. 20
70   Ipsos MORI (research sponsored by Deloitte) 2013 Data Nation 2013: Balancing Growth and Responsibility. (see https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/data-nation-2013-balancing-growth-and-responsibility.pdf, accessed 12 June 2017)
71   Op. cit. 23

In addition, participants of an ESRC dialogue[72] were reassured by the profile of the organisation and the professional attitude displayed by researchers who were going to be handling data about them.

Actions damaging trust

Generally, individuals were mistrustful when an organisation sold or shared data with a third party organisation without consent, or when there was a lack of communication or clarity from organisations in their motives for the use of personal data.

Several studies[73,74] found a strong sentiment that data should only be used for its primary purpose and should not be sold or shared to third party organisations. For example, though the data a supermarket collects about retail purchases could seem relatively harmless, there were concerns amongst the Royal Society dialogue participants about what would happen if that data was sold on in ways that had not been anticipated.

Mistrust in organisations can also be caused by a lack of clear communication about the reason or relevance for data collection or use. Some of the more sceptical participants of the Royal Society and Ipsos MORI public dialogue on machine learning[75] raised that a lack of communication from organisations left them concerned that they were trying to hide something.

Trust in public bodies

A Scottish Government Social Research report[76] found that the moral code of NHS employees and professionals was reassuring to individuals, however certain actions were causing some participants to be less trusting of the NHS. A common trust issue individuals have regarding their medical data is shared by the NHS or whether it could be a government body insurance companies or private with medical organisations. This often leads to concerns about unwanted contact from insurers or other medical organisations[77]. This was seen as a betrayal and raised concerns about potential discrimination.

The privatisation of parts of the NHS and the increasing involvement of private companies with the NHS for services and research was leading to 'blurred boundaries' and was a leading concern among participants of several studies[78,79].

Trust in private sector organisations

There was a trust deficit with commercial organisations and how they protect and use data. In a Scottish government study[80], many participants said that they generally trusted public bodies more than commercial organisations with information about them. A 2014 Ipsos MORI and Deloitte study found that only one third of UK adult internet users agreed that privacy policies were clear about how companies intended to use people's data, despite the majority of organisations adhering to best practice guidelines[81].

72   Op. cit. 9
73   Op. cit. 14
74   Op. cit. 28
75   Op. cit. 28
76   Op. cit. 43
77   Op. cit. 46
78   Op cit. 43
79   Op. cit. 14
80   Op. cit. 9
81   Ipsos MORI (research sponsored by Deloitte) 2014 Data Nation 2014: putting customers first.
     (see https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/deloitte-uk-data-nation-2014.pdf, accessed 12 June 2017)

In the Royal Society dialogue on machine learning, participants felt that public and private sectors should work together to develop the technology, progressing both commercial applications and broader research symbiotically and with approriate scrutiny[82]. Some participants were highly opposed to companies passing or selling on data about them to third party organisations – participants were keen that the data should be used for its original purpose – except if the data was passed on to universities or charities for research purposes.

A 2012 Demos and Populus study asked respondents a series of questions about the extent to which they were comfortable with how personal information and behavioural data were being used[83]. The highest level of comfort was for supermarket loyalty schemes: 27% respondents were comfortable with Tesco Clubcards, but only 10% were comfortable with Gmail scanning email content for the purposes of targeted advertising. Nearly one in two adults expected to be sharing more personal data with companies in ten years' time; similar numbers expected to do so with the government; two in ten expected to share about the same amount of data, and another two in ten expected to share less. There was a discrepancy between the consumers' concerns and their behaviours. Significant numbers of respondents shared information anyway; 85% of them used store loyalty cards, despite their worries. Some ethnographic studies have considered the use of data from sensors, for example in smart cities, to engage people in a different way, and gain insights into their behaviours[84].

D. New applications of data: machine learning and data science

Two studies, sponsored by the Royal Society and Government Data Science Partnership respectively, have investigated the awareness, understanding and acceptance of novel applications of data using new technologies such as machine learning[85,86]. Support for data science and machine learning applications depended on the specific use cases.

Participants could see that data science and machine learning would have benefits as well as risks for society and for individuals. The GDS Partnership study revealed that 47% of adults surveyed were comfortable with the government exploring new applications for data science, while 31% were cautious and believed that the government should not explore data science due to privacy risks. Demonstrating the potential impact of data science through real life case studies was crucial to engaging the public in discussions about opportunities for data science and machine learning. Participants in the Royal Society workshops on machine learning were hopeful that the technology could be used to address a number of key challenges such as climate change.

Most participants in the Royal Society workshops were less interested in the mechanics of machine learning than what it is being used for. They assumed that if algorithms did not work, then they would not be used. They wanted extensive testing of the performance of new applications of machine learning, to make sure for instance that driverless cars performed well under a range of conditions and to minimise any potential harm to humans.

82 Op. cit. 28

83 Op. cit. 7

84 Dourish P et al. 2016 Humbling data in a Playful World. DATA ETHNOGRAPHIES 3. (see https://dataethnographies.com/paper-iv-data-stories/, accessed 12 June 2017); Nafus D. 2016 Quantified: Biosensing Technologies in Everyday Life. MIT Press. ; Gabrys J. 2016 Program Earth: Environmental Sensing Technology and the Making of a Computational Planet. University of Minnesota Press. ; Couldry N and Powell A. 2014 Big Data from the Bottom Up. Big Data & Society. (DOI: 10.1177/2053951714539277)

85 Op. cit. 22

86 Op. cit. 28

Participants in the Royal Society's dialogue on machine learning recognised the potential of this technology to provide more accurate and less biassed analysis. In some applications – for example where the issue at hand was personally sensitive – they felt a human should be 'in the loop' for the final decision.

Another area of concern that came up during discussions of machine learning was about privacy, which has been a recurring theme in studies of how people have used and engaged with data. Of note, concerns about privacy began well before they became associated with computer surveillance[87].

E. Ethical and social challenges

Key social and ethical issues raised during a number of surveys and public dialogue activities fall into privacy, freedom, autonomy and self-determination, with some concerns specific to the use of new technologies such as machine learning.

Privacy

Privacy was a primary focus in a number of public surveys and dialogues, with three main themes: privacy and consent; privacy and national security; privacy and personalisation.

Consent and privacy issues are dominant in many studies. A survey found that 73% of participants agreed that the use of personal information without consent violates privacy[88]. Discussions of machine learning technologies brought up issues of privacy and consent[89]. Younger generations wanted to be able to control who accessed their health data, according to studies by the Royal Academy of Engineering and by Ipsos MORI and the Wellcome Trust[90,91]. According to an EU Barometer survey on *e-Privacy*, 54% UK respondents wanted a website to ask them for permission to use their personal data the first time they used it; 39% wanted the website to ask them each time[92]. The EU Barometer survey on *e-Privacy* found that 7 in 10 UK respondents wanted the default settings from their web browser to stop their information from being shared[93].

Surveys also explored to some extent the tensions between privacy and security. As quoted above about the use of data for surveillance, a study found half of the respondents were not comfortable with surveillance targeting their phones, emails and browsing history without their consent[94]. According to an EU Barometer survey on cybersecurity, 63% of UK respondents had changed one or several of their passwords online in the past year – the fourth highest rate among EU countries[95].

Studies showed mixed views about the use of individual data for the personalisation of services: not all uses justified the use of personal data. A GDS and Ipsos MORI study[96] highlighted that there is a fine line between what workshop participants consider an invasion of privacy and what they deem an acceptable use for personalisation – and it depends on the nature of the service. For example, the study found that using individual data for personalised careers services was more acceptable than using the same data to improve transport services.

---

87 The British Academy and The Royal Society 2017 Data management and use: Governance in the 21st century. (see https://royalsociety.org/topics-policy/projects/data-governance/, accessed June 2017); Agar J. 2003 The Government Machine: a Revolutionary History of the Computer. Cambridge MA: MIT Press.
88 Op. cit. 23
89 Op. cit. 28
90 Op. cit. 20
91 Op. cit. 14
92 Op. cit. 15
93 Op. cit. 15
94 Op. cit. 60
95 Op. cit. 68
96 Op. cit. 22

Freedom of choice

Personalisation also led into the discussion of choice and self-determination. Certain new data-powered technologies such as machine learning are able to personalise options and suggest those most likely to be chosen by an individual, by learning about their preferences. There was a concern from some participants of the Royal Society and Ipsos MORI dialogue exercise that this would restrict their freedom of choice. However, conversely, some participants believed that machine learning would help improve their experience when it came to making choices[97].

Further concern about choice can be highlighted by the findings of a Royal Statistical Society and Ipsos MORI study[98] where participants did not want organisations to use their browsing data to send targeted adverts.

Potential discrimination

The potential use of data science to profile and target certain sectors of society raised concerns.

This came up particularly in cases using administrative data[99,100], and the use of data to segment groups[101].

Agency

When data science and machine learning are used to make decisions that are critical for an individual or society, the studies that explored such scenarios revealed that participants wanted a human to be involved in the final decision[102,103].

F. Regulation and governance of data

The ICO and other organisations have monitored the views of the general public on current regulations and governance of data over the period 2008-16. They also looked at the public's suggestions on how to improve safeguards and regulations, and how to increase trust in organisations and participation in research studies.

Generally the ICO found there was low confidence in current laws and practices. Only 1 in 3 survey respondents, in 2011 and 2012-2013, believed that the laws and practices are providing sufficient protection[104,105].

Similarly, the 2012 Demos *Data Dialogue* found that there was high demand for a variety of reassurance measures overall[106]. Participants wanted in particular the 'ability to withdraw data' (73% respondents agreed) and to 'see what information [was] held on [them]' (70%).

Data Protection Act

The fraction of participants reporting having requested personal information that organisations might hold about them was generally low, with a peak at 23% in 2012-2013[107].

There were concerns raised in the Dialogue on Data study by Ipsos MORI[108] that the Data Protection Act (DPA) was not being enforced.

---

97 Op. cit. 28
98 Op. cit. 34
99 Op. cit. 13
100 Op. cit. 43
101 Op. cit. 14
102 Op. cit. 22
103 Op. cit. 28
104 Social and Market Strategic Research (research sponsored by the Information Commissioner's Office) 2011 Report on the Findings of the Information Commissioner's Office Annual Track 2011. (see https://ico.org.uk/media/about-the-ico/documents/1042360/annual-track-2011-individuals.pdf, accessed 12 June 2017)
105 Opinion Leader (research sponsored by the Information Commissioner's Office) 2013 Annual Track 2013: Individuals. (see https://ico.org.uk/media/about-the-ico/documents/1042195/annual-track-2012-individuals.pdf, accessed 12 June 2017)
106 Op. cit. 7
107 Op. cit. 105
108 Op. cit. 13

Freedom of Information Act

In 2014, only 8%[109] of those surveyed by the ICO had exercised their rights under the Freedom of Information Act (FOIA). However, the existence of the FOIA has increased confidence in public organisations (2014: 41% agreed; 2016: 57%), and trust in them (2014: 45%; 2016: 57%)[110]. Surveyed respondents agreed that it promoted accountability and transparency (2014: 68%; 2016: 71%) and increased their knowledge of what organisations do (2014: 63%; 2016: 70%).

75% of respondents very much agreed that it was important for private companies who are conducting work on behalf of public authorities to be subject to the FOIA[111].

Data science and machine learning

While participants in the Royal Society's dialogues on machine learning expressed a general desire for some form of scrutiny of the advancement of machine learning, there was no clear consensus about what this should look like in practice[112]. The breadth of applications of machine learning made it difficult to come to a general view on such oversight. Participants saw roles for the private sector, government and independent actors in providing this.

Suggestions for future governance and regulations

81% of participants in the 2014 ICO Annual Track survey supported the introduction of a new certification mark for data protection[113].

Several studies indicated that people wanted more transparency[114,115,116,117]. Some members of the public involved mentioned the need to find ways to track exactly who has data about them and why. Participants of the Ipsos MORI and GDS study welcomed the fact that the government committed to transparency regarding their ethical guidelines around data science, and the opportunity to discuss during the workshop the potential use of data science within government[120].

In the Wellcome Trust *One-Way Mirror* study, participants discussed how the process of giving consent to data sharing could be improved. The members of the public involved wished that healthcare professionals such as GPs would be trained to explain how research and consent work[121].

109 ComRes (research sponsored by the Information Commissioner's Office) 2014 Annual Track 2014: Individuals (Topline findings). (see https://ico.org.uk/media/about-the-ico/documents/1043485/annual-track-september-2014-individuals.pdf, accessed 12 June 2017)

110 Op. cit. 32

111 Op. cit. 109

112 Op. cit. 28

113 Op. cit. 109

114 National Data Guardian 2016 Review of Data Security, Consent and Opt-Outs. (see https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF, accessed 12 June 2017)

115 Op. cit. 28

116 Op. cit. 22

117 Illuminas (research sponsored by Citizens Advice) 2016 Consumer expectations for personal data management in the digital world. (see https://www.citizensadvice.org.uk/Global/CitizensAdvice/Consumer%20publications/Personal%20data%20consumer%20expectations%20research.docx.pdf, accessed 12 June 2017)

118 Op. cit. 117

119 Op. cit. 28

120 Op. cit. 22

121 Op. cit. 14

G. Segmentation in public views
Studies on the public perception on data and its use have identified segments of the public that tend to display certain behaviours or hold certain opinions.

Attitudes also depended on the previous exposure to examples of data use. For example, a study highlighted that there were no differences linked with age when it came to the acceptability of the use of data science in government[122]. Instead, acceptability depended on whether individuals used government services and on their views regarding the relationship between the citizen and the state.

Variety and diversity of attitudes and behaviours
The 2012 Demos *Data Dialogue* found that there was no single attitude to sharing personal information. Members of the public surveyed fell into one of five categories, each characterised by a distinct set of views about personal information: around 30% were 'non sharers'; 22% were 'sceptics'; 20% were 'pragmatists'; 19% were 'value hunters'; 8% were 'enthusiastic sharers'[123]. The GDS 2016 study on government using data science identified four categories along the same spectrum: the 'data wary', 'data pragmatists', 'data adapters' and 'data adopters'[124]. The 'data adopters' (23% of adults) supported using data science for research purposes and saw the value in how individual level data can generate better insight; the 'data adapters' (28% of adults) responded best to uses which improved services for individuals and uses of non-sensitive data; the 'data pragmatists' (27% of adults) were more ambivalent in their views, wanting government to explore new ways of using data but were most comfortable using data for high-level statistics rather than advanced data science; the 'data wary' (22% of adults) were the least likely to identify opportunities for data science, they applied caution to the principle of data science, based on concerns around privacy and effectiveness, or a desire for further information.

Digital footprint and digital dependence
The size of the digital footprint[125] and dependence on digital services[126] differed by age, a SAS study found[127]. Young millennials (18-24) tended to have the highest digital dependency compared to older millennials (25-29), but older millennials tended to have larger digital footprints. The difference in digital footprint between young millennials and older millennials in the study was put down to loyalty programmes – older millennials were more likely to sign up thus generating a larger digital footprint as defined in this study. For individuals over 30, the footprint and interaction with digital services decreased with age, and was particularly low for those aged 60+.

Awareness and exposure to data use
More than a quarter of 15 to 34 year-olds involved in a 2013 Ipsos MORI and Deloitte study were not aware that companies collect data about them and their activities despite these people generating sizeable digital footprints[128].

The Wellcome Trust and Ipsos MORI *One-Way Mirror* study[129] found that younger generations wanted more control over information about them, but trends in opinions about health data sharing were not restricted to specific generations. Instead they were increasingly dependent on exposure and knowledge of a specific use of data. They found that the more knowledge or exposure with a certain aspect of data use, such as the conditions for sharing with commercial organisations, the greater the acceptance.

Health professionals have a greater awareness of data permissions and ownership of data compared to the general public. With increased awareness of data and the issues surrounding it, Ipsos MORI[130] also found that both health care professionals and patients were better able to tease out issues such as quality of data and data collection.

---

122 Op. cit. 22

123 Op. cit. 7

124 Op. cit. 22

125 The digital footprint was quantified based on multiple parameters including the survey respondents' use of social media, smart phones and tablets, their online purchasing habits, and their participation in loyalty programs.

126 The dependence on digital service is understood here as a qualitative assessment of the reliance on digital services in general.

127 Op. cit. 23

128 Op. cit. 70

129 Op. cit. 14

130 Op. cit. 14

Generational differences were not seen in discussions on the support for data science applications in government, however clear differences in opinion were associated with varying degrees of use of government services and differing views about the relationship between the citizen and the state[131].

Trust in the organisations and public bodies
A Royal Statistical Society and Ipsos MORI study showed that younger generations had greater trust in both government and commercial internet companies (4.8 in 10 16-to-24 year-olds trust the UK government compared to 4 in 10 55-to-75 year-olds; 4.5 in 10 16-to-24 year-olds trusted internet companies compared to 3.4 in 10 55-to-75 year-olds)[132].

Privacy
Regarding privacy and consent, the youngest EU respondents (aged 15-24) of the EU Barometer *e-Privacy* survey were the least likely to prefer to be asked for permission each time they enter a website (31% versus 39%-42% for other age categories), and the most likely to prefer to be asked the first time they enter the website, with the option to change their mind (61% versus 38%-51%)[133]. The oldest respondents (55+) were the most likely to say they do not want to share their personal information (16% versus 7%-9%). Of note, the study did not address underlying differences between the groups, and it is unclear in particular whether these are due to generational differences or cohort effects.

In the UK, the ICO Annual Track 2016 survey showed that older generations were more likely to actively protect their data than millennials: baby boomers were more likely than millennials to check bank statements for irregular activity (87% versus 56%), use different passwords and PINs online (74% versus 44%), or use anti-virus, firewalls or anti-spam software (84% versus 47%)[134].

When it comes to surveillance and security, a study conducted by Cardiff University revealed that 18-to-59 year-olds tend to have a greater concern over state surveillance of emails, browsing history, mobile phones and social media, whereas those aged over 60 had a more positive and accepting view of surveillance[135].

Knowledge of Regulations and Governance
The ICO annual track survey in 2016[136] found the following segmentation in awareness and use of the Data Protection Act (DPA): more baby boomers (58%) were aware of rights under the DPA compared with millennials (47%) and generation X (49%). People from social grades[137] at the upper end of the scale were more likely to have requested personal information from organisations (18% of grade AB; 13% grade C1; 12% C2; 10% DE).

131 Op. cit. 22
132 Op. cit. 34
133 Op. cit. 15
134 Op. cit. 32
135 Op. cit. 8
136 Op. cit. 32
137 Ipsos MORI 2015 Social Grade: a Classification Tool. (see https://www.ipsos.com/sites/default/files/publication/6800-03/MediaCT_thoughtpiece_Social_Grade_July09_V3_WEB.pdf, accessed 12 June 2017)

The number of participants who agreed that information about public organisations are available and accessible was higher in Scotland (45%) than in other UK countries (Wales: 40%; England: 33%; Northern Ireland: 27%). The proportion was also higher among baby boomers (38% versus 30% for millennials, 34% for generation X), and among social grades at the upper end of the scale (AB: 38%; C1: 36%; C2: 29%; DE: 33%).

The ICO annual track survey in 2011[138] found that only 27% of participants were aware of the right to request information held by the government (23% of participants aged 18-24). When prompted, those aged 18-24 had a lower awareness of this right, with 67% recognising it compared to 83% of those aged 55-64, and an average of 77% for all groups combined. Respondents from social-economic group AB tended to have a higher spontaneous and prompted awareness of the FOIA in 2008 and 2009[139].

In 2016, only 62% of millennials and 61% of respondents from social grades C2/D/E thought it was important to have independent regulation of the DPA (compared with 66% of all groups considered)[140]. Similarly, only 63% of millennials and 65% of respondents from social grades C2/D/E deemed it important that the FOIA had an independent regulator (compared with 71% all groups considered).

---

138 Op. cit. 104

139 Social and Market Strategic Research (research sponsored by the Information Commissioner's Office) 2009 Report on the Findings of the Information Commissioner's Office Annual Track 2009. (see https://ico.org.uk/media/about-the-ico/documents/1042379/ico-annual-tracking-individuals-final-report2009.pdf, accessed 12 June 2017)

140 Op. cit. 32

The Royal Society is a self-governing Fellowship of many of the world's most distinguished scientists drawn from all areas of science, engineering, and medicine. The Society's fundamental purpose, as it has been since its foundation in 1660, is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

The Society's strategic priorities emphasise its commitment to the highest quality science, to curiosity-driven research, and to the development and use of science for the benefit of society. These priorities are:

- Promoting excellence in science
- Supporting international collaboration
- Demonstrating the importance of science to everyone

For further information
The Royal Society
6 – 9 Carlton House Terrace
London SW1Y 5AG

T   +44 20 7451 2500
E   science.policy@royalsociety.org
W   royalsociety.org

Registered Charity No 207043

The British Academy is the UK's national body for the humanities and social sciences – the study of peoples, cultures and societies, past, present and future. We have three principle roles: as an independent Fellowship of world-leading scholars and researchers; a Funding Body that supports the best ideas, nationally and internationally; and a Forum for debate and engagement – a voice that champions the humanities and social sciences.

For further information
The British Academy
10 – 11 Carlton House Terrace
London SW1Y 5AH

T   +44 20 7969 5200
W   britishacademy.ac.uk

 @britac_news
 @TheBritishAcademy
 Britacfilm
 BritishAcademy

Registered Charity No 233176